Bruce K. Driver

Analysis Tools with Examples

March 10, 2004 *File:tempanal.tex*

Springer
Berlin Heidelberg New York
Hong Kong London
Milan Paris Tokyo

Contents

Part I

Back Ground Material

1

Introduction / User Guide

Not written as of yet. Topics to mention.

1. A better and more general integral.
   a) Convergence Theorems
   b) Integration over diverse collection of sets. (See probability theory.)
   c) Integration relative to different weights or densities including singular weights.
   d) Characterization of dual spaces.
   e) Completeness.
2. Infinite dimensional Linear algebra.
3. ODE and PDE.
4. Harmonic and Fourier Analysis.
5. Probability Theory

*** Beginning of WORK material. ***

1.1 Topology beginnings

Recall the notion of a topology by extrapolating from the open sets on $\mathbb{R}^2$. Also recall what it means to be continuous, namely $f : X \to \mathbb{R}$ is continuous at x if for all $\varepsilon > 0$ there exists $V \in \tau_x$ such that

$$f(V) \subset f(x) + (-\varepsilon, \varepsilon).$$

1.2 A Better Integral and an Introduction to Measure Theory

Let $a, b \in \mathbb{R}$ with $a < b$ and let

$$I^0(f) := \int_a^b f(t)dt \text{ for all } f \in C([a,b])$$

denote the Riemann integral. Also let $\mathcal{H}$ denote the smallest **linear subspace** of bounded functions on $[a,b]$ which is closed under bounded convergence and contains $C([a,b])$. Such a space exists since we can take the intersection over all such spaces of functions.

Theorem 1.1. *There is an extension I of I^0 to $\mathcal{H}$ such that I is still linear and $\lim_{n\to\infty} I(f_n) = I(f)$ for all $f_n \in \mathcal{H}$ with $f_n \to f$ boundedly. Moreover this extension is unique and is* ***positive*** *in the sense that $I(f) \geq 0$ if $f \in \mathcal{H}$ and $f \geq 0$.*

Proof. We will only prove the uniqueness here. Suppose that J and I are two such extensions and let

$$\mathcal{K} := \{f \in \mathcal{H} : J(f) = I(f)\}.$$

Then $\mathcal{K}$ is a linear subspace closed under bounded convergence which contains $C([a,b])$ and hence $\mathcal{K} = \mathcal{H}$. The existence of I is the hard part. The positivity of I can be seen from the existence construction. ■

Example 1.2. Here are some examples of functions in $\mathcal{H}$ and their integrals:

1. Suppose $[\alpha,\beta] \subset [a,b]$, then $1_{[\alpha,\beta]} \in \mathcal{H}$ and $I\left(1_{[\alpha,\beta]}\right) = \beta - \alpha$.(Draw a picture.)
2. $I\left(1_{\{\alpha\}}\right) = 0$.
3. The space $\mathcal{H}$ is an algebra, i.e. if $f, g \in \mathcal{H}$ then $fg \in \mathcal{H}$. To prove this, first assume that $f \in C([a,b])$ and let
$$\mathcal{H}_f := \{g \in \mathcal{H} : fg \in \mathcal{H}\}.$$
Then $\mathcal{H}_f$ is closed under bounded convergence and contains $C([a,b])$ and hence $\mathcal{H}_f = \mathcal{H}$, i.e. the product of a continuous function and an element in $\mathcal{H}$ is back in $\mathcal{H}$.
Now suppose that $f \in \mathcal{H}$ and again let $\mathcal{H}_f$ be as above. By the same reasoning we may show again that $\mathcal{H}_f = \mathcal{H}$ and this proves the assertion.
4. If $f \in \mathcal{H}$ and $\phi \in C(\mathbb{R})$, then $\phi \circ f \in \mathcal{H}$. This a consequence of the Weierstrass approximation Theorem 8.34. In particular $|f| \in \mathcal{H}$ and $f_\pm := \frac{|f| \pm f}{2} \in \mathcal{H}$ if $f \in \mathcal{H}$.
5. If $f_n \in \mathcal{H}$, $f_n \geq 0$ and $f = \sum_{n=1}^\infty f_n$ is a bounded function, then $f \in \mathcal{H}$ and
$$I(f) = \sum_{n=1}^\infty I(f_n). \tag{1.1}$$
To prove Eq. (1.1) we have
$$\sum_{n=1}^\infty I(f_n) = \lim_{N\to\infty} I\left(\sum_{n=1}^N f_n\right) = I(f).$$

6. As an example of item 4., $1_{\mathbb{Q}\cap[a,b]} = \sum_{n=1}^{\infty} 1_{\{\alpha_n\}} \in \mathcal{H}$ and $I\left(1_{\mathbb{Q}\cap[a,b]}\right) = 0$. Here $\{\alpha_n\}_{n=1}^{\infty}$ is an enumeration of the rational numbers in the interval $[a,b]$.
7. Let $\mathcal{M} := \{A \subset [a,b] : 1_A \in \mathcal{H}\}$ and for $A \in \mathcal{M}$ let $m(A) := I(1_A)$. Then $\mathcal{M}$ and m have the following properties:
   a) $\emptyset, [a,b] \in \mathcal{M}$ and $m(\emptyset) = 0$ and $m([a,b]) = b-a$. Moreover $m(A) \geq 0$ for all $A \in \mathcal{M}$.
   b) If $A \in \mathcal{M}$ then $A^c \in \mathcal{M}$ and $m(A^c) = b - a - m(A)$. This follows from the fact that $1_{A^c} = 1 - 1_A$.
   c) If $A, B \in \mathcal{M}$, then $A \cap B \in \mathcal{M}$ since if $1_{A\cap B} = 1_A \cdot 1_B$ and $\mathcal{H}$ is an algebra.
   Definition: a collection of sets $\mathcal{M}$ satisfying a) – c) is called an **algebra** of subsets of $[a,b]$.
   d) More generally if $A_n \in \mathcal{M}$ then $\cap A_n \in \mathcal{M}$ since $1_{\cap A_n} = \lim_{N\to\infty} 1_{A_1} \cdots 1_{A_N}$ and the convergence is bounded.
   Definition: a collection of sets $\mathcal{M}$ satisfying a) – d) is called an σ – **algebra.**
   e) If $A_n \in \mathcal{M}$, then $\cup A_n \in \mathcal{M}$. Indeed we know $\cup A_n \in \mathcal{M}$ iff $(\cup A_n)^c \in \mathcal{M}$. But
   $$(\cup A_n)^c = \cap A_n^c \in \mathcal{M}$$
   by item d. above.
   f) If $A_n \in \mathcal{M}$ are pairwise disjoint, then
   $$m(\cup A_n) = \sum_{n=1}^{\infty} m(A_n).$$
   To prove this it suffices to observe that $1_{\cup A_n} = \sum_{n=1}^{\infty} 1_{A_n}$.
   g) $\mathcal{M}$ is not $2^{[a,b]}$, i.e. $\mathcal{M}$ is not all subset of $[a,b]$. This is not obvious and it is not possible to really write down an "explicit" subset $[a,b]$ which is not in $\mathcal{M}$. We will prove the existence of such sets later.
8. **Fact:** $\mathcal{M}$ is the smallest σ – algebra on $[a,b]$ which contains all subintervals of $[a,b]$.
9. **Fact:** A bounded function $f : [a,b] \to \mathbb{R}$ is in $\mathcal{H}$ iff $\{f > \alpha\} \in \mathcal{M}$ for all $\alpha \in \mathbb{R}$.
10. **Fact:** The integral I may be recovered from the measure m by the formula
$$I(f) = \lim_{\text{mesh}\to 0} \sum_{0<\alpha_1<\alpha_2<\alpha_3<\ldots}^{\infty} \alpha_i m\left(\{x \in [a,b] : \alpha_i < f(x) \leq \alpha_i\}\right).$$

We will prove items 8. – 10. later in the course. The proof if Items 9. and 10. is not so hard and the energetic reader may wish to give them a try.

Notation 1.3 *The collection of sets $\mathcal{M}$ is called the Borel σ – algebra on $[a,b]$ and the function $m : \mathcal{M} \to \mathbb{R}$ is called Lebesgue measure. We will usually* *write* $I(f)$ *as* $\int_{[a,b]} f dm$ *and* $I(f)$ *will be called the Lebesgue integral of* f*. This integral may be extended to all positive functions* f *such that* $f 1_{|f| \le M} \in \mathcal{H}$ *for all* M *by*

$$I(f) = \lim_{M\to\infty} I\left(f 1_{|f|\le M}\right).$$

Again, we will come back to all of this again later.

*** End of WORK material. ***

2

Set Operations

Let $\mathbb{N}$ denote the positive integers, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ be the non-negative integers and $\mathbb{Z} = \mathbb{N}_0 \cup (-\mathbb{N})$ – the positive and negative integers including 0, $\mathbb{Q}$ the rational numbers, $\mathbb{R}$ the real numbers (see Chapter 3 below), and $\mathbb{C}$ the complex numbers. We will also use $\mathbb{F}$ to stand for either of the fields $\mathbb{R}$ or $\mathbb{C}$.

Notation 2.1 *Given two sets X and Y, let Y^X denote the collection of all functions $f : X \to Y$. If $X = \mathbb{N}$, we will say that $f \in Y^{\mathbb{N}}$ is a sequence with values in Y and often write f_n for $f(n)$ and express f as $\{f_n\}_{n=1}^{\infty}$. If $X = \{1, 2, \ldots, N\}$, we will write Y^N in place of $Y^{\{1,2,\ldots,N\}}$ and denote $f \in Y^N$ by $f = (f_1, f_2, \ldots, f_N)$ where $f_n = f(n)$.*

Notation 2.2 *More generally if $\{X_\alpha : \alpha \in A\}$ is a collection of non-empty sets, let $X_A = \prod_{\alpha \in A} X_\alpha$ and $\pi_\alpha : X_A \to X_\alpha$ be the canonical projection map defined by $\pi_\alpha(x) = x_\alpha$. If If $X_\alpha = X$ for some fixed space X, then we will write $\prod_{\alpha \in A} X_\alpha$ as X^A rather than X_A.*

Recall that an element $x \in X_A$ is a "**choice function**," i.e. an assignment $x_\alpha := x(\alpha) \in X_\alpha$ for each $\alpha \in A$. The **axiom of choice** (See Appendix B.) states that $X_A \neq \emptyset$ provided that $X_\alpha \neq \emptyset$ for each $\alpha \in A$.

Notation 2.3 *Given a set X, let 2^X denote the **power set** of X – the collection of all subsets of X including the empty set.*

The reason for writing the power set of X as 2^X is that if we think of 2 meaning $\{0, 1\}$, then an element of $a \in 2^X = \{0, 1\}^X$ is completely determined by the set

$$A := \{x \in X : a(x) = 1\} \subset X.$$

In this way elements in $\{0, 1\}^X$ are in one to one correspondence with subsets of X.

For $A \in 2^X$ let

$$A^c := X \setminus A = \{x \in X : x \notin A\}$$

and more generally if $A, B \subset X$ let

$$B \setminus A := \{x \in B : x \notin A\} = A \cap B^c.$$

We also define the symmetric difference of A and B by

$$A \triangle B := (B \setminus A) \cup (A \setminus B).$$

As usual if $\{A_\alpha\}_{\alpha \in I}$ is an indexed collection of subsets of X we define the union and the intersection of this collection by

$$\cup_{\alpha \in I} A_\alpha := \{x \in X : \exists\, \alpha \in I \ \ni x \in A_\alpha\} \text{ and}$$
$$\cap_{\alpha \in I} A_\alpha := \{x \in X : x \in A_\alpha \, \forall\, \alpha \in I\}.$$

Notation 2.4 *We will also write* $\coprod_{\alpha \in I} A_\alpha$ *for* $\cup_{\alpha \in I} A_\alpha$ *in the case that* $\{A_\alpha\}_{\alpha \in I}$ *are pairwise disjoint, i.e.* $A_\alpha \cap A_\beta = \emptyset$ *if* $\alpha \neq \beta$.

Notice that $\cup$ is closely related to $\exists$ and $\cap$ is closely related to $\forall$. For example let $\{A_n\}_{n=1}^\infty$ be a sequence of subsets from X and define

$$\{A_n \text{ i.o.}\} := \{x \in X : \#\{n : x \in A_n\} = \infty\} \text{ and}$$
$$\{A_n \text{ a.a.}\} := \{x \in X : x \in A_n \text{ for all } n \text{ sufficiently large}\}.$$

(One should read $\{A_n \text{ i.o.}\}$ as A_n infinitely often and $\{A_n \text{ a.a.}\}$ as A_n almost always.) Then $x \in \{A_n \text{ i.o.}\}$ iff

$$\forall N \in \mathbb{N}\, \exists\, n \geq N \ni x \in A_n$$

and this may be expressed as

$$\{A_n \text{ i.o.}\} = \cap_{N=1}^\infty \cup_{n \geq N} A_n.$$

Similarly, $x \in \{A_n \text{ a.a.}\}$ iff

$$\exists\, N \in \mathbb{N} \ni \forall\, n \geq N, \ \ x \in A_n$$

which may be written as

$$\{A_n \text{ a.a.}\} = \cup_{N=1}^\infty \cap_{n \geq N} A_n.$$

Definition 2.5. *A set* X *is said to be* ***countable*** *if is empty or there is an injective function* $f : X \to \mathbb{N}$, *otherwise* X *is said to be* ***uncountable.***

Lemma 2.6 (Basic Properties of Countable Sets).

1. *If* $A \subset X$ *is a subset of a countable set* X *then* A *is countable.*
2. *Any infinite subset* $\Lambda \subset \mathbb{N}$ *is in one to one correspondence with* $\mathbb{N}$.

3. *A non-empty set X is countable iff there exists a surjective map, $g : \mathbb{N} \to X$.*
4. *If X and Y are countable then $X \times Y$ is countable.*
5. *Suppose for each $m \in \mathbb{N}$ that A_m is a countable subset of a set X, then $A = \cup_{m=1}^{\infty} A_m$ is countable. In short, the countable union of countable sets is still countable.*
6. *If X is an infinite set and Y is a set with at least two elements, then Y^X is uncountable. In particular 2^X is uncountable for any infinite set X.*

Proof. 1. If $f : X \to \mathbb{N}$ is an injective map then so is the restriction, $f|_A$, of f to the subset A. 2. Let $f(1) = \min \Lambda$ and define f inductively by

$$f(n+1) = \min \Lambda \setminus \{f(1), \ldots, f(n)\}.$$

Since Λ is infinite the process continues indefinitely. The function $f : \mathbb{N} \to \Lambda$ defined this way is a bijection. 3. If $g : \mathbb{N} \to X$ is a surjective map, let

$$f(x) = \min g^{-1}(\{x\}) = \min \{n \in \mathbb{N} : f(n) = x\}.$$

Then $f : X \to \mathbb{N}$ is injective which combined with item 2. (taking $\Lambda = f(X)$) shows X is countable. Conversely if $f : X \to \mathbb{N}$ is injective let $x_0 \in X$ be a fixed point and define $g : \mathbb{N} \to X$ by $g(n) = f^{-1}(n)$ for $n \in f(X)$ and $g(n) = x_0$ otherwise. 4. Let us first construct a bijection, h, from $\mathbb{N}$ to $\mathbb{N} \times \mathbb{N}$. To do this put the elements of $\mathbb{N} \times \mathbb{N}$ into an array of the form

$$\begin{pmatrix} (1,1) & (1,2) & (1,3) & \ldots \\ (2,1) & (2,2) & (2,3) & \ldots \\ (3,1) & (3,2) & (3,3) & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

and then "count" these elements by counting the sets $\{(i,j) : i + j = k\}$ one at a time. For example let $h(1) = (1,1)$, $h(2) = (2,1)$, $h(3) = (1,2)$, $h(4) = (3,1)$, $h(5) = (2,2)$, $h(6) = (1,3)$, etc. etc. If $f : \mathbb{N} \to X$ and $g : \mathbb{N} \to Y$ are surjective functions, then the function $(f \times g) \circ h : \mathbb{N} \to X \times Y$ is surjective where $(f \times g)(m,n) := (f(m), g(n))$ for all $(m,n) \in \mathbb{N} \times \mathbb{N}$. 5. If $A = \emptyset$ then A is countable by definition so we may assume $A \neq \emptyset$. With out loss of generality we may assume $A_1 \neq \emptyset$ and by replacing A_m by A_1 if necessary we may also assume $A_m \neq \emptyset$ for all m. For each $m \in \mathbb{N}$ let $a_m : \mathbb{N} \to A_m$ be a surjective function and then define $f : \mathbb{N} \times \mathbb{N} \to \cup_{m=1}^{\infty} A_m$ by $f(m,n) := a_m(n)$. The function f is surjective and hence so is the composition, $f \circ h : \mathbb{N} \to X \times Y$, where $h : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ is the bijection defined above. 6. Let us begin by showing $2^{\mathbb{N}} = \{0,1\}^{\mathbb{N}}$ is uncountable. For sake of contradiction suppose $f : \mathbb{N} \to \{0,1\}^{\mathbb{N}}$ is a surjection and write $f(n)$ as $(f_1(n), f_2(n), f_3(n), \ldots)$. Now define $a \in \{0,1\}^{\mathbb{N}}$ by $a_n := 1 - f_n(n)$. By construction $f_n(n) \neq a_n$ for all n and so $a \notin f(\mathbb{N})$. This contradicts the assumption that f is surjective and shows $2^{\mathbb{N}}$ is uncountable. For the general case, since $Y_0^X \subset Y^X$ for any subset $Y_0 \subset Y$, if Y_0^X is uncountable then so is Y^X. In this way we may assume Y_0 is a two point set which may as well be $Y_0 = \{0,1\}$. Moreover, since X is an infinite set we may find an injective map $x : \mathbb{N} \to X$ and use this to set up an injection, $i : 2^{\mathbb{N}} \to 2^X$ by setting $i(a)(x_n) = a_n$ for all $n \in \mathbb{N}$ and $i(a)(x) = 0$ if $x \notin \{x_n : n \in \mathbb{N}\}$. If 2^X were countable we could find a surjective map $f : 2^X \to \mathbb{N}$ in which case $f \circ i : 2^{\mathbb{N}} \to \mathbb{N}$ would be surjective as well. However this is impossible since we have already seed that $2^{\mathbb{N}}$ is uncountable. ■

We end this section with some notation which will be used frequently in the sequel.

Notation 2.7 *If $f : X \to Y$ is a function and $\mathcal{E} \subset 2^Y$ let*

$$f^{-1}\mathcal{E} := f^{-1}(\mathcal{E}) := \{f^{-1}(E) | E \in \mathcal{E}\}.$$

If $\mathcal{G} \subset 2^X$, let

$$f_*\mathcal{G} := \{A \in 2^Y | f^{-1}(A) \in \mathcal{G}\}.$$

Definition 2.8. *Let $\mathcal{E} \subset 2^X$ be a collection of sets, $A \subset X$, $i_A : A \to X$ be the **inclusion map** ($i_A(x) = x$ for all $x \in A$) and*

$$\mathcal{E}_A = i_A^{-1}(\mathcal{E}) = \{A \cap E : E \in \mathcal{E}\}.$$

2.1 Exercises

Let $f : X \to Y$ be a function and $\{A_i\}_{i\in I}$ be an indexed family of subsets of Y, verify the following assertions.

Exercise 2.1. $(\cap_{i\in I} A_i)^c = \cup_{i\in I} A_i^c$.

Exercise 2.2. Suppose that $B \subset Y$, show that $B \setminus (\cup_{i\in I} A_i) = \cap_{i\in I}(B \setminus A_i)$.

Exercise 2.3. $f^{-1}(\cup_{i\in I} A_i) = \cup_{i\in I} f^{-1}(A_i)$.

Exercise 2.4. $f^{-1}(\cap_{i\in I} A_i) = \cap_{i\in I} f^{-1}(A_i)$.

Exercise 2.5. Find a counter example which shows that $f(C \cap D) = f(C) \cap f(D)$ need not hold.

3

A Brief Review of Real and Complex Numbers

Although it is assumed that the reader of this book is familiar with the properties of the real numbers, $\mathbb{R}$, nevertheless I feel it is instructive to define them here and sketch the development of their basic properties. It will most certainly be assumed that the reader is familiar with basic algebraic properties of the natural numbers $\mathbb{N}$ and the ordered field of rational numbers,

$$\mathbb{Q} = \left\{ \frac{m}{n} : m, n \in \mathbb{Z} : n \neq 0 \right\}.$$

As usual, for $q \in \mathbb{Q}$, we define

$$|q| = \begin{cases} q & \text{if } q \geq 0 \\ -q & \text{if } q \leq 0. \end{cases}$$

Notice that if $q \in \mathbb{Q}$ and $|q| \leq \frac{1}{n}$ for all n, then $q = 0$. Indeed $q \neq 0$ then $|q| = \frac{m}{n}$ for some $m, n \in \mathbb{N}$ and hence $|q| \geq \frac{1}{n}$. A similar argument shows $q \geq 0$ iff $q \geq -\frac{1}{n}$ for all $n \in \mathbb{N}$. These trivial remarks will be used in the future without further reference.

Definition 3.1. *A sequence* $\{q_n\}_{n=1}^{\infty} \subset \mathbb{Q}$ ***converges*** *to* $q \in \mathbb{Q}$ *if* $|q - q_n| \to 0$ *as* $n \to \infty$, *i.e. if for all* $N \in \mathbb{N}$, $|q - q_n| \leq \frac{1}{N}$ *for a.a.* n. *As usual if* $\{q_n\}_{n=1}^{\infty}$ *converges to* q *we will write* $q_n \to q$ *as* $n \to \infty$ *or* $q = \lim_{n\to\infty} q_n$.

Definition 3.2. *A sequence* $\{q_n\}_{n=1}^{\infty} \subset \mathbb{Q}$ *is* ***Cauchy*** *if* $|q_n - q_m| \to 0$ *as* $m, n \to \infty$. *More precisely we require for each* $N \in \mathbb{N}$ *that* $|q_m - q_n| \leq \frac{1}{N}$ *for a.a. pairs* (m, n).

Exercise 3.1. Show that all convergent sequences $\{q_n\}_{n=1}^{\infty} \subset \mathbb{Q}$ are Cauchy and that all Cauchy sequences $\{q_n\}_{n=1}^{\infty}$ are bounded – i.e. there exists $M \in \mathbb{N}$ such that

$$|q_n| \leq M \text{ for all } n \in \mathbb{N}.$$

Exercise 3.2. Suppose $\{q_n\}_{n=1}^{\infty}$ and $\{r_n\}_{n=1}^{\infty}$ are Cauchy sequences in $\mathbb{Q}$.

1. Show $\{q_n + r_n\}_{n=1}^{\infty}$ and $\{q_n \cdot r_n\}_{n=1}^{\infty}$ are Cauchy.
   Now assume that $\{q_n\}_{n=1}^{\infty}$ and $\{r_n\}_{n=1}^{\infty}$ are convergent sequences in $\mathbb{Q}$.
2. Show $\{q_n + r_n\}_{n=1}^{\infty}$ $\{q_n \cdot r_n\}_{n=1}^{\infty}$ are convergent in $\mathbb{Q}$ and
$$\lim_{n\to\infty}(q_n + r_n) = \lim_{n\to\infty} q_n + \lim_{n\to\infty} r_n \text{ and}$$
$$\lim_{n\to\infty}(q_n r_n) = \lim_{n\to\infty} q_n \cdot \lim_{n\to\infty} r_n.$$
3. If we further assume $q_n \le r_n$ for all n, show $\lim_{n\to\infty} q_n \le \lim_{n\to\infty} r_n$. (It suffices to consider the case where $q_n = 0$ for all n.)

The rational numbers $\mathbb{Q}$ suffer from the defect that they are not complete, i.e. not all Cauchy sequences are convergent. In fact, according to Corollary 3.14 below, "most" Cauchy sequences of rational numbers do not converge to a rational number.

Exercise 3.3. Use the following outline to construct a Cauchy sequence $\{q_n\}_{n=1}^{\infty} \subset \mathbb{Q}$ which is **not** convergent in $\mathbb{Q}$.

1. Recall that there is no element $q \in \mathbb{Q}$ such that $q^2 = 2$.[1] To each $n \in \mathbb{N}$ let $m_n \in \mathbb{N}$ be chosen so that
$$\frac{m_n^2}{n^2} < 2 < \frac{(m_n+1)^2}{n^2} \tag{3.1}$$
and let $q_n := \frac{m_n}{n}$.
2. Verify that $q_n^2 \to 2$ as $n \to \infty$ and that $\{q_n\}_{n=1}^{\infty}$ is a Cauchy sequence in $\mathbb{Q}$.
3. Show $\{q_n\}_{n=1}^{\infty}$ does not have a limit in $\mathbb{Q}$.

3.1 The Real Numbers

Let $\mathcal{C}$ denote the collection of Cauchy sequences $a = \{a_n\}_{n=1}^{\infty} \subset \mathbb{Q}$ and say $a, b \in \mathcal{C}$ are equivalent (write $a \sim b$) iff $\lim_{n\to\infty} |a_n - b_n| = 0$. (The reader should check that " $\sim$ " is an equivalence relation.)

Definition 3.3. *A* ***real number*** *is an equivalence class,* $\bar{a} := \{b \in \mathcal{C} : b \sim a\}$ *associated to some element* $a \in \mathcal{C}$*. The collection of real numbers will be denoted by* $\mathbb{R}$*. For* $q \in \mathbb{Q}$*, let* $i(q) = \bar{a}$ *where* a *is the constant sequence* $a_n = q$ *for all* $n \in \mathbb{N}$*. We will simply write* 0 *for* $i(0)$ *and* 1 *for* $i(1)$*.*

Exercise 3.4. Given $\bar{a}, \bar{b} \in \mathbb{R}$ show that the definitions
$$-\bar{a} = \overline{(-a)},\ \bar{a} + \bar{b} := \overline{(a+b)} \text{ and } \bar{a} \cdot \bar{b} := \overline{a \cdot b}$$

[1] This fact also shows that the intermediate value theorem, (See Theorem 10.50 below.) fails when working with continuous functions defined over $\mathbb{Q}$.

are well defined. Here $-a$, $a+b$ and $a \cdot b$ denote the sequences $\{-a_n\}_{n=1}^{\infty}$, $\{a_n + b_n\}_{n=1}^{\infty}$ and $\{a_n \cdot b_n\}_{n=1}^{\infty}$ respectively. Further verify that with these operations, $\mathbb{R}$ becomes a field and the map $i : \mathbb{Q} \to \mathbb{R}$ is injective homomorphism of fields. **Hint:** if $\bar{a} \neq 0$ show that $\bar{a}$ may be represented by a sequence $a \in \mathcal{C}$ with $|a_n| \geq \frac{1}{N}$ for all n and some $N \in \mathbb{N}$. For this representative show the sequence $a^{-1} := \left\{a_n^{-1}\right\}_{n=1}^{\infty} \in \mathcal{C}$. The multiplicative inverse to $\bar{a}$ may now be constructed as: $\frac{1}{\bar{a}} = \bar{a}^{-1} := \overline{\left\{a_n^{-1}\right\}_{n=1}^{\infty}}$.

Definition 3.4. *Let $\bar{a}, \bar{b} \in \mathbb{R}$. Then*

1. $\bar{a} > 0$ *if there exists an* $N \in \mathbb{N}$ *such that* $a_n > \frac{1}{N}$ *for a.a.* n.
2. $\bar{a} \geq 0$ *iff either* $\bar{a} > 0$ *or* $\bar{a} = 0$. *Equivalently (as the reader should verify),* $\bar{a} \geq 0$ *iff for all* $N \in \mathbb{N}$, $a_n \geq -\frac{1}{N}$ *for a.a.* n.
3. *Write* $\bar{a} > \bar{b}$ *or* $\bar{b} < \bar{a}$ *if* $\bar{a} - \bar{b} > 0$
4. *Write* $\bar{a} \geq \bar{b}$ *or* $\bar{b} \leq \bar{a}$ *if* $\bar{a} - \bar{b} \geq 0$.

Exercise 3.5. Show " $\geq$ " make $\mathbb{R}$ into a linearly ordered field and the map $i : \mathbb{Q} \to \mathbb{R}$ preserves order. Namely if $\bar{a}, \bar{b} \in \mathbb{R}$ then

1. exactly one of the following relations hold: $\bar{a} < \bar{b}$ or $\bar{a} > \bar{b}$ or $\bar{a} = \bar{b}$.
2. If $\bar{a} \geq 0$ and $\bar{b} \geq 0$ then $\bar{a} + \bar{b} \geq 0$ and $\bar{a} \cdot \bar{b} \geq 0$.
3. If $q, r \in Q$ then $q \leq r$ iff $i(q) \leq i(r)$.

The **absolute value** of a real number $\bar{a}$ is defined analogously to that of a rational number by

$$|\bar{a}| = \begin{cases} \bar{a} & \text{if } \bar{a} \geq 0 \\ -\bar{a} & \text{if } \bar{a} < 0 \end{cases}.$$

Observe this definition is consistent with our previous definition of the absolute value on $\mathbb{Q}$, namely $i(|q|) = |i(q)|$. Also notice that $\bar{a} = 0$ (i.e. $a \sim 0$ where 0 denotes the constant sequence of all zeros) iff for all $N \in \mathbb{N}$, $|a_n| \leq \frac{1}{N}$ for a.a. n. This is equivalent to saying $|\bar{a}| \leq i\left(\frac{1}{N}\right)$ for all $N \in \mathbb{N}$ iff $\bar{a} = 0$.

Exercise 3.6. Given $\bar{a}, \bar{b} \in \mathbb{R}$ show

$$\left|\bar{a}\bar{b}\right| = |\bar{a}| \left|\bar{b}\right| \text{ and } \left|\bar{a} + \bar{b}\right| \leq |\bar{a}| + \left|\bar{b}\right|.$$

The latter inequality being referred to as the **triangle inequality**.

By exercise 3.6,

$$|\bar{a}| = \left|\bar{a} - \bar{b} + \bar{b}\right| \leq \left|\bar{a} - \bar{b}\right| + \left|\bar{b}\right|$$

and hence

$$|\bar{a}| - \left|\bar{b}\right| \leq \left|\bar{a} - \bar{b}\right|$$

and by reversing the roles of $\bar{a}$ and $\bar{b}$ we also have

$$-\left(|\bar{a}| - \left|\bar{b}\right|\right) = \left|\bar{b}\right| - |\bar{a}| \leq \left|\bar{b} - \bar{a}\right| = \left|\bar{a} - \bar{b}\right|.$$

Therefore $\left||\bar{a}|-|\bar{b}|\right| \leq \left|\bar{a}-\bar{b}\right|$ and in particular if $\{\bar{a}_n\}_{n=1}^{\infty} \subset \mathbb{R}$ **converges** to $\bar{a} \in \mathbb{R}$ then

$$||\bar{a}_n| - |\bar{a}|| \leq |\bar{a}_n - \bar{a}| \to 0 \text{ as } n \to \infty.$$

Definition 3.5. *A sequence $\{\bar{a}_n\}_{n=1}^{\infty} \subset \mathbb{R}$ **converges** to $\bar{a} \in \mathbb{R}$ if $|\bar{a} - \bar{a}_n| \to 0$ as $n \to \infty$, i.e. if for all $N \in \mathbb{N}$, $|\bar{a} - \bar{a}_n| \leq i\left(\frac{1}{N}\right)$ for a.a. n. As before if $\{\bar{a}_n\}_{n=1}^{\infty}$ converges to $\bar{a}$ we will write $\bar{a}_n \to \bar{a}$ as $n \to \infty$ or $\bar{a} = \lim_{n\to\infty} \bar{a}_n$.*

Remark 3.6. The field $i(\mathbb{Q})$ is **dense** in $\mathbb{R}$ in the sense that if $\bar{a} \in \mathbb{R}$ there exists $\{q_n\}_{n=1}^{\infty} \subset \mathbb{Q}$ such that $i(q_n) \to \bar{a}$ as $n \to \infty$. Indeed, simply let $q_n = a_n$ where a represents $\bar{a}$. Since a is a Cauchy sequence, to any $N \in \mathbb{N}$ there exits $M \in \mathbb{N}$ such that

$$-\frac{1}{N} \leq a_m - a_n \leq \frac{1}{N} \text{ for all } m, n \geq M$$

and therefore

$$-i\left(\frac{1}{N}\right) \leq i(a_m) - \bar{a} \leq i\left(\frac{1}{N}\right) \text{ for all } m \geq M.$$

This shows

$$|i(q_m) - \bar{a}| = |i(a_m) - \bar{a}| \leq i\left(\frac{1}{N}\right) \text{ for all } m \geq M$$

and since N is arbitrary that $i(q_m) \to \bar{a}$ as $m \to \infty$.

Definition 3.7. *A sequence $\{\bar{a}_n\}_{n=1}^{\infty} \subset \mathbb{R}$ is **Cauchy** if $|\bar{a}_n - \bar{a}_m| \to 0$ as $m, n \to \infty$. More precisely we require for each $N \in \mathbb{N}$ that $|\bar{a}_m - \bar{a}_n| \leq i\left(\frac{1}{N}\right)$ for a.a. pairs (m, n).*

Exercise 3.7. The analogues of the results in Exercises 3.1 and 3.2 hold with $\mathbb{Q}$ replaced by $\mathbb{R}$. (We now say a subset $\Lambda \subset \mathbb{R}$ is bounded if there exists $M \in \mathbb{N}$ such that $|\lambda| \leq i(M)$ for all $\lambda \in \Lambda$.)

For the purposes of real analysis the most important property of $\mathbb{R}$ is that it is "complete."

Theorem 3.8. *The ordered field $\mathbb{R}$ is **complete**, i.e. all Cauchy sequences in $\mathbb{R}$ are convergent.*

Proof. Suppose that $\{\bar{a}(m)\}_{m=1}^{\infty}$ is a Cauchy sequence in $\mathbb{R}$. By Remark 3.6, we may choose $q_m \in \mathbb{Q}$ such that

$$|\bar{a}(m) - i(q_m)| \leq i\left(m^{-1}\right) \text{ for all } m \in \mathbb{N}.$$

Given $N \in \mathbb{N}$, choose $M \in \mathbb{N}$ such that $|\bar{a}(m) - \bar{a}(n)| \leq i\left(N^{-1}\right)$ for all $m, n \geq M$. Then

$$\begin{aligned}|i(q_m) - i(q_n)| &\le |i(q_m) - \bar{a}(m)| + |\bar{a}(m) - \bar{a}(n)| + |\bar{a}(n) - i(q_n)| \\ &\le i\left(m^{-1}\right) + i\left(n^{-1}\right) + i\left(N^{-1}\right)\end{aligned}$$

and therefore

$$|q_m - q_n| \le m^{-1} + n^{-1} + N^{-1} \text{ for all } m, n \ge M.$$

It now follows that $q = \{q_m\}_{m=1}^\infty \in \mathcal{C}$ and therefore q represents a point $\bar{q} \in \mathbb{R}$. Using Remark 3.6 and the triangle inequality,

$$\begin{aligned}|\bar{a}(m) - \bar{q}| &\le |\bar{a}(m) - i(q_m)| + |i(q_m) - \bar{q}| \\ &\le i\left(m^{-1}\right) + |i(q_m) - \bar{q}| \to 0 \text{ as } m \to \infty\end{aligned}$$

and therefore $\lim_{m\to\infty} \bar{a}(m) = \bar{q}$. ∎

Definition 3.9. *A number $M \in \mathbb{R}$ is an **upper bound** for a set $\Lambda \subset \mathbb{R}$ if $\lambda \le M$ for all $\lambda \in \Lambda$ and a number $m \in \mathbb{R}$ is an **lower bound** for a set $\Lambda \subset \mathbb{R}$ if $\lambda \ge m$ for all $\lambda \in \Lambda$. Upper and lower bounds need not exist. If Λ has upper (lower) bound, Λ is said to be **bounded from above (below)**.*

Theorem 3.10. *To each non-empty set $\Lambda \subset \mathbb{R}$ which is bounded from above (below) there is a unique **least upper bound** denoted by* $\sup \Lambda \in \mathbb{R}$ *(respectively **greatest lower bound** denoted by* $\inf \Lambda \in \mathbb{R}$).

Proof. Suppose Λ is bounded from above and for each $n \in \mathbb{N}$, let $m_n \in \mathbb{Z}$ be the smallest integer such that $i\left(\frac{m_n}{2^n}\right)$ is an upper bound for Λ. The sequence $q_n := \frac{m_n}{2^n}$ is Cauchy because $q_m \in [q_n - 2^{-n}, q_n] \cap \mathbb{Q}$ for all $m \ge n$, i.e.

$$|q_m - q_n| \le 2^{-\min(m,n)} \to 0 \text{ as } m, n \to \infty.$$

Passing to the limit, $n \to \infty$, in the inequality $i(q_n) \ge \lambda$, which is valid for all $\lambda \in \Lambda$ implies

$$\bar{q} = \lim_{n\to\infty} i(q_n) \ge \lambda \text{ for all } \lambda \in \Lambda.$$

Thus $\bar{q}$ is an upper bound for Λ. If there were another upper bound $M \in \mathbb{R}$ for Λ such that $M < \bar{q}$, it would follow that $M \le i(q_n) < \bar{q}$ for some n. But this is a contradiction because $\{q_n\}_{n=1}^\infty$ is a decreasing sequence, $i(q_n) \ge i(q_m)$ for all $m \ge n$ and therefore $i(q_n) \ge \bar{q}$ for all n. Therefore $\bar{q}$ is the unique least upper bound for Λ. The existence of lower bounds is proved analogously. ∎

Proposition 3.11. *If $\{a_n\}_{n=1}^\infty \subset \mathbb{R}$ is an increasing (decreasing) sequence which is bounded from above (below), then $\{a_n\}_{n=1}^\infty$ is convergent and*

$$\lim_{n\to\infty} a_n = \sup\{a_n : n \in \mathbb{N}\} \ \left(\lim_{n\to\infty} a_n = \inf\{a_n : n \in \mathbb{N}\}\right).$$

If $\Lambda \subset \mathbb{R}$ is a set bounded from above then there exists $\{\lambda_n\} \subset \Lambda$ such that $\lambda_n \uparrow M := \sup \Lambda$, as $n \to \infty$, i.e. $\{\lambda_n\}$ is increasing and $\lim_{n\to\infty} \lambda_n = M$.

Proof. Let $M := \sup \{a_n : n \in \mathbb{N}\}$, then for each $N \in \mathbb{N}$ there must exist $m \in \mathbb{N}$ such that $M - i\left(N^{-1}\right) < a_m \leq M$. Since a_n is increasing, it follows that

$$M - i\left(N^{-1}\right) < a_n \leq M \text{ for all } n \geq m.$$

From this we conclude that $\lim a_n$ exists and $\lim a_n = M$. If $M = \sup \Lambda$, for each $n \in \mathbb{N}$ we may choose $\lambda_n \in \Lambda$ such that

$$M - i\left(n^{-1}\right) < \lambda_n \leq M. \tag{3.2}$$

By replacing λ_n by $\max\{\lambda_1, \ldots, \lambda_n\}$[2] if necessary we may assume that λ_n is increasing in n. It now follows easily from Eq. (3.2) that $\lim_{n\to\infty} \lambda_n = M$. ∎

3.1.1 The Decimal Representation of a Real Number

Let $\alpha \in \mathbb{R}$ or $\alpha \in \mathbb{Q}$, $m, n \in \mathbb{Z}$ and $S := \sum_{k=n}^{m} \alpha^k$. If $\alpha = 1$ then $\sum_{k=n}^{m} \alpha^k = m - n + 1$ while for $\alpha \neq 1$,

$$\alpha S - S = \alpha^{m+1} - \alpha^n$$

and solving for S gives the important geometric summation formula,

$$\sum_{k=n}^{m} \alpha^k = \frac{\alpha^{m+1} - \alpha^n}{\alpha - 1} \text{ if } \alpha \neq 1. \tag{3.3}$$

Taking $\alpha = 10^{-1}$ in Eq. (3.3) implies

$$\sum_{k=n}^{m} 10^{-k} = \frac{10^{-(m+1)} - 10^{-n}}{10^{-1} - 1} = \frac{1}{10^{n-1}} \frac{1 - 10^{-(m-n)}}{9}$$

and in particular, for all $M \geq n$,

$$\lim_{m\to\infty} \sum_{k=n}^{m} 10^{-k} = \frac{1}{9 \cdot 10^{n-1}} \geq \sum_{k=n}^{M} 10^{-k}.$$

Let $\mathbb{D}$ denote those sequences $\alpha \in \{0, 1, 2, \ldots, 9\}^{\mathbb{Z}}$ with the following properties:

1. there exists $N \in \mathbb{N}$ such that $\alpha_{-n} = 0$ for all $n \geq N$ and
2. $\alpha_n \neq 0$ for some $n \in \mathbb{Z}$.

[2] The notation, $\max \Lambda$, denotes $\sup \Lambda$ along with the assertion that $\sup \Lambda \in \Lambda$. Similarly, $\min \Lambda = \inf \Lambda$ along with the assertion that $\inf \Lambda \in \Lambda$.

Associated to each $\alpha \in \mathbb{D}$ is the sequence $a = a(\alpha)$ defined by

$$a_n := \sum_{k=-\infty}^{n} \alpha_k 10^{-k}.$$

Since for $m > n$,

$$|a_m - a_n| = \left| \sum_{k=n+1}^{m} \alpha_k 10^{-k} \right| \le 9 \sum_{k=n+1}^{m} 10^{-k} \le 9 \frac{1}{9 \cdot 10^n} = \frac{1}{10^n},$$

it follows that

$$|a_m - a_n| \le \frac{1}{10^{\min(m,n)}} \to 0 \text{ as } m, n \to \infty.$$

Therefore $a = \underline{a(\alpha)} \in \mathcal{C}$ and we may define a map $D : \{\pm 1\} \times \mathbb{D} \to \mathbb{R}$ defined by $D(\varepsilon, \alpha) = \varepsilon \overline{a(\alpha)}$. As is customary we will denote $D(\varepsilon, \alpha) = \varepsilon \overline{a(\alpha)}$ as

$$\varepsilon \cdot \alpha_m \dots \alpha_0 . \alpha_1 \alpha_2 \dots \alpha_n \dots \tag{3.4}$$

where m is the largest integer in $\mathbb{Z}$ such that $\alpha_k = 0$ for all $k < m$. If $m > 0$ the expression in Eq. (3.4) should be interpreted as

$$\varepsilon \cdot 0.0 \dots 0 \alpha_m \alpha_{m+1} \dots .$$

An element $\alpha \in \mathbb{D}$ has a tail of all 9's starting at $N \in \mathbb{N}$ if $\alpha_n = 9$ and for all $n \ge N$ and $\alpha_{N-1} \ne 9$. If α has a tail of 9's starting at $N \in \mathbb{N}$, then for $n > N$,

$$\begin{aligned} a_n(\alpha) &= \sum_{k=-\infty}^{N-1} \alpha_k 10^{-k} + 9 \sum_{k=N}^{n} 10^{-k} \\ &= \sum_{k=-\infty}^{N-1} \alpha_k 10^{-k} + \frac{9}{10^{N-1}} \cdot \frac{1 - 10^{-(n-N)}}{9} \\ &\to \sum_{k=-\infty}^{N-1} \alpha_k 10^{-k} + 10^{-(N-1)} \text{ as } n \to \infty. \end{aligned}$$

If α' is the digits in the decimal expansion of $\sum_{k=-\infty}^{N-1} \alpha_k 10^{-k} + 10^{-(N-1)}$, then

$$\alpha' \in \mathbb{D}' := \{\alpha \in \mathbb{D} : \alpha \text{ does not have a tail of all 9's}\}.$$

and we have just shown that $D(\varepsilon, \alpha) = D(\varepsilon, \alpha')$. In particular this implies

$$D(\{\pm 1\} \times \mathbb{D}') = D(\{\pm 1\} \times \mathbb{D}). \tag{3.5}$$

Theorem 3.12 (Decimal Representation). *The map*

$$D : \{\pm 1\} \times \mathbb{D}' \to \mathbb{R} \setminus \{0\}$$

is a bijection.

Proof. Suppose $D(\varepsilon, \alpha) = D(\delta, \beta)$ for some (ε, α) and (δ, β) in $\{\pm 1\} \times \mathbb{D}$. Since $D(\varepsilon, \alpha) > 0$ if $\varepsilon = 1$ and $D(\varepsilon, \alpha) < 0$ if $\varepsilon = -1$ it follows that $\varepsilon = \delta$. Let $a = a(\alpha)$ and $b = a(\beta)$ be the sequences associated to α and β respectively. Suppose that $\alpha \neq \beta$ and let $j \in \mathbb{Z}$ be the position where α and β first disagree, i.e. $\alpha_n = \beta_n$ for all $n < j$ while $\alpha_j \neq \beta_j$. For sake of definiteness suppose $\beta_j > \alpha_j$. Then for $n > j$ we have

$$\begin{aligned} b_n - a_n &= (\beta_j - \alpha_j) 10^{-j} + \sum_{k=j+1}^{n} (\beta_k - \alpha_k) 10^{-k} \\ &\geq 10^{-j} - 9 \sum_{k=j+1}^{n} 10^{-k} \geq 10^{-j} - 9 \frac{1}{9 \cdot 10^j} = 0. \end{aligned}$$

Therefore $b_n - a_n \geq 0$ for all n and $\lim (b_n - a_n) = 0$ iff $\beta_j = \alpha_j + 1$ and $\beta_k = 9$ and $\alpha_k = 0$ for all $k > j$. In summary, $D(\varepsilon, \alpha) = D(\delta, \beta)$ with $\alpha \neq \beta$ implies either α or β has an infinite tail of nines which shows that D is injective when restricted to $\{\pm 1\} \times \mathbb{D}'$. To see that D is surjective it suffices to show any $\bar{b} \in \mathbb{R}$ with $0 < \bar{b} < 1$ is in the range of D. For each $n \in \mathbb{N}$, let $a_n = .\alpha_1 \dots \alpha_n$ with $\alpha_i \in \{0, 1, 2, \dots, 9\}$ such that

$$i(a_n) < \bar{b} \leq i(a_n) + i\left(10^{-n}\right). \tag{3.6}$$

Since $a_{n+1} = a_n + \alpha_{n+1} 10^{-(n+1)}$ for some $\alpha_{n+1} \in \{0, 1, 2, \dots, 9\}$, we see that $a_{n+1} = .\alpha_1 \dots \alpha_n \alpha_{n+1}$, i.e. the first n digits in the decimal expansion of a_{n+1} are the same as in the decimal expansion of a_n. Hence this defines α_n uniquely for all $n \geq 1$. By setting $\alpha_n = 0$ when $n \leq 0$, we have constructed from $\bar{b}$ an element $\alpha \in \mathbb{D}$. Because of Eq. (3.6), $D(1, \alpha) = \bar{b}$. ∎

Notation 3.13 *From now on we will identify $\mathbb{Q}$ with $i(\mathbb{Q}) \subset \mathbb{R}$ and elements in $\mathbb{R}$ with their decimal expansions.*

To summarize, we have constructed a complete ordered field $\mathbb{R}$ "containing" $\mathbb{Q}$ as a dense subset. Moreover every element in $\mathbb{R}$ (modulo those of the form $m10^{-n}$ for some $m \in \mathbb{Z}$ and $n \in \mathbb{N}$) has a unique decimal expansion.

Corollary 3.14. *The set $(0, 1) := \{a \in \mathbb{R} : 0 < a < 1\}$ is uncountable while $\mathbb{Q} \cap (0, 1)$ is countable.*

Proof. By Theorem 3.12, the set $\{0, 1, 2 \dots, 8\}^{\mathbb{N}}$ can be mapped injectively into $(0, 1)$ and therefore it follows from Lemma 2.6 that $(0, 1)$ is uncountable. For each $m \in \mathbb{N}$, let $A_m := \left\{\frac{n}{m} : n \in \mathbb{N} \text{ with } n < m\right\}$. Since $\mathbb{Q} \cap (0, 1) = \cup_{m=1}^{\infty} A_m$ and $\#(A_m) < \infty$ for all m, another application of Lemma 2.6 shows $\mathbb{Q} \cap (0, 1)$ is countable. ∎

3.2 The Complex Numbers

Definition 3.15 (Complex Numbers). *Let $\mathbb{C} = \mathbb{R}^2$ equipped with multiplication rule*

$$(a,b)(c,d) := (ac - bd, bc + ad) \tag{3.7}$$

and the usual rule for vector addition. As is standard we will write $0 = (0,0)$, $1 = (1,0)$ and $i = (0,1)$ so that every element z of $\mathbb{C}$ may be written as $z = x1 + yi$ which in the future will be written simply as $z = x + iy$. If $z = x + iy$, let $\operatorname{Re} z = x$ and $\operatorname{Im} z = y$.

Writing $z = a + ib$ and $w = c + id$, the multiplication rule in Eq. (3.7) becomes

$$(a + ib)(c + id) := (ac - bd) + i(bc + ad) \tag{3.8}$$

and in particular $1^2 = 1$ and $i^2 = -1$.

Proposition 3.16. *The complex numbers $\mathbb{C}$ with the above multiplication rule satisfies the usual definitions of a field. For example $wz = zw$ and $z(w_1 + w_2) = zw_1 + zw_2$, etc. Moreover if $z \neq 0$, z has a multiplicative inverse given by*

$$z^{-1} = \frac{a}{a^2 + b^2} - i\frac{b}{a^2 + b^2}. \tag{3.9}$$

Proof. The proof is a straightforward verification. Only the last assertion will be verified here. Suppose $z = a + ib \neq 0$, we wish to find $w = c + id$ such that $zw = 1$ and this happens by Eq. (3.8) iff

$$ac - bd = 1 \text{ and} \tag{3.10}$$

$$bc + ad = 0. \tag{3.11}$$

Solving these equations for c and d gives $c = \frac{a}{a^2+b^2}$ and $d = -\frac{b}{a^2+b^2}$ as claimed. ■

Notation 3.17 (Conjugation and Modulus) *If $z = a + ib$ with $a, b \in \mathbb{R}$ let $\bar{z} = a - ib$ and*

$$|z| := \sqrt{z\bar{z}} = \sqrt{a^2 + b^2} = \sqrt{|\operatorname{Re} z|^2 + |\operatorname{Im} z|^2}.$$

See Exercise 3.8 for the existence of the square root as a positive real number.

Notice that

$$\operatorname{Re} z = \frac{1}{2}(z + \bar{z}) \text{ and } \operatorname{Im} z = \frac{1}{2i}(z - \bar{z}). \tag{3.12}$$

Proposition 3.18. *Complex conjugation and the modulus operators satisfy the following properties.*

1. $\bar{\bar{z}} = z$,

2. $\overline{zw} = \bar{z}\bar{w}$ *and* $\bar{z} + \bar{w} = \overline{z + w}$.
3. $|\bar{z}| = |z|$
4. $|zw| = |z|\,|w|$ *and in particular* $|z^n| = |z|^n$ *for all* $n \in \mathbb{N}$.
5. $|\mathrm{Re}\, z| \leq |z|$ *and* $|\mathrm{Im}\, z| \leq |z|$
6. $|z + w| \leq |z| + |w|$.
7. $z = 0$ *iff* $|z| = 0$.
8. *If* $z \neq 0$ *then* $z^{-1} := \frac{\bar{z}}{|z|^2}$ *(also written as* $\frac{1}{z}$*) is the inverse of* z.
9. $\left|z^{-1}\right| = |z|^{-1}$ *and more generally* $|z^n| = |z|^n$ *for all* $n \in \mathbb{Z}$.

Proof. All of these properties are direct computations except for possibly the triangle inequality in item 6 which is verified by the following computation;

$$\begin{aligned} |z+w|^2 &= (z+w)\left(\overline{z+w}\right) = |z|^2 + |w|^2 + w\bar{z} + \bar{w}z \\ &= |z|^2 + |w|^2 + w\bar{z} + \overline{w\bar{z}} \\ &= |z|^2 + |w|^2 + 2\,\mathrm{Re}\,(w\bar{z}) \leq |z|^2 + |w|^2 + 2\,|z|\,|w| \\ &= (|z| + |w|)^2 . \end{aligned}$$

■

Definition 3.19. *A sequence* $\{z_n\}_{n=1}^{\infty} \subset \mathbb{C}$ *is* ***Cauchy*** *if* $|z_n - z_m| \to 0$ *as* $m, n \to \infty$ *and is* ***convergent*** *to* $z \in \mathbb{C}$ *if* $|z - z_n| \to 0$ *as* $n \to \infty$. *As usual if* $\{z_n\}_{n=1}^{\infty}$ *converges to* z *we will write* $z_n \to z$ *as* $n \to \infty$ *or* $z = \lim_{n\to\infty} z_n$.

Theorem 3.20. *The complex numbers are complete,i.e. all Cauchy sequences are convergent.*

Proof. This follows from the completeness of real numbers and the easily proved observations that if $z_n = a_n + ib_n \in \mathbb{C}$, then

1. $\{z_n\}_{n=1}^{\infty} \subset \mathbb{C}$ is Cauchy iff $\{a_n\}_{n=1}^{\infty} \subset \mathbb{R}$ and $\{b_n\}_{n=1}^{\infty} \subset \mathbb{R}$ are Cauchy and
2. $z_n \to z = a + ib$ as $n \to \infty$ iff $a_n \to a$ and $b_n \to b$ as $n \to \infty$.

■

3.3 Exercises

Exercise 3.8. Show to every $a \in \mathbb{R}$ with $a \geq 0$ there exists a unique number $b \in \mathbb{R}$ such that $b \geq 0$ and $b^2 = a$. Of course we will call $b = \sqrt{a}$. Also show that $a \to \sqrt{a}$ is an increasing function on $[0, \infty)$. **Hint:** To construct $b = \sqrt{a}$ for $a > 0$, to each $n \in \mathbb{N}$ let $m_n \in \mathbb{N}_0$ be chosen so that

$$\frac{m_n^2}{n^2} < a \leq \frac{(m_n+1)^2}{n^2} \text{ i.e. } i\left(\frac{m_n^2}{n^2}\right) < a \leq i\left(\frac{(m_n+1)^2}{n^2}\right)$$

and let $q_n := \frac{m_n}{n}$. Then show $b = \overline{\{q_n\}_{n=1}^{\infty}} \in \mathbb{R}$ satisfies $b > 0$ and $b^2 = a$.

4

Limits and Sums

4.1 Limsups, Liminfs and Extended Limits

Notation 4.1 *The* ***extended real numbers*** *is the set* $\bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$, *i.e. it is* $\mathbb{R}$ *with two new points called* ∞ *and* $-\infty$. *We use the following conventions,* $\pm\infty \cdot 0 = 0$, $\pm\infty + a = \pm\infty$ *for any* $a \in \mathbb{R}$, $\infty + \infty = \infty$ *and* $-\infty - \infty = -\infty$ *while* $\infty - \infty$ *is not defined. A sequence* $a_n \in \bar{\mathbb{R}}$ *is said to converge to* ∞ $(-\infty)$ *if for all* $M \in \mathbb{R}$ *there exists* $m \in \mathbb{N}$ *such that* $a_n \geq M$ $(a_n \leq M)$ *for all* $n \geq m$.

Lemma 4.2. *Suppose* $\{a_n\}_{n=1}^{\infty}$ *and* $\{b_n\}_{n=1}^{\infty}$ *are convergent sequences in* $\bar{\mathbb{R}}$, *then:*

1. *If* $a_n \leq b_n$ *for a.a.* n *then* $\lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n$.
2. *If* $c \in \mathbb{R}$, $\lim_{n\to\infty} (ca_n) = c \lim_{n\to\infty} a_n$.
3. *If* $\{a_n + b_n\}_{n=1}^{\infty}$ *is convergent and*
$$\lim_{n\to\infty} (a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n \tag{4.1}$$
provided the right side is not of the form $\infty - \infty$.
4. $\{a_n b_n\}_{n=1}^{\infty}$ *is convergent and*
$$\lim_{n\to\infty} (a_n b_n) = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n \tag{4.2}$$
provided the right hand side is not of the for $\infty \cdot 0$.

Before going to the proof consider the simple example where $a_n = n$ and $b_n = -\alpha n$ with $\alpha > 0$. Then

$$\lim (a_n + b_n) = \begin{cases} \infty & \text{if } \alpha < 1 \\ 0 & \text{if } \alpha = 1 \\ -\infty & \text{if } \alpha > 1 \end{cases}$$

while

$$\lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n \text{“} = \text{”} \infty - \infty.$$

This shows that the requirement that the right side of Eq. (4.1) is not of form $\infty - \infty$ is necessary in Lemma 4.2. Similarly by considering the examples $a_n = n$ and $b_n = n^{-\alpha}$ with $\alpha > 0$ shows the necessity for assuming right hand side of Eq. (4.2) is not of the form $\infty \cdot 0$.

Proof. The proofs of items 1. and 2. are left to the reader.

Proof of Eq. (4.1). Let $a := \lim_{n\to\infty} a_n$ and $b = \lim_{n\to\infty} b_n$. Case 1., suppose $b = \infty$ in which case we must assume $a > -\infty$. In this case, for every $M > 0$, there exists N such that $b_n \geq M$ and $a_n \geq a - 1$ for all $n \geq N$ and this implies

$$a_n + b_n \geq M + a - 1 \text{ for all } n \geq N.$$

Since M is arbitrary it follows that $a_n + b_n \to \infty$ as $n \to b = \infty$. The cases where $b = -\infty$ or $a = \pm\infty$ are handled similarly. Case 2. If $a, b \in \mathbb{R}$, then for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$|a - a_n| \leq \varepsilon \text{ and } |b - b_n| \leq \varepsilon \text{ for all } n \geq N.$$

Therefore,

$$|a + b - (a_n + b_n)| = |a - a_n + b - b_n| \leq |a - a_n| + |b - b_n| \leq 2\varepsilon$$

for all $n \geq N$. Since n is arbitrary, it follows that $\lim_{n\to\infty} (a_n + b_n) = a + b$.

Proof of Eq. (4.2). It will be left to the reader to prove the case where $\lim a_n$ and $\lim b_n$ exist in $\mathbb{R}$. I will only consider the case where $a = \lim_{n\to\infty} a_n \neq 0$ and $\lim_{n\to\infty} b_n = \infty$ here. Let us also suppose that $a > 0$ (the case $a < 0$ is handled similarly) and let $\alpha := \min\left(\frac{a}{2}, 1\right)$. Given any $M < \infty$, there exists $N \in \mathbb{N}$ such that $a_n \geq \alpha$ and $b_n \geq M$ for all $n \geq N$ and for this choice of N, $a_n b_n \geq M\alpha$ for all $n \geq N$. Since $\alpha > 0$ is fixed and M is arbitrary it follows that $\lim_{n\to\infty} (a_n b_n) = \infty$ as desired. ∎

For any subset $\Lambda \subset \bar{\mathbb{R}}$, let $\sup \Lambda$ and $\inf \Lambda$ denote the least upper bound and greatest lower bound of Λ respectively. The convention being that $\sup \Lambda = \infty$ if $\infty \in \Lambda$ or Λ is not bounded from above and $\inf \Lambda = -\infty$ if $-\infty \in \Lambda$ or Λ is not bounded from below. We will also use the **conventions** that $\sup \emptyset = -\infty$ and $\inf \emptyset = +\infty$.

Notation 4.3 *Suppose that* $\{x_n\}_{n=1}^\infty \subset \bar{\mathbb{R}}$ *is a sequence of numbers. Then*

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \inf\{x_k : k \geq n\} \text{ and} \tag{4.3}$$

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup\{x_k : k \geq n\}. \tag{4.4}$$

We will also write $\underline{\lim}$ *for* $\liminf$ *and* $\overline{\lim}$ *for* $\limsup$.

Remark 4.4. Notice that if $a_k := \inf\{x_k : k \geq n\}$ and $b_k := \sup\{x_k : k \geq n\}$, then $\{a_k\}$ is an increasing sequence while $\{b_k\}$ is a decreasing sequence. Therefore the limits in Eq. (4.3) and Eq. (4.4) always exist in $\bar{\mathbb{R}}$ and

$$\liminf_{n\to\infty} x_n = \sup_n \inf\{x_k : k \geq n\} \text{ and}$$
$$\limsup_{n\to\infty} x_n = \inf_n \sup\{x_k : k \geq n\}.$$

The following proposition contains some basic properties of liminfs and limsups.

Proposition 4.5. *Let $\{a_n\}_{n=1}^\infty$ and $\{b_n\}_{n=1}^\infty$ be two sequences of real numbers. Then*

1. $\liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n$ *and* $\lim_{n\to\infty} a_n$ *exists in* $\bar{\mathbb{R}}$ *iff*
$$\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n \in \bar{\mathbb{R}}.$$
2. *There is a subsequence* $\{a_{n_k}\}_{k=1}^\infty$ *of* $\{a_n\}_{n=1}^\infty$ *such that* $\lim_{k\to\infty} a_{n_k} = \limsup_{n\to\infty} a_n$.
3. $$\limsup_{n\to\infty} (a_n + b_n) \leq \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \tag{4.5}$$
whenever the right side of this equation is not of the form $\infty - \infty$.
4. *If* $a_n \geq 0$ *and* $b_n \geq 0$ *for all* $n \in \mathbb{N}$, *then*
$$\limsup_{n\to\infty} (a_n b_n) \leq \limsup_{n\to\infty} a_n \cdot \limsup_{n\to\infty} b_n, \tag{4.6}$$
provided the right hand side of (4.6) is not of the form $0 \cdot \infty$ *or* $\infty \cdot 0$.

Proof. Item 1. will be proved here leaving the remaining items as an exercise to the reader. Since

$$\inf\{a_k : k \geq n\} \leq \sup\{a_k : k \geq n\} \;\forall\, n,$$
$$\liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n.$$

Now suppose that $\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n = a \in \mathbb{R}$. Then for all $\varepsilon > 0$, there is an integer N such that

$$a - \varepsilon \leq \inf\{a_k : k \geq N\} \leq \sup\{a_k : k \geq N\} \leq a + \varepsilon,$$

i.e.

$$a - \varepsilon \leq a_k \leq a + \varepsilon \text{ for all } k \geq N.$$

Hence by the definition of the limit, $\lim_{k\to\infty} a_k = a$. If $\liminf_{n\to\infty} a_n = \infty$, then we know for all $M \in (0, \infty)$ there is an integer N such that

$$M \leq \inf\{a_k : k \geq N\}$$

and hence $\lim_{n\to\infty} a_n = \infty$. The case where $\limsup_{n\to\infty} a_n = -\infty$ is handled similarly.

Conversely, suppose that $\lim_{n\to\infty} a_n = A \in \bar{\mathbb{R}}$ exists. If $A \in \mathbb{R}$, then for every $\varepsilon > 0$ there exists $N(\varepsilon) \in \mathbb{N}$ such that $|A - a_n| \le \varepsilon$ for all $n \ge N(\varepsilon)$, i.e.

$$A - \varepsilon \le a_n \le A + \varepsilon \text{ for all } n \ge N(\varepsilon).$$

From this we learn that

$$A - \varepsilon \le \liminf_{n\to\infty} a_n \le \limsup_{n\to\infty} a_n \le A + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, it follows that

$$A \le \liminf_{n\to\infty} a_n \le \limsup_{n\to\infty} a_n \le A,$$

i.e. that $A = \liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n$. If $A = \infty$, then for all $M > 0$ there exists $N(M)$ such that $a_n \ge M$ for all $n \ge N(M)$. This show that $\liminf_{n\to\infty} a_n \ge M$ and since M is arbitrary it follows that

$$\infty \le \liminf_{n\to\infty} a_n \le \limsup_{n\to\infty} a_n.$$

The proof for the case $A = -\infty$ is analogous to the $A = \infty$ case. ∎

4.2 Sums of positive functions

In this and the next few sections, let X and Y be two sets. We will write $\alpha \subset\subset X$ to denote that α is a **finite** subset of X and write 2^X_f for those $\alpha \subset\subset X$.

Definition 4.6. *Suppose that $a : X \to [0,\infty]$ is a function and $F \subset X$ is a subset, then*

$$\sum_F a = \sum_{x\in F} a(x) := \sup\left\{\sum_{x\in\alpha} a(x) : \alpha \subset\subset F\right\}.$$

Remark 4.7. Suppose that $X = \mathbb{N} = \{1,2,3,\dots\}$ and $a : X \to [0,\infty]$, then

$$\sum_{\mathbb{N}} a = \sum_{n=1}^{\infty} a(n) := \lim_{N\to\infty} \sum_{n=1}^{N} a(n).$$

Indeed for all N, $\sum_{n=1}^N a(n) \le \sum_{\mathbb{N}} a$, and thus passing to the limit we learn that

$$\sum_{n=1}^{\infty} a(n) \le \sum_{\mathbb{N}} a.$$

Conversely, if $\alpha \subset\subset \mathbb{N}$, then for all N large enough so that $\alpha \subset \{1,2,\dots,N\}$, we have $\sum_\alpha a \le \sum_{n=1}^N a(n)$ which upon passing to the limit implies that

$$\sum_\alpha a \le \sum_{n=1}^\infty a(n).$$

Taking the supremum over α in the previous equation shows

$$\sum_{\mathbb{N}} a \le \sum_{n=1}^\infty a(n).$$

Remark 4.8. Suppose $a : X \to [0, \infty]$ and $\sum_X a < \infty$, then $\{x \in X : a(x) > 0\}$ is at most countable. To see this first notice that for any $\varepsilon > 0$, the set $\{x : a(x) \ge \varepsilon\}$ must be finite for otherwise $\sum_X a = \infty$. Thus

$$\{x \in X : a(x) > 0\} = \bigcup\nolimits_{k=1}^\infty \{x : a(x) \ge 1/k\}$$

which shows that $\{x \in X : a(x) > 0\}$ is a countable union of finite sets and thus countable by Lemma 2.6.

Lemma 4.9. *Suppose that $a, b : X \to [0, \infty]$ are two functions, then*

$$\sum_X (a + b) = \sum_X a + \sum_X b \text{ and}$$
$$\sum_X \lambda a = \lambda \sum_X a$$

for all $\lambda \ge 0$.

I will only prove the first assertion, the second being easy. Let $\alpha \subset\subset X$ be a finite set, then

$$\sum_\alpha (a + b) = \sum_\alpha a + \sum_\alpha b \le \sum_X a + \sum_X b$$

which after taking sups over α shows that

$$\sum_X (a + b) \le \sum_X a + \sum_X b.$$

Similarly, if $\alpha, \beta \subset\subset X$, then

$$\sum_\alpha a + \sum_\beta b \le \sum_{\alpha\cup\beta} a + \sum_{\alpha\cup\beta} b = \sum_{\alpha\cup\beta} (a + b) \le \sum_X (a + b).$$

Taking sups over α and β then shows that

$$\sum_X a + \sum_X b \le \sum_X (a + b).$$

Lemma 4.10. *Let X and Y be sets, $R \subset X \times Y$ and suppose that $a : R \to \bar{\mathbb{R}}$ is a function. Let ${}_xR := \{y \in Y : (x,y) \in R\}$ and $R_y := \{x \in X : (x,y) \in R\}$. Then*

$$\sup_{(x,y)\in R} a(x,y) = \sup_{x\in X} \sup_{y\in {}_xR} a(x,y) = \sup_{y\in Y} \sup_{x\in R_y} a(x,y) \text{ and}$$
$$\inf_{(x,y)\in R} a(x,y) = \inf_{x\in X} \inf_{y\in {}_xR} a(x,y) = \inf_{y\in Y} \inf_{x\in R_y} a(x,y).$$

(Recall the conventions: $\sup \emptyset = -\infty$ *and* $\inf \emptyset = +\infty$.*)*

Proof. Let $M = \sup_{(x,y)\in R} a(x,y)$, $N_x := \sup_{y\in {}_xR} a(x,y)$. Then $a(x,y) \le M$ for all $(x,y) \in R$ implies $N_x = \sup_{y\in {}_xR} a(x,y) \le M$ and therefore that

$$\sup_{x\in X} \sup_{y\in {}_xR} a(x,y) = \sup_{x\in X} N_x \le M. \tag{4.7}$$

Similarly for any $(x,y) \in R$,

$$a(x,y) \le N_x \le \sup_{x\in X} N_x = \sup_{x\in X} \sup_{y\in {}_xR} a(x,y)$$

and therefore

$$M = \sup_{(x,y)\in R} a(x,y) \le \sup_{x\in X} \sup_{y\in {}_xR} a(x,y) \tag{4.8}$$

Equations (4.7) and (4.8) show that

$$\sup_{(x,y)\in R} a(x,y) = \sup_{x\in X} \sup_{y\in {}_xR} a(x,y).$$

The assertions involving infimums are proved analogously or follow from what we have just proved applied to the function $-a$. ■

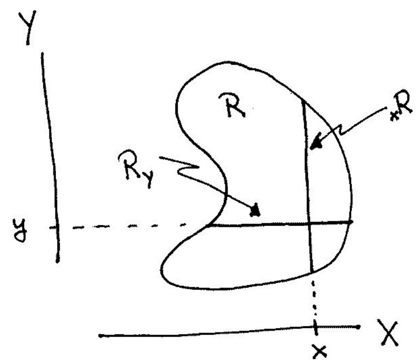


Fig. 4.1. The x and y – slices of a set $R \subset X \times Y$.

Theorem 4.11 (Monotone Convergence Theorem for Sums). *Suppose that $f_n : X \to [0,\infty]$ is an increasing sequence of functions and*

$$f(x) := \lim_{n\to\infty} f_n(x) = \sup_n f_n(x).$$

Then

$$\lim_{n\to\infty} \sum_X f_n = \sum_X f$$

Proof. We will give two proves.
First proof. Let

$$2_f^X := \{A \subset X : A \subset\subset X\}.$$

Then

$$\begin{aligned} \lim_{n\to\infty} \sum_X f_n &= \sup_n \sum_X f_n = \sup_n \sup_{\alpha \in 2_f^X} \sum_\alpha f_n = \sup_{\alpha\in 2_f^X} \sup_n \sum_\alpha f_n \\ &= \sup_{\alpha\in 2_f^X} \lim_{n\to\infty} \sum_\alpha f_n = \sup_{\alpha\in 2_f^X} \sum_\alpha \lim_{n\to\infty} f_n \\ &= \sup_{\alpha\in 2_f^X} \sum_\alpha f = \sum_X f. \end{aligned}$$

Second Proof. Let $S_n = \sum_X f_n$ and $S = \sum_X f$. Since $f_n \le f_m \le f$ for all $n \le m$, it follows that

$$S_n \le S_m \le S$$

which shows that $\lim_{n\to\infty} S_n$ exists and is less that S, i.e.

$$A := \lim_{n\to\infty} \sum_X f_n \le \sum_X f. \tag{4.9}$$

Noting that $\sum_\alpha f_n \le \sum_X f_n = S_n \le A$ for all $\alpha \subset\subset X$ and in particular,

$$\sum_\alpha f_n \le A \text{ for all } n \text{ and } \alpha \subset\subset X.$$

Letting n tend to infinity in this equation shows that

$$\sum_\alpha f \le A \text{ for all } \alpha \subset\subset X$$

and then taking the sup over all $\alpha \subset\subset X$ gives

$$\sum_X f \le A = \lim_{n\to\infty} \sum_X f_n \tag{4.10}$$

which combined with Eq. (4.9) proves the theorem. ∎

Lemma 4.12 (Fatou's Lemma for Sums). *Suppose that $f_n : X \to [0,\infty]$ is a sequence of functions, then*

$$\sum_X \liminf_{n\to\infty} f_n \le \liminf_{n\to\infty} \sum_X f_n.$$

Proof. Define $g_k := \inf_{n\ge k} f_n$ so that $g_k \uparrow \liminf_{n\to\infty} f_n$ as $k \to \infty$. Since $g_k \le f_n$ for all $k \le n$,

$$\sum_X g_k \le \sum_X f_n \text{ for all } n \ge k$$

and therefore

$$\sum_X g_k \le \liminf_{n\to\infty} \sum_X f_n \text{ for all } k.$$

We may now use the monotone convergence theorem to let $k \to \infty$ to find

$$\sum_X \liminf_{n\to\infty} f_n = \sum_X \lim_{k\to\infty} g_k \overset{\text{MCT}}{=} \lim_{k\to\infty} \sum_X g_k \le \liminf_{n\to\infty} \sum_X f_n.$$

■

Remark 4.13. If $A = \sum_X a < \infty$, then for all $\varepsilon > 0$ there exists $\alpha_\varepsilon \subset\subset X$ such that

$$A \ge \sum_\alpha a \ge A - \varepsilon$$

for all $\alpha \subset\subset X$ containing α_ε or equivalently,

$$\left| A - \sum_\alpha a \right| \le \varepsilon \tag{4.11}$$

for all $\alpha \subset\subset X$ containing α_ε. Indeed, choose α_ε so that $\sum_{\alpha_\varepsilon} a \ge A - \varepsilon$.

4.3 Sums of complex functions

Definition 4.14. *Suppose that $a : X \to \mathbb{C}$ is a function, we say that*

$$\sum_X a = \sum_{x\in X} a(x)$$

exists and is equal to $A \in \mathbb{C}$, if for all $\varepsilon > 0$ there is a finite subset $\alpha_\varepsilon \subset X$ such that for all $\alpha \subset\subset X$ containing α_ε we have

$$\left| A - \sum_\alpha a \right| \le \varepsilon.$$

The following lemma is left as an exercise to the reader.

Lemma 4.15. *Suppose that $a, b : X \to \mathbb{C}$ are two functions such that $\sum_X a$ and $\sum_X b$ exist, then $\sum_X (a + \lambda b)$ exists for all $\lambda \in \mathbb{C}$ and*

$$\sum_X (a + \lambda b) = \sum_X a + \lambda \sum_X b.$$

Definition 4.16 (Summable). *We call a function $a : X \to \mathbb{C}$* ***summable*** *if*

$$\sum_X |a| < \infty.$$

Proposition 4.17. *Let $a : X \to \mathbb{C}$ be a function, then $\sum_X a$ exists iff $\sum_X |a| < \infty$, i.e. iff a is summable. Moreover if a is summable, then*

$$\left| \sum_X a \right| \le \sum_X |a| .$$

Proof. If $\sum_X |a| < \infty$, then $\sum_X (\operatorname{Re} a)^{\pm} < \infty$ and $\sum_X (\operatorname{Im} a)^{\pm} < \infty$ and hence by Remark 4.13 these sums exists in the sense of Definition 4.14. Therefore by Lemma 4.15, $\sum_X a$ exists and

$$\sum_X a = \sum_X (\operatorname{Re} a)^+ - \sum_X (\operatorname{Re} a)^- + i \left(\sum_X (\operatorname{Im} a)^+ - \sum_X (\operatorname{Im} a)^- \right).$$

Conversely, if $\sum_X |a| = \infty$ then, because $|a| \le |\operatorname{Re} a| + |\operatorname{Im} a|$, we must have

$$\sum_X |\operatorname{Re} a| = \infty \text{ or } \sum_X |\operatorname{Im} a| = \infty.$$

Thus it suffices to consider the case where $a : X \to \mathbb{R}$ is a real function. Write $a = a^+ - a^-$ where

$$a^+(x) = \max(a(x), 0) \text{ and } a^-(x) = \max(-a(x), 0). \tag{4.12}$$

Then $|a| = a^+ + a^-$ and

$$\infty = \sum_X |a| = \sum_X a^+ + \sum_X a^-$$

which shows that either $\sum_X a^+ = \infty$ or $\sum_X a^- = \infty$. Suppose, with out loss of generality, that $\sum_X a^+ = \infty$. Let $X' := \{x \in X : a(x) \ge 0\}$, then we know that $\sum_{X'} a = \infty$ which means there are finite subsets $\alpha_n \subset X' \subset X$ such that $\sum_{\alpha_n} a \ge n$ for all n. Thus if $\alpha \subset\subset X$ is any finite set, it follows that $\lim_{n\to\infty} \sum_{\alpha_n \cup \alpha} a = \infty$, and therefore $\sum_X a$ can not exist as a number in $\mathbb{R}$. Finally if a is summable, write $\sum_X a = \rho e^{i\theta}$ with $\rho \ge 0$ and $\theta \in \mathbb{R}$, then

$$\begin{aligned}\left|\sum_X a\right| &= \rho = e^{-i\theta}\sum_X a = \sum_X e^{-i\theta} a \\ &= \sum_X \operatorname{Re}\left[e^{-i\theta}a\right] \le \sum_X \left(\operatorname{Re}\left[e^{-i\theta}a\right]\right)^+ \\ &\le \sum_X \left|\operatorname{Re}\left[e^{-i\theta}a\right]\right| \le \sum_X \left|e^{-i\theta}a\right| \le \sum_X |a|\,.\end{aligned}$$

Alternatively, this may be proved by approximating $\sum_X a$ by a finite sum and then using the triangle inequality of $|\cdot|$. ■

Remark 4.18. Suppose that $X = \mathbb{N}$ and $a : \mathbb{N} \to \mathbb{C}$ is a sequence, then it is not necessarily true that

$$\sum_{n=1}^{\infty} a(n) = \sum_{n\in\mathbb{N}} a(n). \tag{4.13}$$

This is because

$$\sum_{n=1}^{\infty} a(n) = \lim_{N\to\infty} \sum_{n=1}^{N} a(n)$$

depends on the ordering of the sequence a where as $\sum_{n\in\mathbb{N}} a(n)$ does not. For example, take $a(n) = (-1)^n/n$ then $\sum_{n\in\mathbb{N}} |a(n)| = \infty$ i.e. $\sum_{n\in\mathbb{N}} a(n)$ does **not** exist while $\sum_{n=1}^{\infty} a(n)$ does exist. On the other hand, if

$$\sum_{n\in\mathbb{N}} |a(n)| = \sum_{n=1}^{\infty} |a(n)| < \infty$$

then Eq. (4.13) is valid.

Theorem 4.19 (Dominated Convergence Theorem for Sums). *Suppose that $f_n : X \to \mathbb{C}$ is a sequence of functions on X such that $f(x) = \lim_{n\to\infty} f_n(x) \in \mathbb{C}$ exists for all $x \in X$. Further assume there is a* ***dominating function*** *$g : X \to [0,\infty)$ such that*

$$|f_n(x)| \le g(x) \text{ for all } x \in X \text{ and } n \in \mathbb{N} \tag{4.14}$$

and that g is summable. Then

$$\lim_{n\to\infty} \sum_{x\in X} f_n(x) = \sum_{x\in X} f(x). \tag{4.15}$$

Proof. Notice that $|f| = \lim |f_n| \le g$ so that f is summable. By considering the real and imaginary parts of f separately, it suffices to prove the theorem in the case where f is real. By Fatou's Lemma,

$$\sum_X (g \pm f) = \sum_X \liminf_{n\to\infty} (g \pm f_n) \le \liminf_{n\to\infty} \sum_X (g \pm f_n)$$
$$= \sum_X g + \liminf_{n\to\infty} \left(\pm \sum_X f_n \right).$$

Since $\liminf_{n\to\infty}(-a_n) = -\limsup_{n\to\infty} a_n$, we have shown,

$$\sum_X g \pm \sum_X f \le \sum_X g + \begin{cases} \liminf_{n\to\infty} \sum_X f_n \\ -\limsup_{n\to\infty} \sum_X f_n \end{cases}$$

and therefore

$$\limsup_{n\to\infty} \sum_X f_n \le \sum_X f \le \liminf_{n\to\infty} \sum_X f_n.$$

This shows that $\lim\limits_{n\to\infty} \sum_X f_n$exists and is equal to $\sum_X f$. ∎

Proof. (Second Proof.) Passing to the limit in Eq. (4.14) shows that $|f| \le g$ and in particular that f is summable. Given $\varepsilon > 0$, let $\alpha \subset\subset X$ such that

$$\sum_{X \setminus \alpha} g \le \varepsilon.$$

Then for $\beta \subset\subset X$ such that $\alpha \subset \beta$,

$$\begin{aligned} \left| \sum_\beta f - \sum_\beta f_n \right| &= \left| \sum_\beta (f - f_n) \right| \\ &\le \sum_\beta |f - f_n| = \sum_\alpha |f - f_n| + \sum_{\beta\setminus\alpha} |f - f_n| \\ &\le \sum_\alpha |f - f_n| + 2 \sum_{\beta\setminus\alpha} g \\ &\le \sum_\alpha |f - f_n| + 2\varepsilon. \end{aligned}$$

and hence that

$$\left| \sum_\beta f - \sum_\beta f_n \right| \le \sum_\alpha |f - f_n| + 2\varepsilon.$$

Since this last equation is true for all such $\beta \subset\subset X$, we learn that

$$\left| \sum_X f - \sum_X f_n \right| \le \sum_\alpha |f - f_n| + 2\varepsilon$$

which then implies that

$$\begin{aligned}\limsup_{n\to\infty}\left|\sum_X f - \sum_X f_n\right| &\le \limsup_{n\to\infty}\sum_\alpha |f-f_n| + 2\varepsilon \\ &= 2\varepsilon.\end{aligned}$$

Because $\varepsilon > 0$ is arbitrary we conclude that

$$\limsup_{n\to\infty}\left|\sum_X f - \sum_X f_n\right| = 0.$$

which is the same as Eq. (4.15). ■

Remark 4.20. Theorem 4.19 may easily be generalized as follows. Suppose f_n, g_n, g are summable functions on X such that $f_n \to f$ and $g_n \to g$ pointwise, $|f_n| \le g_n$ and $\sum_X g_n \to \sum_X g$ as $n \to \infty$. Then f is summable and Eq. (4.15) still holds. For the proof we use Fatou's Lemma to again conclude

$$\begin{aligned}\sum_X (g \pm f) &= \sum_X \liminf_{n\to\infty}(g_n \pm f_n) \le \liminf_{n\to\infty}\sum_X (g_n \pm f_n) \\ &= \sum_X g + \liminf_{n\to\infty}\left(\pm\sum_X f_n\right)\end{aligned}$$

and then proceed exactly as in the first proof of Theorem 4.19.

4.4 Iterated sums and the Fubini and Tonelli Theorems

Let X and Y be two sets. The proof of the following lemma is left to the reader.

Lemma 4.21. *Suppose that $a : X \to \mathbb{C}$ is function and $F \subset X$ is a subset such that $a(x) = 0$ for all $x \notin F$. Then $\sum_F a$ exists iff $\sum_X a$ exists and when the sums exists,*

$$\sum_X a = \sum_F a.$$

Theorem 4.22 (Tonelli's Theorem for Sums). *Suppose that $a : X \times Y \to [0, \infty]$, then*

$$\sum_{X\times Y} a = \sum_X\sum_Y a = \sum_Y\sum_X a.$$

Proof. It suffices to show, by symmetry, that

$$\sum_{X\times Y} a = \sum_X\sum_Y a$$

Let $\Lambda \subset\subset X \times Y$. The for any $\alpha \subset\subset X$ and $\beta \subset\subset Y$ such that $\Lambda \subset \alpha \times \beta$, we have

$$\sum_{\Lambda} a \leq \sum_{\alpha \times \beta} a = \sum_{\alpha} \sum_{\beta} a \leq \sum_{\alpha} \sum_{Y} a \leq \sum_{X} \sum_{Y} a,$$

i.e. $\sum_{\Lambda} a \leq \sum_X \sum_Y a$. Taking the sup over Λ in this last equation shows

$$\sum_{X \times Y} a \leq \sum_X \sum_Y a.$$

For the reverse inequality, for each $x \in X$ choose $\beta_n^x \subset\subset X$ such that $\beta_n^x \uparrow$ as $n \uparrow$ and

$$\sum_{y \in Y} a(x,y) = \lim_{n \to \infty} \sum_{y \in \beta_n^x} a(x,y).$$

If $\alpha \subset\subset X$ is a given finite subset of X, then

$$\sum_{y \in Y} a(x,y) = \lim_{n \to \infty} \sum_{y \in \beta_n} a(x,y) \text{ for all } x \in \alpha$$

where $\beta_n := \cup_{x \in \alpha} \beta_n^x \subset\subset X$. Hence

$$\begin{aligned} \sum_{x \in \alpha} \sum_{y \in Y} a(x,y) &= \sum_{x \in \alpha} \lim_{n \to \infty} \sum_{y \in \beta_n} a(x,y) = \lim_{n \to \infty} \sum_{x \in \alpha} \sum_{y \in \beta_n} a(x,y) \\ &= \lim_{n \to \infty} \sum_{(x,y) \in \alpha \times \beta_n} a(x,y) \leq \sum_{X \times Y} a. \end{aligned}$$

Since α is arbitrary, it follows that

$$\sum_{x \in X} \sum_{y \in Y} a(x,y) = \sup_{\alpha \subset\subset X} \sum_{x \in \alpha} \sum_{y \in Y} a(x,y) \leq \sum_{X \times Y} a$$

which completes the proof. ∎

Theorem 4.23 (Fubini's Theorem for Sums). *Now suppose that $a : X \times Y \to \mathbb{C}$ is a summable function, i.e. by Theorem 4.22 any one of the following equivalent conditions hold:*

1. $\sum_{X \times Y} |a| < \infty$,
2. $\sum_X \sum_Y |a| < \infty$ *or*
3. $\sum_Y \sum_X |a| < \infty$.
   Then

$$\sum_{X \times Y} a = \sum_X \sum_Y a = \sum_Y \sum_X a.$$

Proof. If $a : X \to \mathbb{R}$ is real valued the theorem follows by applying Theorem 4.22 to $a^{\pm}$ – the positive and negative parts of a. The general result holds for complex valued functions a by applying the real version just proved to the real and imaginary parts of a. ∎

4.5 Exercises

Exercise 4.1. Now suppose for each $n \in \mathbb{N} := \{1, 2, \ldots\}$ that $f_n : X \to \mathbb{R}$ is a function. Let

$$D := \{x \in X : \lim_{n\to\infty} f_n(x) = +\infty\}$$

show that

$$D = \cap_{M=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n \geq N} \{x \in X : f_n(x) \geq M\}. \tag{4.16}$$

Exercise 4.2. Let $f_n : X \to \mathbb{R}$ be as in the last problem. Let

$$C := \{x \in X : \lim_{n\to\infty} f_n(x) \text{ exists in } \mathbb{R}\}.$$

Find an expression for C similar to the expression for D in (4.16). (Hint: use the Cauchy criteria for convergence.)

4.5.1 Limit Problems

Exercise 4.3. Show $\liminf_{n\to\infty}(-a_n) = -\limsup_{n\to\infty} a_n$.

Exercise 4.4. Suppose that $\limsup_{n\to\infty} a_n = M \in \bar{\mathbb{R}}$, show that there is a subsequence $\{a_{n_k}\}_{k=1}^{\infty}$ of $\{a_n\}_{n=1}^{\infty}$ such that $\lim_{k\to\infty} a_{n_k} = M$.

Exercise 4.5. Show that

$$\limsup_{n\to\infty}(a_n + b_n) \leq \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \tag{4.17}$$

provided that the right side of Eq. (4.17) is well defined, i.e. no $\infty - \infty$ or $-\infty + \infty$ type expressions. (It is OK to have $\infty + \infty = \infty$ or $-\infty - \infty = -\infty$, etc.)

Exercise 4.6. Suppose that $a_n \geq 0$ and $b_n \geq 0$ for all $n \in \mathbb{N}$. Show

$$\limsup_{n\to\infty}(a_n b_n) \leq \limsup_{n\to\infty} a_n \cdot \limsup_{n\to\infty} b_n, \tag{4.18}$$

provided the right hand side of (4.18) is not of the form $0 \cdot \infty$ or $\infty \cdot 0$.

Exercise 4.7. Prove Lemma 4.15.

Exercise 4.8. Prove Lemma 4.21.

Let $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ be two sequences of real numbers.

4.5.2 Dominated Convergence Theorem Problems

Notation 4.24 *For $u_0 \in \mathbb{R}^n$ and $\delta > 0$, let $B_{u_0}(\delta) := \{x \in \mathbb{R}^n : |x - u_0| < \delta\}$ be the ball in $\mathbb{R}^n$ centered at u_0 with radius δ.*

Exercise 4.9. Suppose $U \subset \mathbb{R}^n$ is a set and $u_0 \in U$ is a point such that $U \cap (B_{u_0}(\delta) \setminus \{u_0\}) \neq \emptyset$ for all $\delta > 0$. Let $G : U \setminus \{u_0\} \to \mathbb{C}$ be a function on $U \setminus \{u_0\}$. Show that $\lim_{u \to u_0} G(u)$ exists and is equal to $\lambda \in \mathbb{C}$,[1] iff for all sequences $\{u_n\}_{n=1}^{\infty} \subset U \setminus \{u_0\}$ which converge to u_0 (i.e. $\lim_{n\to\infty} u_n = u_0$) we have $\lim_{n\to\infty} G(u_n) = \lambda$.

Exercise 4.10. Suppose that Y is a set, $U \subset \mathbb{R}^n$ is a set, and $f : U \times Y \to \mathbb{C}$ is a function satisfying:

1. For each $y \in Y$, the function $u \in U \to f(u, y)$ is continuous on U.[2]
2. There is a summable function $g : Y \to [0, \infty)$ such that

$$|f(u, y)| \le g(y) \text{ for all } y \in Y \text{ and } u \in U.$$

Show that
$$F(u) := \sum_{y \in Y} f(u, y) \tag{4.19}$$
is a continuous function for $u \in U$.

Exercise 4.11. Suppose that Y is a set, $J = (a, b) \subset \mathbb{R}$ is an interval, and $f : J \times Y \to \mathbb{C}$ is a function satisfying:

1. For each $y \in Y$, the function $u \to f(u, y)$ is differentiable on J,
2. There is a summable function $g : Y \to [0, \infty)$ such that

$$\left| \frac{\partial}{\partial u} f(u, y) \right| \le g(y) \text{ for all } y \in Y \text{ and } u \in J.$$

3. There is a $u_0 \in J$ such that $\sum_{y \in Y} |f(u_0, y)| < \infty$.

Show:

a) for all $u \in J$ that $\sum_{y \in Y} |f(u, y)| < \infty$.

---

[1] More explicitly, $\lim_{u \to u_0} G(u) = \lambda$ means for every every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|G(u) - \lambda| < \epsilon \text{ whenerver } u \in U \cap (B_{u_0}(\delta) \setminus \{u_0\}).$$

[2] To say $g := f(\cdot, y)$ is continuous on U means that $g : U \to \mathbb{C}$ is continuous relative to the metric on $\mathbb{R}^n$ restricted to U.

b) Let $F(u) := \sum_{y\in Y} f(u,y)$, show F is differentiable on J and that

$$\dot{F}(u) = \sum_{y\in Y} \frac{\partial}{\partial u} f(u,y).$$

(Hint: Use the mean value theorem.)

Exercise 4.12 (Differentiation of Power Series). Suppose $R > 0$ and $\{a_n\}_{n=0}^{\infty}$ is a sequence of complex numbers such that $\sum_{n=0}^{\infty} |a_n| r^n < \infty$ for all $r \in (0,R)$. Show, using Exercise 4.11, $f(x) := \sum_{n=0}^{\infty} a_n x^n$ is continuously differentiable for $x \in (-R,R)$ and

$$f'(x) = \sum_{n=0}^{\infty} n a_n x^{n-1} = \sum_{n=1}^{\infty} n a_n x^{n-1}.$$

Exercise 4.13. Show the functions

$$e^x := \sum_{n=0}^{\infty} \frac{x^n}{n!}, \tag{4.20}$$

$$\sin x := \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} \text{ and} \tag{4.21}$$

$$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} \tag{4.22}$$

are infinitely differentiable and they satisfy

$$\frac{d}{dx} e^x = e^x \text{ with } e^0 = 1$$
$$\frac{d}{dx} \sin x = \cos x \text{ with } \sin(0) = 0$$
$$\frac{d}{dx} \cos x = -\sin x \text{ with } \cos(0) = 1.$$

Exercise 4.14. Continue the notation of Exercise 4.13.

1. Use the product and the chain rule to show,

$$\frac{d}{dx}\left[e^{-x} e^{(x+y)}\right] = 0$$

   and conclude from this, that $e^{-x} e^{(x+y)} = e^y$ for all $x, y \in \mathbb{R}$. In particular taking $y = 0$ this implies that $e^{-x} = 1/e^x$ and hence that $e^{(x+y)} = e^x e^y$. Use this result to show $e^x \uparrow \infty$ as $x \uparrow \infty$ and $e^x \downarrow 0$ as $x \downarrow -\infty$.
   Remark: since $e^x \geq \sum_{n=0}^{N} \frac{x^n}{n!}$ when $x \geq 0$, it follows that $\lim_{x\to\infty} \frac{x^n}{e^x} = 0$ for any $n \in \mathbb{N}$, i.e. e^x grows at a rate faster than any polynomial in x as $x \to \infty$.

2. Use the product rule to show

$$\frac{d}{dx}\left(\cos^2 x + \sin^2 x\right) = 0$$

and use this to conclude that $\cos^2 x + \sin^2 x = 1$ for all $x \in \mathbb{R}$.

Exercise 4.15. Let $\{a_n\}_{n=-\infty}^{\infty}$ be a summable sequence of complex numbers, i.e. $\sum_{n=-\infty}^{\infty} |a_n| < \infty$. For $t \geq 0$ and $x \in \mathbb{R}$, define

$$F(t,x) = \sum_{n=-\infty}^{\infty} a_n e^{-tn^2} e^{inx},$$

where as usual $e^{ix} = \cos(x) + i\sin(x)$, this is motivated by replacing x in Eq. (4.20) by ix and comparing the result to Eqs. (4.21) and (4.22).

1. $F(t,x)$ is continuous for $(t,x) \in [0,\infty)\times\mathbb{R}$. **Hint:** Let $Y = \mathbb{Z}$ and $u = (t,x)$ and use Exercise 4.10.
2. $\partial F(t,x)/\partial t$, $\partial F(t,x)/\partial x$ and $\partial^2 F(t,x)/\partial x^2$ exist for $t > 0$ and $x \in \mathbb{R}$. **Hint:** Let $Y = \mathbb{Z}$ and $u = t$ for computing $\partial F(t,x)/\partial t$ and $u = x$ for computing $\partial F(t,x)/\partial x$ and $\partial^2 F(t,x)/\partial x^2$. See Exercise 4.11.
3. F satisfies the heat equation, namely

$$\partial F(t,x)/\partial t = \partial^2 F(t,x)/\partial x^2 \text{ for } t > 0 \text{ and } x \in \mathbb{R}.$$

5

ℓ^p – spaces, Minkowski and Holder Inequalities

In this chapter, let $\mu : X \to (0,\infty)$ be a given function. Let $\mathbb{F}$ denote either $\mathbb{R}$ or $\mathbb{C}$. For $p \in (0,\infty)$ and $f : X \to \mathbb{F}$, let

$$\|f\|_p := (\sum_{x\in X} |f(x)|^p \, \mu(x))^{1/p}$$

and for $p = \infty$ let

$$\|f\|_\infty = \sup\{|f(x)| : x \in X\}.$$

Also, for $p > 0$, let

$$\ell^p(\mu) = \{f : X \to \mathbb{F} : \|f\|_p < \infty\}.$$

In the case where $\mu(x) = 1$ for all $x \in X$ we will simply write $\ell^p(X)$ for $\ell^p(\mu)$.

Definition 5.1. *A **norm** on a vector space Z is a function $\|\cdot\| : Z \to [0,\infty)$ such that*

1. *(Homogeneity) $\|\lambda f\| = |\lambda| \, \|f\|$ for all $\lambda \in \mathbb{F}$ and $f \in Z$.*
2. *(Triangle inequality) $\|f + g\| \le \|f\| + \|g\|$ for all $f, g \in Z$.*
3. *(Positive definite) $\|f\| = 0$ implies $f = 0$.*

*A pair $(Z, \|\cdot\|)$ where Z is a vector space and $\|\cdot\|$ is a norm on Z is called a **normed vector space.***

The rest of this section is devoted to the proof of the following theorem.

Theorem 5.2. *For $p \in [1,\infty]$, $(\ell^p(\mu), \|\cdot\|_p)$ is a normed vector space.*

Proof. The only difficulty is the proof of the triangle inequality which is the content of Minkowski's Inequality proved in Theorem 5.8 below. ∎

Proposition 5.3. *Let $f:[0,\infty)\to[0,\infty)$ be a continuous strictly increasing function such that $f(0)=0$ (for simplicity) and $\lim_{s\to\infty} f(s)=\infty$. Let $g=f^{-1}$ and for $s,t\geq 0$ let*

$$F(s)=\int_0^s f(s')ds' \text{ and } G(t)=\int_0^t g(t')dt'.$$

Then for all $s,t\geq 0$,

$$st\leq F(s)+G(t)$$

and equality holds iff $t=f(s)$.

Proof. Let

$$\begin{aligned} A_s &:= \{(\sigma,\tau):0\leq\tau\leq f(\sigma) \text{ for } 0\leq\sigma\leq s\} \text{ and} \\ B_t &:= \{(\sigma,\tau):0\leq\sigma\leq g(\tau) \text{ for } 0\leq\tau\leq t\} \end{aligned}$$

then as one sees from Figure 5.1, $[0,s]\times[0,t]\subset A_s\cup B_t$. (In the figure: $s=3$, $t=1$, A_3 is the region under $t=f(s)$ for $0\leq s\leq 3$ and B_1 is the region to the left of the curve $s=g(t)$ for $0\leq t\leq 1$.) Hence if m denotes the area of a region in the plane, then

$$st=m\left([0,s]\times[0,t]\right)\leq m(A_s)+m(B_t)=F(s)+G(t).$$

As it stands, this proof is a bit on the intuitive side. However, it will become rigorous if one takes m to be Lebesgue measure on the plane which will be introduced later. We can also give a calculus proof of this theorem under the additional assumption that f is C^1. (This restricted version of the theorem is all we need in this section.) To do this fix $t\geq 0$ and let

$$h(s)=st-F(s)=\int_0^s (t-f(\sigma))d\sigma.$$

If $\sigma>g(t)=f^{-1}(t)$, then $t-f(\sigma)<0$ and hence if $s>g(t)$, we have

$$\begin{aligned} h(s) &= \int_0^s (t-f(\sigma))d\sigma=\int_0^{g(t)}(t-f(\sigma))d\sigma+\int_{g(t)}^s (t-f(\sigma))d\sigma \\ &\leq \int_0^{g(t)}(t-f(\sigma))d\sigma=h(g(t)). \end{aligned}$$

Combining this with $h(0)=0$ we see that $h(s)$ takes its maximum at some point $s\in(0,t]$ and hence at a point where $0=h'(s)=t-f(s)$. The only solution to this equation is $s=g(t)$ and we have thus shown

$$st-F(s)=h(s)\leq\int_0^{g(t)}(t-f(\sigma))d\sigma=h(g(t))$$

with equality when $s = g(t)$. To finish the proof we must show $\int_0^{g(t)}(t - f(\sigma))d\sigma = G(t)$. This is verified by making the change of variables $\sigma = g(\tau)$ and then integrating by parts as follows:

$$\begin{aligned}\int_0^{g(t)}(t - f(\sigma))d\sigma &= \int_0^t (t - f(g(\tau)))g'(\tau)d\tau = \int_0^t (t-\tau)g'(\tau)d\tau \\ &= \int_0^t g(\tau)d\tau = G(t).\end{aligned}$$

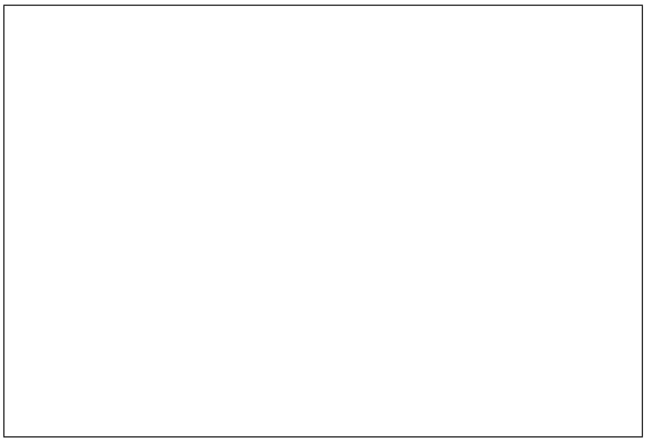


Fig. 5.1. A picture proof of Proposition 5.3.

■

Definition 5.4. *The conjugate exponent* $q \in [1, \infty]$ *to* $p \in [1, \infty]$ *is* $q := \frac{p}{p-1}$ *with the conventions that* $q = \infty$ *if* $p = 1$ *and* $q = 1$ *if* $p = \infty$. *Notice that* q *is characterized by any of the following identities:*

$$\frac{1}{p} + \frac{1}{q} = 1,\ 1 + \frac{q}{p} = q,\ p - \frac{p}{q} = 1 \text{ and } q(p-1) = p. \tag{5.1}$$

Lemma 5.5. *Let* $p \in (1, \infty)$ *and* $q := \frac{p}{p-1} \in (1, \infty)$ *be the conjugate exponent. Then*

$$st \le \frac{s^p}{p} + \frac{t^q}{q} \text{ for all } s, t \ge 0$$

with equality if and only if $t^q = s^p$.

Proof. Let $F(s) = \frac{s^p}{p}$ for $p > 1$. Then $f(s) = s^{p-1} = t$ and $g(t) = t^{\frac{1}{p-1}} = t^{q-1}$, wherein we have used $q - 1 = p/(p-1) - 1 = 1/(p-1)$. Therefore $G(t) = t^q/q$ and hence by Proposition 5.3,

$$st \leq \frac{s^p}{p} + \frac{t^q}{q}$$

with equality iff $t = s^{p-1}$, i.e. $t^q = s^{q(p-1)} = s^p$. For those who do not want to use Proposition 5.3, here is a direct calculus proof. Fix $t > 0$ and let

$$h(s) := st - \frac{s^p}{p}.$$

Then $h(0) = 0$, $\lim_{s\to\infty} h(s) = -\infty$ and $h'(s) = t - s^{p-1}$ which equals zero iff $s = t^{\frac{1}{p-1}}$. Since

$$h\left(t^{\frac{1}{p-1}}\right) = t^{\frac{1}{p-1}}t - \frac{t^{\frac{p}{p-1}}}{p} = t^{\frac{p}{p-1}} - \frac{t^{\frac{p}{p-1}}}{p} = t^q\left(1 - \frac{1}{p}\right) = \frac{t^q}{q},$$

it follows from the first derivative test that

$$\max h = \max\left\{h(0), h\left(t^{\frac{1}{p-1}}\right)\right\} = \max\left\{0, \frac{t^q}{q}\right\} = \frac{t^q}{q}.$$

So we have shown

$$st - \frac{s^p}{p} \leq \frac{t^q}{q} \text{ with equality iff } t = s^{p-1}.$$

■

Theorem 5.6 (Hölder's inequality). *Let $p, q \in [1, \infty]$ be conjugate exponents. For all $f, g : X \to \mathbb{F}$,*

$$\|fg\|_1 \leq \|f\|_p \cdot \|g\|_q. \tag{5.2}$$

If $p \in (1, \infty)$ and f and g are not identically zero, then equality holds in Eq. (5.2) iff

$$\left(\frac{|f|}{\|f\|_p}\right)^p = \left(\frac{|g|}{\|g\|_q}\right)^q. \tag{5.3}$$

Proof. The proof of Eq. (5.2) for $p \in \{1, \infty\}$ is easy and will be left to the reader. The cases where $\|f\|_q = 0$ or ∞ or $\|g\|_p = 0$ or ∞ are easily dealt with and are also left to the reader. So we will assume that $p \in (1, \infty)$ and $0 < \|f\|_q, \|g\|_p < \infty$. Letting $s = |f(x)| / \|f\|_p$ and $t = |g|/\|g\|_q$ in Lemma 5.5 implies

$$\frac{|f(x)\, g(x)|}{\|f\|_p \|g\|_q} \leq \frac{1}{p} \frac{|f(x)|^p}{\|f\|_p} + \frac{1}{q} \frac{|g(x)|^q}{\|g\|^q}$$

with equality iff

$$\frac{|f(x)|^p}{\|f\|_p} = s^p = t^q = \frac{|g(x)|^q}{\|g\|^q}. \tag{5.4}$$

Multiplying this equation by $\mu(x)$ and then summing on x gives

$$\frac{\|fg\|_1}{\|f\|_p \|g\|_q} \leq \frac{1}{p} + \frac{1}{q} = 1$$

with equality iff Eq. (5.4) holds for all $x \in X$, i.e. iff Eq. (5.3) holds. ■

Definition 5.7. *For a complex number* $\lambda \in \mathbb{C}$, *let*

$$\operatorname{sgn}(\lambda) = \begin{cases} \frac{\lambda}{|\lambda|} & \text{if } \lambda \neq 0 \\ 0 & \text{if } \lambda = 0. \end{cases}$$

For $\lambda, \mu \in \mathbb{C}$ *we will write* $\operatorname{sgn}(\lambda) \overset{\circ}{=} \operatorname{sgn}(\mu)$ *if either* $\lambda\mu = 0$ *or* $\lambda\mu \neq 0$ *and* $\operatorname{sgn}(\lambda) = \operatorname{sgn}(\mu)$.

Theorem 5.8 (Minkowski's Inequality). *If* $1 \leq p \leq \infty$ *and* $f, g \in \ell^p(\mu)$ *then*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p. \tag{5.5}$$

Moreover, assuming f *and* g *are not identically zero, equality holds in Eq. (5.5) iff*

$$\begin{aligned} \operatorname{sgn}(f) \overset{\circ}{=} \operatorname{sgn}(g) & \text{ when } p = 1 \text{ and} \\ f = cg \text{ for some } c > 0 & \text{ when } p \in (1, \infty). \end{aligned}$$

Proof. For $p = 1$,

$$\|f + g\|_1 = \sum_X |f + g|\mu \leq \sum_X (|f|\,\mu + |g|\mu) = \sum_X |f|\,\mu + \sum_X |g|\mu$$

with equality iff

$$|f| + |g| = |f + g| \iff \operatorname{sgn}(f) \overset{\circ}{=} \operatorname{sgn}(g).$$

For $p = \infty$,

$$\begin{aligned} \|f + g\|_\infty &= \sup_X |f + g| \leq \sup_X (|f| + |g|) \\ &\leq \sup_X |f| + \sup_X |g| = \|f\|_\infty + \|g\|_\infty. \end{aligned}$$

Now assume that $p \in (1, \infty)$. Since

$$|f + g|^p \leq (2\max(|f|, |g|))^p = 2^p \max(|f|^p, |g|^p) \leq 2^p (|f|^p + |g|^p)$$

it follows that

$$\|f + g\|_p^p \leq 2^p \left(\|f\|_p^p + \|g\|_p^p\right) < \infty.$$

Eq. (5.5) is easily verified if $\|f + g\|_p = 0$, so we may assume $\|f + g\|_p > 0$. Multiplying the inequality,

$$|f + g|^p = |f + g||f + g|^{p-1} \leq (|f| + |g|)|f + g|^{p-1} \tag{5.6}$$

by μ, then summing on x and applying Holder's inequality two times gives

$$\begin{aligned} \sum_X |f + g|^p \mu &\leq \sum_X |f|\ |f + g|^{p-1}\mu + \sum_X |g|\ |f + g|^{p-1}\mu \\ &\leq (\|f\|_p + \|g\|_p)\, \|\, |f + g|^{p-1}\, \|_q. \end{aligned} \tag{5.7}$$

Since $q(p-1) = p$, as in Eq. (5.1),

$$\| |f+g|^{p-1} \|_q^q = \sum_X (|f+g|^{p-1})^q \mu = \sum_X |f+g|^p \mu = \|f+g\|_p^p. \tag{5.8}$$

Combining Eqs. (5.7) and (5.8) shows

$$\|f+g\|_p^p \le (\|f\|_p + \|g\|_p) \|f+g\|_p^{p/q} \tag{5.9}$$

and solving this equation for $\|f+g\|_p$ (making use of Eq. (5.1)) implies Eq. (5.5). Now suppose that f and g are not identically zero and $p \in (1, \infty)$. Equality holds in Eq. (5.5) iff equality holds in Eq. (5.9) iff equality holds in Eq. (5.7) and Eq. (5.6). The latter happens iff

$$\begin{gathered} \operatorname{sgn}(f) \doteq \operatorname{sgn}(g) \text{ and} \\ \left(\frac{|f|}{\|f\|_p}\right)^p = \frac{|f+g|^p}{\|f+g\|_p^p} = \left(\frac{|g|}{\|g\|_p}\right)^p. \end{gathered} \tag{5.10}$$

wherein we have used

$$\left(\frac{|f+g|^{p-1}}{\| |f+g|^{p-1} \|_q}\right)^q = \frac{|f+g|^p}{\|f+g\|_p^p}.$$

Finally Eq. (5.10) is equivalent $|f| = c\,|g|$ with $c = (\|f\|_p / \|g\|_p) > 0$ and this equality along with $\operatorname{sgn}(f) \doteq \operatorname{sgn}(g)$ implies $f = cg$. ■

5.1 Exercises

Exercise 5.1. Generalize Proposition 5.3 as follows. Let $a \in [-\infty, 0]$ and $f : \mathbb{R} \cap [a, \infty) \to [0, \infty)$ be a continuous strictly increasing function such that $\lim_{s\to\infty} f(s) = \infty$, $f(a) = 0$ if $a > -\infty$ or $\lim_{s\to-\infty} f(s) = 0$ if $a = -\infty$. Also let $g = f^{-1}$, $b = f(0) \ge 0$,

$$F(s) = \int_0^s f(s')ds' \text{ and } G(t) = \int_0^t g(t')dt'.$$

Then for all $s, t \ge 0$,

$$st \le F(s) + G(t \vee b) \le F(s) + G(t)$$

and equality holds iff $t = f(s)$. In particular, taking $f(s) = e^s$, prove Young's inequality stating

$$st \le e^s + (t \vee 1) \ln (t \vee 1) - (t \vee 1) \le e^s + t \ln t - t.$$

Hint: Refer to Figures 5.2 and 5.3..

Fig. 5.2. Comparing areas when $t \geq b$ goes the same way as in the text.

Fig. 5.3. When $t \leq b$, notice that $g(t) \leq 0$ but $G(t) \geq 0$. Also notice that $G(t)$ is no longer needed to estimate st.

Part II

Metric and Banach Space Basics

6

Metric Spaces

Definition 6.1. *A function $d : X \times X \to [0, \infty)$ is called a metric if*

1. *(Symmetry) $d(x, y) = d(y, x)$ for all $x, y \in X$*
2. *(Non-degenerate) $d(x, y) = 0$ if and only if $x = y \in X$*
3. *(Triangle inequality) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.*

As primary examples, any normed space $(X, \|\cdot\|)$ (see Definition 5.1) is a metric space with $d(x, y) := \|x - y\|$. Thus the space $\ell^p(\mu)$ (as in Theorem 5.2) is a metric space for all $p \in [1, \infty]$. Also any subset of a metric space is a metric space. For example a surface Σ in $\mathbb{R}^3$ is a metric space with the distance between two points on Σ being the usual distance in $\mathbb{R}^3$.

Definition 6.2. *Let (X, d) be a metric space. The **open ball** $B(x, \delta) \subset X$ centered at $x \in X$ with radius $\delta > 0$ is the set*

$$B(x, \delta) := \{y \in X : d(x, y) < \delta\}.$$

*We will often also write $B(x, \delta)$ as $B_x(\delta)$. We also define the **closed ball** centered at $x \in X$ with radius $\delta > 0$ as the set $C_x(\delta) := \{y \in X : d(x, y) \leq \delta\}$.*

Definition 6.3. *A sequence $\{x_n\}_{n=1}^{\infty}$ in a metric space (X, d) is said to be **convergent** if there exists a point $x \in X$ such that $\lim_{n\to\infty} d(x, x_n) = 0$. In this case we write $\lim_{n\to\infty} x_n = x$ of $x_n \to x$ as $n \to \infty$.*

Exercise 6.1. Show that x in Definition 6.3 is necessarily unique.

Definition 6.4. *A set $E \subset X$ is **bounded** if $E \subset B(x, R)$ for some $x \in X$ and $R < \infty$. A set $F \subset X$ is **closed** iff every convergent sequence $\{x_n\}_{n=1}^{\infty}$ which is contained in F has its limit back in F. A set $V \subset X$ is **open** iff V^c is closed. We will write $F \sqsubset X$ to indicate the F is a closed subset of X and $V \subset_o X$ to indicate the V is an open subset of X. We also let τ_d denote the collection of open subsets of X relative to the metric d.*

Definition 6.5. *A subset $A \subset X$ is a* ***neighborhood*** *of x if there exists an open set $V \subset_o X$ such that $x \in V \subset A$. We will say that $A \subset X$ is an* ***open neighborhood*** *of x if A is open and $x \in A$.*

Exercise 6.2. Let $\mathcal{F}$ be a collection of closed subsets of X, show $\cap\mathcal{F} := \cap_{F\in\mathcal{F}}F$ is closed. Also show that finite unions of closed sets are closed, i.e. if $\{F_k\}_{k=1}^n$ are closed sets then $\cup_{k=1}^n F_k$ is closed. (By taking complements, this shows that the collection of open sets, τ_d, is closed under finite intersections and arbitrary unions.)

The following "continuity" facts of the metric d will be used frequently in the remainder of this book.

Lemma 6.6. *For any non empty subset $A \subset X$, let $d_A(x) := \inf\{d(x,a)|a \in A\}$, then*

$$|d_A(x) - d_A(y)| \le d(x,y)\ \forall\, x, y \in X \tag{6.1}$$

and in particular if $x_n \to x$ in X then $d_A\left(x_n\right) \to d_A\left(x\right)$ as $n \to \infty$. Moreover the set $F_\varepsilon := \{x \in X|d_A(x) \ge \varepsilon\}$ is closed in X.

Proof. Let $a \in A$ and $x, y \in X$, then

$$d(x,a) \le d(x,y) + d(y,a).$$

Take the inf over a in the above equation shows that

$$d_A(x) \le d(x,y) + d_A(y) \quad \forall x, y \in X.$$

Therefore, $d_A(x) - d_A(y) \le d(x,y)$ and by interchanging x and y we also have that $d_A(y) - d_A(x) \le d(x,y)$ which implies Eq. (6.1). If $x_n \to x \in X$, then by Eq. (6.1),

$$|d_A(x) - d_A(x_n)| \le d(x,x_n) \to 0 \text{ as } n \to \infty$$

so that $\lim_{n\to\infty} d_A\left(x_n\right) = d_A\left(x\right).$ Now suppose that $\{x_n\}_{n=1}^\infty \subset F_\varepsilon$ and $x_n \to x$ in X, then

$$d_A\left(x\right) = \lim_{n\to\infty} d_A\left(x_n\right) \ge \varepsilon$$

since $d_A\left(x_n\right) \ge \varepsilon$ for all n. This shows that $x \in F_\varepsilon$ and hence F_ε is closed. ∎

Corollary 6.7. *The function d satisfies,*

$$|d(x,y) - d(x',y')| \le d(y,y') + d(x,x').$$

In particular $d : X \times X \to [0,\infty)$ is "continuous" in the sense that $d(x,y)$ is close to $d(x',y')$ if x is close to x' and y is close to y'. (The notion of continuity will be developed shortly.)

Proof. By Lemma 6.6 for single point sets and the triangle inequality for the absolute value of real numbers,

$$\begin{aligned} |d(x,y) - d(x',y')| &\le |d(x,y) - d(x,y')| + |d(x,y') - d(x',y')| \\ &\le d(y,y') + d(x,x'). \end{aligned}$$

■

Example 6.8. Let $x \in X$ and $\delta > 0$, then $C_x(\delta)$ and $B_x(\delta)^c$ are closed subsets of X. For example if $\{y_n\}_{n=1}^{\infty} \subset C_x(\delta)$ and $y_n \to y \in X$, then $d(y_n, x) \le \delta$ for all n and using Corollary 6.7 it follows $d(y,x) \le \delta$, i.e. $y \in C_x(\delta)$. A similar proof shows $B_x(\delta)^c$ is open, see Exercise 6.3.

Exercise 6.3. Show that $V \subset X$ is open iff for every $x \in V$ there is a $\delta > 0$ such that $B_x(\delta) \subset V$. In particular show $B_x(\delta)$ is open for all $x \in X$ and $\delta > 0$. **Hint:** by definition V is not open iff V^c is not closed.

Lemma 6.9 (Approximating open sets from the inside by closed sets). *Let A be a closed subset of X and $F_\varepsilon := \{x \in X | d_A(x) \ge \varepsilon\} \sqsubset X$ be as in Lemma 6.6. Then $F_\varepsilon \uparrow A^c$ as $\varepsilon \downarrow 0$.*

Proof. It is clear that $d_A(x) = 0$ for $x \in A$ so that $F_\varepsilon \subset A^c$ for each $\varepsilon > 0$ and hence $\cup_{\varepsilon>0} F_\varepsilon \subset A^c$. Now suppose that $x \in A^c \subset_o X$. By Exercise 6.3 there exists an $\varepsilon > 0$ such that $B_x(\varepsilon) \subset A^c$, i.e. $d(x,y) \ge \varepsilon$ for all $y \in A$. Hence $x \in F_\varepsilon$ and we have shown that $A^c \subset \cup_{\varepsilon>0} F_\varepsilon$. Finally it is clear that $F_\varepsilon \subset F_{\varepsilon'}$ whenever $\varepsilon' \le \varepsilon$. ■

Definition 6.10. *Given a set A contained a metric space X, let $\bar{A} \subset X$ be the **closure of** A defined by*

$$\bar{A} := \{x \in X : \exists \, \{x_n\} \subset A \ni x = \lim_{n\to\infty} x_n\}.$$

*That is to say $\bar{A}$ contains all **limit points** of A. We say A is **dense in** X if $\bar{A} = X$, i.e. every element $x \in X$ is a limit of a sequence of elements from A.*

Exercise 6.4. Given $A \subset X$, show $\bar{A}$ is a closed set and in fact

$$\bar{A} = \cap\{F : A \subset F \subset X \text{ with } F \text{ closed}\}. \tag{6.2}$$

That is to say $\bar{A}$ is the smallest closed set containing A.

6.1 Continuity

Suppose that (X, ρ) and (Y, d) are two metric spaces and $f : X \to Y$ is a function.

Definition 6.11. *A function $f : X \to Y$ is continuous at $x \in X$ if for all $\varepsilon > 0$ there is a $\delta > 0$ such that*

$$d(f(x), f(x')) < \varepsilon \text{ provided that } \rho(x, x') < \delta. \tag{6.3}$$

The function f is said to be continuous if f is continuous at all points $x \in X$.

The following lemma gives two other characterizations of continuity of a function at a point.

Lemma 6.12 (Local Continuity Lemma). *Suppose that (X, ρ) and (Y, d) are two metric spaces and $f : X \to Y$ is a function defined in a neighborhood of a point $x \in X$. Then the following are equivalent:*

1. *f is continuous at $x \in X$.*
2. *For all neighborhoods $A \subset Y$ of $f(x)$, $f^{-1}(A)$ is a neighborhood of $x \in X$.*
3. *For all sequences $\{x_n\}_{n=1}^\infty \subset X$ such that $x = \lim_{n\to\infty} x_n$, $\{f(x_n)\}$ is convergent in Y and*

$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right).$$

Proof. $1 \implies 2$. If $A \subset Y$ is a neighborhood of $f(x)$, there exists $\varepsilon > 0$ such that $B_{f(x)}(\varepsilon) \subset A$ and because f is continuous there exists a $\delta > 0$ such that Eq. (6.3) holds. Therefore

$$B_x(\delta) \subset f^{-1}\left(B_{f(x)}(\varepsilon)\right) \subset f^{-1}(A)$$

showing $f^{-1}(A)$ is a neighborhood of x. $2 \implies 3$. Suppose that $\{x_n\}_{n=1}^\infty \subset X$ and $x = \lim_{n\to\infty} x_n$. Then for any $\varepsilon > 0$, $B_{f(x)}(\varepsilon)$ is a neighborhood of $f(x)$ and so $f^{-1}\left(B_{f(x)}(\varepsilon)\right)$ is a neighborhood of x which must containing $B_x(\delta)$ for some $\delta > 0$. Because $x_n \to x$, it follows that $x_n \in B_x(\delta) \subset f^{-1}\left(B_{f(x)}(\varepsilon)\right)$ for a.a. n and this implies $f(x_n) \in B_{f(x)}(\varepsilon)$ for a.a. n, i.e. $d(f(x), f(x_n)) < \varepsilon$ for a.a. n. Since $\varepsilon > 0$ is arbitrary it follows that $\lim_{n\to\infty} f(x_n) = f(x)$. $3. \implies 1$. We will show not $1. \implies$ not 3. If f is not continuous at x, there exists an $\varepsilon > 0$ such that for all $n \in \mathbb{N}$ there exists a point $x_n \in X$ with $\rho(x_n, x) < \frac{1}{n}$ yet $d(f(x_n), f(x)) \geq \varepsilon$. Hence $x_n \to x$ as $n \to \infty$ yet $f(x_n)$ does not converge to $f(x)$. ■

Here is a global version of the previous lemma.

Lemma 6.13 (Global Continuity Lemma). *Suppose that (X, ρ) and (Y, d) are two metric spaces and $f : X \to Y$ is a function defined on all of X. Then the following are equivalent:*

1. *f is continuous.*
2. *$f^{-1}(V) \in \tau_\rho$ for all $V \in \tau_d$, i.e. $f^{-1}(V)$ is open in X if V is open in Y.*
3. *$f^{-1}(C)$ is closed in X if C is closed in Y.*

4. For all convergent sequences $\{x_n\} \subset X$, $\{f(x_n)\}$ is convergent in Y and

$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right).$$

Proof. Since $f^{-1}(A^c) = \left[f^{-1}(A)\right]^c$, it is easily seen that 2. and 3. are equivalent. So because of Lemma 6.12 it only remains to show 1. and 2. are equivalent. If f is continuous and $V \subset Y$ is open, then for every $x \in f^{-1}(V)$, V is a neighborhood of $f(x)$ and so $f^{-1}(V)$ is a neighborhood of x. Hence $f^{-1}(V)$ is a neighborhood of all of its points and from this and Exercise 6.3 it follows that $f^{-1}(V)$ is open. Conversely if $x \in X$ and $A \subset Y$ is a neighborhood of $f(x)$, then there exists $V \subset_o X$ such that $f(x) \in V \subset A$. Hence $x \in f^{-1}(V) \subset f^{-1}(A)$ and by assumption $f^{-1}(V)$ is open showing $f^{-1}(A)$ is a neighborhood of x. Therefore f is continuous at x and since $x \in X$ was arbitrary, f is continuous. ■

Example 6.14. The function d_A defined in Lemma 6.6 is continuous for each $A \subset X$. In particular, if $A = \{x\}$, it follows that $y \in X \to d(y,x)$ is continuous for each $x \in X$.

Exercise 6.5. Use Example 6.14 and Lemma 6.13 to recover the results of Example 6.8.

The next result shows that there are lots of continuous functions on a metric space (X,d).

Lemma 6.15 (Urysohn's Lemma for Metric Spaces). *Let (X,d) be a metric space and suppose that A and B are two disjoint closed subsets of X. Then*

$$f(x) = \frac{d_B(x)}{d_A(x) + d_B(x)} \text{ for } x \in X \tag{6.4}$$

defines a continuous function, $f : X \to [0,1]$, such that $f(x) = 1$ for $x \in A$ and $f(x) = 0$ if $x \in B$.

Proof. By Lemma 6.6, d_A and d_B are continuous functions on X. Since A and B are closed, $d_A(x) > 0$ if $x \notin A$ and $d_B(x) > 0$ if $x \notin B$. Since $A \cap B = \emptyset$, $d_A(x) + d_B(x) > 0$ for all x and $(d_A + d_B)^{-1}$ is continuous as well. The remaining assertions about f are all easy to verify. ■

Sometimes Urysohn's lemma will be use in the following form. Suppose $F \subset V \subset X$ with F being closed and V being open, then there exists $f \in C(X,[0,1]))$ such that $f = 1$ on F while $f = 0$ on V^c. This of course follows from Lemma 6.15 by taking $A = F$ and $B = V^c$.

6.2 Completeness in Metric Spaces

Definition 6.16 (Cauchy sequences). *A sequence $\{x_n\}_{n=1}^\infty$ in a metric space (X,d) is **Cauchy** provided that*

$$\lim_{m,n\to\infty} d(x_n, x_m) = 0.$$

Exercise 6.6. Show that convergent sequences are always Cauchy sequences. The converse is not always true. For example, let $X = \mathbb{Q}$ be the set of rational numbers and $d(x,y) = |x-y|$. Choose a sequence $\{x_n\}_{n=1}^\infty \subset \mathbb{Q}$ which converges to $\sqrt{2} \in \mathbb{R}$, then $\{x_n\}_{n=1}^\infty$ is $(\mathbb{Q}, d)$ – Cauchy but not $(\mathbb{Q}, d)$ – convergent. The sequence does converge in $\mathbb{R}$ however.

Definition 6.17. *A metric space* (X,d) *is* ***complete*** *if all Cauchy sequences are convergent sequences.*

Exercise 6.7. Let (X,d) be a complete metric space. Let $A \subset X$ be a subset of X viewed as a metric space using $d|_{A\times A}$. Show that $(A, d|_{A\times A})$ is complete iff A is a closed subset of X.

Example 6.18. Examples 2. – 4. of complete metric spaces will be verified in Chapter 7 below.

1. $X = \mathbb{R}$ and $d(x,y) = |x-y|$, see Theorem 3.8 above.
2. $X = \mathbb{R}^n$ and $d(x,y) = \|x-y\|_2 = \sum_{i=1}^n (x_i - y_i)^2$.
3. $X = \ell^p(\mu)$ for $p \in [1,\infty]$ and any weight function $\mu : X \to (0,\infty)$.
4. $X = C([0,1], \mathbb{R})$ – the space of continuous functions from $[0,1]$ to $\mathbb{R}$ and
$$d(f,g) := \max_{t\in[0,1]} |f(t) - g(t)|.$$
This is a special case of Lemma 7.3 below.
5. Let $X = C([0,1], \mathbb{R})$ and
$$d(f,g) := \int_0^1 |f(t) - g(t)|\, dt.$$
You are asked in Exercise 7.10 to verify that (X,d) is a metric space which is **not** complete.

Exercise 6.8 (Completions of Metric Spaces). Suppose that (X,d) is a (not necessarily complete) metric space. Using the following outline show there exists a complete metric space $\left(\bar{X}, \bar{d}\right)$ and an isometric map $i : X \to \bar{X}$ such that $i(X)$ is dense in $\bar{X}$, see Definition 6.10.

1. Let $\mathcal{C}$ denote the collection of Cauchy sequences $a = \{a_n\}_{n=1}^\infty \subset X$. Given two element $a, b \in \mathcal{C}$ show
$$d_{\mathcal{C}}(a,b) := \lim_{n\to\infty} d(a_n, b_n) \text{ exists,}$$
$d_{\mathcal{C}}(a,b) \geq 0$ for all $a, b \in \mathcal{C}$ and $d_{\mathcal{C}}$ satisfies the triangle inequality,
$$d_{\mathcal{C}}(a,c) \leq d_{\mathcal{C}}(a,b) + d_{\mathcal{C}}(b,c) \text{ for all } a,b,c \in \mathcal{C}.$$
Thus $(\mathcal{C}, d_{\mathcal{C}})$ would be a metric space if it were true that $d_{\mathcal{C}}(a,b) = 0$ iff $a = b$. This however is false, for example if $a_n = b_n$ for all $n \geq 100$, then $d_{\mathcal{C}}(a,b) = 0$ while a need not equal b.

2. Define two elements $a, b \in \mathcal{C}$ to be equivalent (write $a \sim b$) whenever $d_{\mathcal{C}}(a,b) = 0$. Show " $\sim$ " is an equivalence relation on $\mathcal{C}$ and that $d_{\mathcal{C}}(a',b') = d_{\mathcal{C}}(a,b)$ if $a \sim a'$ and $b \sim b'$. (**Hint:** see Corollary 6.7.)
3. Given $a \in \mathcal{C}$ let $\bar{a} := \{b \in \mathcal{C} : b \sim a\}$ denote the equivalence class containing a and let $\bar{X} := \{\bar{a} : a \in \mathcal{C}\}$ denote the collection of such equivalence classes. Show that $\bar{d}\left(\bar{a}, \bar{b}\right) := d_{\mathcal{C}}(a,b)$ is well defined on $\bar{X} \times \bar{X}$ and verify $\left(\bar{X}, \bar{d}\right)$ is a metric space.
4. For $x \in X$ let $i(x) = \bar{a}$ where a is the constant sequence, $a_n = x$ for all n. Verify that $i : X \to \bar{X}$ is an isometric map and that $i(X)$ is dense in $\bar{X}$.
5. Verify $\left(\bar{X}, \bar{d}\right)$ is complete. **Hint:** if $\{\bar{a}(m)\}_{m=1}^{\infty}$ is a Cauchy sequence in $\bar{X}$ choose $b_m \in X$ such that $\bar{d}\left(i(b_m), \bar{a}(m)\right) \le 1/m$. Then show $\bar{a}(m) \to \bar{b}$ where $b = \{b_m\}_{m=1}^{\infty}$.

6.3 Supplementary Remarks

6.3.1 Word of Caution

Example 6.19. Let (X,d) be a metric space. It is always true that $\overline{B_x(\varepsilon)} \subset C_x(\varepsilon)$ since $C_x(\varepsilon)$ is a closed set containing $B_x(\varepsilon)$. However, it is not always true that $\overline{B_x(\varepsilon)} = C_x(\varepsilon)$. For example let $X = \{1,2\}$ and $d(1,2) = 1$, then $B_1(1) = \{1\}$, $\overline{B_1(1)} = \{1\}$ while $C_1(1) = X$. For another counter example, take

$$X = \left\{(x,y) \in \mathbb{R}^2 : x = 0 \text{ or } x = 1\right\}$$

with the usually Euclidean metric coming from the plane. Then

$$\begin{aligned} B_{(0,0)}(1) &= \left\{(0,y) \in \mathbb{R}^2 : |y| < 1\right\}, \\ \overline{B_{(0,0)}(1)} &= \left\{(0,y) \in \mathbb{R}^2 : |y| \le 1\right\}, \text{ while} \\ C_{(0,0)}(1) &= \overline{B_{(0,0)}(1)} \cup \{(0,1)\}. \end{aligned}$$

In spite of the above examples, Lemmas 6.20 and 6.21 below shows that for certain metric spaces of interest it is true that $\overline{B_x(\varepsilon)} = C_x(\varepsilon)$.

Lemma 6.20. *Suppose that $(X, |\cdot|)$ is a normed vector space and d is the metric on X defined by $d(x,y) = |x-y|$. Then*

$$\begin{aligned} \overline{B_x(\varepsilon)} &= C_x(\varepsilon) \text{ and} \\ \mathrm{bd}(B_x(\varepsilon)) &= \{y \in X : d(x,y) = \varepsilon\}. \end{aligned}$$

where the boundary operation, $\mathrm{bd}(\cdot)$ *is defined in Definition 10.29 below.*

Proof. We must show that $C := C_x(\varepsilon) \subset \overline{B_x(\varepsilon)} =: \bar{B}$. For $y \in C$, let $v = y - x$, then

$$|v| = |y - x| = d(x,y) \le \varepsilon.$$

Let $\alpha_n = 1 - 1/n$ so that $\alpha_n \uparrow 1$ as $n \to \infty$. Let $y_n = x + \alpha_n v$, then $d(x, y_n) = \alpha_n d(x,y) < \varepsilon$, so that $y_n \in B_x(\varepsilon)$ and $d(y, y_n) = 1 - \alpha_n \to 0$ as $n \to \infty$. This shows that $y_n \to y$ as $n \to \infty$ and hence that $y \in \bar{B}$. ■

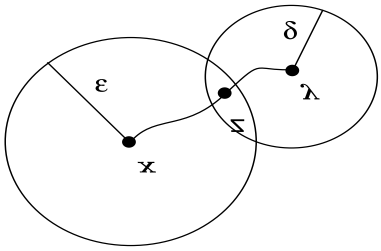


Fig. 6.1. An almost length minimizing curve joining x to y.

6.3.2 Riemannian Metrics

This subsection is not completely self contained and may safely be skipped.

Lemma 6.21. *Suppose that X is a Riemannian (or sub-Riemannian) manifold and d is the metric on X defined by*

$$d(x, y) = \inf \{\ell(\sigma) : \sigma(0) = x \text{ and } \sigma(1) = y\}$$

where $\ell(\sigma)$ is the length of the curve σ. We define $\ell(\sigma) = \infty$ if σ is not piecewise smooth.

Then

$$\overline{B_x(\varepsilon)} = C_x(\varepsilon) \text{ and}$$
$$\mathrm{bd}(B_x(\varepsilon)) = \{y \in X : d(x, y) = \varepsilon\}$$

where the boundary operation, $\mathrm{bd}(\cdot)$ *is defined in Definition 10.29 below.*

Proof. Let $C := C_x(\varepsilon) \subset \overline{B_x(\varepsilon)} =: \bar{B}$. We will show that $C \subset \bar{B}$ by showing $\bar{B}^c \subset C^c$. Suppose that $y \in \bar{B}^c$ and choose $\delta > 0$ such that $B_y(\delta) \cap \bar{B} = \emptyset$. In particular this implies that

$$B_y(\delta) \cap B_x(\varepsilon) = \emptyset.$$

We will finish the proof by showing that $d(x, y) \geq \varepsilon + \delta > \varepsilon$ and hence that $y \in C^c$. This will be accomplished by showing: if $d(x, y) < \varepsilon + \delta$ then $B_y(\delta) \cap B_x(\varepsilon) \neq \emptyset$. If $d(x, y) < \max(\varepsilon, \delta)$ then either $x \in B_y(\delta)$ or $y \in B_x(\varepsilon)$. In either case $B_y(\delta) \cap B_x(\varepsilon) \neq \emptyset$. Hence we may assume that $\max(\varepsilon, \delta) \leq d(x, y) < \varepsilon + \delta$. Let $\alpha > 0$ be a number such that

$$\max(\varepsilon, \delta) \leq d(x, y) < \alpha < \varepsilon + \delta$$

and choose a curve σ from x to y such that $\ell(\sigma) < \alpha$. Also choose $0 < \delta' < \delta$ such that $0 < \alpha - \delta' < \varepsilon$ which can be done since $\alpha - \delta < \varepsilon$. Let $k(t) = d(y, \sigma(t))$ a continuous function on $[0, 1]$ and therefore $k([0, 1]) \subset \mathbb{R}$ is a connected set which contains 0 and $d(x,y)$. Therefore there exists $t_0 \in [0,1]$ such that $d(y,\sigma(t_0)) = k(t_0) = \delta'$. Let $z = \sigma(t_0) \in B_y(\delta)$ then

$$d(x,z) \le \ell(\sigma|_{[0,t_0]}) = \ell(\sigma) - \ell(\sigma|_{[t_0,1]}) < \alpha - d(z,y) = \alpha - \delta' < \varepsilon$$

and therefore $z \in B_x(\varepsilon) \cap B_x(\delta) \neq \emptyset$. ■

Remark 6.22. Suppose again that X is a Riemannian (or sub-Riemannian) manifold and

$$d(x,y) = \inf\{\ell(\sigma) : \sigma(0) = x \text{ and } \sigma(1) = y\}.$$

Let σ be a curve from x to y and let $\varepsilon = \ell(\sigma) - d(x,y)$. Then for all $0 \le u < v \le 1$,

$$d(\sigma(u),\sigma(v)) \le \ell(\sigma|_{[u,v]}) + \varepsilon.$$

So if σ is within ε of a length minimizing curve from x to y that $\sigma|_{[u,v]}$ is within ε of a length minimizing curve from $\sigma(u)$ to $\sigma(v)$. In particular if $d(x,y) = \ell(\sigma)$ then $d(\sigma(u),\sigma(v)) = \ell(\sigma|_{[u,v]})$ for all $0 \le u < v \le 1$, i.e. if σ is a length minimizing curve from x to y that $\sigma|_{[u,v]}$ is a length minimizing curve from $\sigma(u)$ to $\sigma(v)$.

To prove these assertions notice that

$$\begin{aligned} d(x,y) + \varepsilon = \ell(\sigma) &= \ell(\sigma|_{[0,u]}) + \ell(\sigma|_{[u,v]}) + \ell(\sigma|_{[v,1]}) \\ &\ge d(x,\sigma(u)) + \ell(\sigma|_{[u,v]}) + d(\sigma(v),y) \end{aligned}$$

and therefore

$$\begin{aligned} \ell(\sigma|_{[u,v]}) &\le d(x,y) + \varepsilon - d(x,\sigma(u)) - d(\sigma(v),y) \\ &\le d(\sigma(u),\sigma(v)) + \varepsilon. \end{aligned}$$

6.4 Exercises

Exercise 6.9. Let (X,d) be a metric space. Suppose that $\{x_n\}_{n=1}^\infty \subset X$ is a sequence and set $\varepsilon_n := d(x_n,x_{n+1})$. Show that for $m > n$ that

$$d(x_n,x_m) \le \sum_{k=n}^{m-1} \varepsilon_k \le \sum_{k=n}^{\infty} \varepsilon_k.$$

Conclude from this that if

$$\sum_{k=1}^\infty \varepsilon_k = \sum_{n=1}^\infty d(x_n,x_{n+1}) < \infty$$

then $\{x_n\}_{n=1}^\infty$ is Cauchy. Moreover, show that if $\{x_n\}_{n=1}^\infty$ is a convergent sequence and $x = \lim_{n\to\infty} x_n$ then

$$d(x,x_n) \le \sum_{k=n}^\infty \varepsilon_k.$$

Exercise 6.10. Show that (X,d) is a complete metric space iff every sequence $\{x_n\}_{n=1}^\infty \subset X$ such that $\sum_{n=1}^\infty d(x_n,x_{n+1}) < \infty$ is a convergent sequence in X. You may find it useful to prove the following statements in the course of the proof.

1. If $\{x_n\}$ is Cauchy sequence, then there is a subsequence $y_j := x_{n_j}$ such that $\sum_{j=1}^\infty d(y_{j+1},y_j) < \infty$.
2. If $\{x_n\}_{n=1}^\infty$ is Cauchy and there exists a subsequence $y_j := x_{n_j}$ of $\{x_n\}$ such that $x = \lim_{j\to\infty} y_j$ exists, then $\lim_{n\to\infty} x_n$ also exists and is equal to x.

Exercise 6.11. Suppose that $f : [0,\infty) \to [0,\infty)$ is a C^2 – function such that $f(0) = 0$, $f' > 0$ and $f'' \le 0$ and (X,ρ) is a metric space. Show that $d(x,y) = f(\rho(x,y))$ is a metric on X. In particular show that

$$d(x,y) := \frac{\rho(x,y)}{1+\rho(x,y)}$$

is a metric on X. (Hint: use calculus to verify that $f(a+b) \le f(a) + f(b)$ for all $a,b \in [0,\infty)$.)

Exercise 6.12. Let $\{(X_n,d_n)\}_{n=1}^\infty$ be a sequence of metric spaces, $X := \prod_{n=1}^\infty X_n$, and for $x = (x(n))_{n=1}^\infty$ and $y = (y(n))_{n=1}^\infty$ in X let

$$d(x,y) = \sum_{n=1}^\infty 2^{-n} \frac{d_n(x(n),y(n))}{1+d_n(x(n),y(n))}.$$

Show:

1. (X,d) is a metric space,
2. a sequence $\{x_k\}_{k=1}^\infty \subset X$ converges to $x \in X$ iff $x_k(n) \to x(n) \in X_n$ as $k \to \infty$ for each $n \in \mathbb{N}$ and
3. X is complete if X_n is complete for all n.

Exercise 6.13. Suppose (X,ρ) and (Y,d) are metric spaces and A is a dense subset of X.

1. Show that if $F : X \to Y$ and $G : X \to Y$ are two continuous functions such that $F = G$ on A then $F = G$ on X. **Hint:** consider the set $C := \{x \in X : F(x) = G(x)\}$.
2. Suppose $f : A \to Y$ is a function which is uniformly continuous, i.e. for every $\varepsilon > 0$ there exists a $\delta > 0$ such that
$$d(f(a), f(b)) < \varepsilon \text{ for all } a,b \in A \text{ with } \rho(a,b) < \delta.$$
Show there is a unique continuous function $F : X \to Y$ such that $F = f$ on A. **Hint:** each point $x \in X$ is a limit of a sequence consisting of elements from A.
3. Let $X = \mathbb{R} = Y$ and $A = \mathbb{Q} \subset X$, find a function $f : \mathbb{Q} \to \mathbb{R}$ which is continuous on $\mathbb{Q}$ but does **not** extend to a continuous function on $\mathbb{R}$.

7

Banach Spaces

Let $(X, \|\cdot\|)$ be a normed vector space and $d(x,y) := \|x-y\|$ be the associated metric on X. We say $\{x_n\}_{n=1}^{\infty} \subset X$ **converges to** $x \in X$ (and write $\lim_{n\to\infty} x_n = x$ or $x_n \to x$) if

$$0 = \lim_{n\to\infty} d(x, x_n) = \lim_{n\to\infty} \|x - x_n\|.$$

Similarly $\{x_n\}_{n=1}^{\infty} \subset X$ is said to be a **Cauchy sequence** if

$$0 = \lim_{m,n\to\infty} d(x_m, x_n) = \lim_{m,n\to\infty} \|x_m - x_n\|.$$

Definition 7.1 (Banach space). *A normed vector space $(X, \|\cdot\|)$ is a **Banach space** if the associated metric space (X, d) is complete, i.e. all Cauchy sequences are convergent.*

Remark 7.2. Since $\|x\| = d(x, 0)$, it follows from Lemma 6.6 that $\|\cdot\|$ is a continuous function on X and that

$$|\|x\| - \|y\|| \le \|x - y\| \text{ for all } x, y \in X.$$

It is also easily seen that the vector addition and scalar multiplication are continuos on any normed space as the reader is asked to verify in Exercise 7.4. These facts will often be used in the sequel without further mention.

7.1 Examples

Lemma 7.3. *Suppose that X is a set then the bounded functions, $\ell^\infty(X)$, on X is a Banach space with the norm*

$$\|f\| = \|f\|_\infty = \sup_{x\in X} |f(x)|.$$

Moreover if X is a metric space (more generally a topological space, see Chapter 10) the set $BC(X) \subset \ell^\infty(X) = B(X)$ is closed subspace of $\ell^\infty(X)$ and hence is also a Banach space.

Proof. Let $\{f_n\}_{n=1}^\infty \subset \ell^\infty(X)$ be a Cauchy sequence. Since for any $x \in X$, we have

$$|f_n(x) - f_m(x)| \le \|f_n - f_m\|_\infty \tag{7.1}$$

which shows that $\{f_n(x)\}_{n=1}^\infty \subset \mathbb{F}$ is a Cauchy sequence of numbers. Because $\mathbb{F}$ ($\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$) is complete, $f(x) := \lim_{n\to\infty} f_n(x)$ exists for all $x \in X$. Passing to the limit $n \to \infty$ in Eq. (7.1) implies

$$|f(x) - f_m(x)| \le \lim\inf_{n\to\infty} \|f_n - f_m\|_\infty$$

and taking the supremum over $x \in X$ of this inequality implies

$$\|f - f_m\|_\infty \le \lim\inf_{n\to\infty} \|f_n - f_m\|_\infty \to 0 \text{ as } m \to \infty$$

showing $f_m \to f$ in $\ell^\infty(X)$. For the second assertion, suppose that $\{f_n\}_{n=1}^\infty \subset BC(X) \subset \ell^\infty(X)$ and $f_n \to f \in \ell^\infty(X)$. We must show that $f \in BC(X)$, i.e. that f is continuous. To this end let $x, y \in X$, then

$$\begin{aligned} |f(x) - f(y)| &\le |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| \\ &\le 2\,\|f - f_n\|_\infty + |f_n(x) - f_n(y)|\,. \end{aligned}$$

Thus if $\varepsilon > 0$, we may choose n large so that $2\,\|f - f_n\|_\infty < \varepsilon/2$ and then for this n there exists an open neighborhood V_x of $x \in X$ such that $|f_n(x) - f_n(y)| < \varepsilon/2$ for $y \in V_x$. Thus $|f(x) - f(y)| < \varepsilon$ for $y \in V_x$ showing the limiting function f is continuous. ■

Here is an application of this theorem.

Theorem 7.4 (Metric Space Tietze Extension Theorem). *Let (X, d) be a metric space, D be a closed subset of X, $-\infty < a < b < \infty$ and $f \in C(D, [a, b])$. (Here we are viewing D as a metric space with metric $d_D := d_{D\times D}$.) Then there exists $F \in C(X, [a, b])$ such that $F|_D = f$.*

Proof.

1. By scaling and translation (i.e. by replacing f by $(b - a)^{-1}(f - a)$), it suffices to prove Theorem 7.4 with $a = 0$ and $b = 1$.
2. Suppose $\alpha \in (0, 1]$ and $f : D \to [0, \alpha]$ is continuous function. Let $A := f^{-1}([0, \frac{1}{3}\alpha])$ and $B := f^{-1}([\frac{2}{3}\alpha, \alpha])$. By Lemma 6.15 there exists a function $\tilde{g} \in C(X, [0, \alpha/3])$ such that $\tilde{g} = 0$ on A and $\tilde{g} = 1$ on B. Letting $g := \frac{\alpha}{3}\tilde{g}$, we have $g \in C(X, [0, \alpha/3])$ such that $g = 0$ on A and $g = \alpha/3$ on B. Further notice that
$$0 \le f(x) - g(x) \le \frac{2}{3}\alpha \text{ for all } x \in D.$$
3. Now suppose $f : D \to [0, 1]$ is a continuous function as in step 1. Let $g_1 \in C(X, [0, 1/3])$ be as in step 2, see Figure 7.1. with $\alpha = 1$ and let $f_1 := f - g_1|_D \in C(D, [0, 2/3])$. Apply step 2. with $\alpha = 2/3$ and $f = f_1$ to

find $g_2 \in C(X,[0,\frac{1}{3}\frac{2}{3}])$ such that $f_2 := f - (g_1+g_2)|_D \in C(D,[0,\left(\frac{2}{3}\right)^2])$. Continue this way inductively to find $g_n \in C(X,[0,\frac{1}{3}\left(\frac{2}{3}\right)^{n-1}])$ such that

$$f - \sum_{n=1}^{N} g_n|_D =: f_N \in C(D,[0,\left(\frac{2}{3}\right)^N]). \tag{7.2}$$

4. Define $F := \sum_{n=1}^{\infty} g_n$. Since

$$\sum_{n=1}^{\infty} \|g_n\|_\infty \le \sum_{n=1}^{\infty} \frac{1}{3}\left(\frac{2}{3}\right)^{n-1} = \frac{1}{3}\frac{1}{1-\frac{2}{3}} = 1,$$

the series defining F is uniformly convergent so $F \in C(X,[0,1])$. Passing to the limit in Eq. (7.2) shows $f = F|_D$.

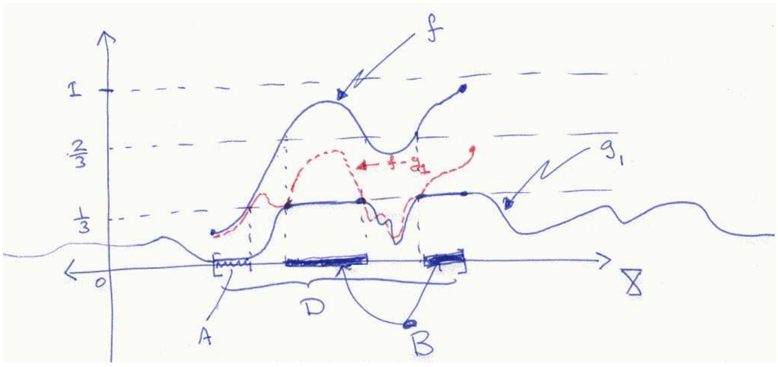


Fig. 7.1. Reducing f by subtracting off a globally defined function $g_1 \in C\left(X,[0,\frac{1}{3}]\right)$.

∎

Theorem 7.5 (Completeness of $\ell^p(\mu)$). *Let X be a set and $\mu : X \to (0,\infty)$ be a given function. Then for any $p \in [1,\infty]$, $(\ell^p(\mu), \|\cdot\|_p)$ is a Banach space.*

Proof. We have already proved this for $p=\infty$ in Lemma 7.3 so we now assume that $p \in [1,\infty)$. Let $\{f_n\}_{n=1}^\infty \subset \ell^p(\mu)$ be a Cauchy sequence. Since for any $x \in X$,

$$|f_n(x) - f_m(x)| \le \frac{1}{\mu(x)} \|f_n - f_m\|_p \to 0 \text{ as } m,n \to \infty$$

it follows that $\{f_n(x)\}_{n=1}^\infty$ is a Cauchy sequence of numbers and $f(x) := \lim_{n\to\infty} f_n(x)$ exists for all $x \in X$. By Fatou's Lemma,

$$\|f_n - f\|_p^p = \sum_X \mu \cdot \lim_{m\to\infty} \inf |f_n - f_m|^p \le \lim_{m\to\infty} \inf \sum_X \mu \cdot |f_n - f_m|^p$$
$$= \lim_{m\to\infty} \inf \|f_n - f_m\|_p^p \to 0 \text{ as } n \to \infty.$$

This then shows that $f = (f - f_n) + f_n \in \ell^p(\mu)$ (being the sum of two ℓ^p – functions) and that $f_n \xrightarrow{\ell^p} f$. ■

Remark 7.6. Let X be a set, Y be a Banach space and $\ell^\infty(X,Y)$ denote the bounded functions $f : X \to Y$ equipped with the norm

$$\|f\| = \|f\|_\infty = \sup_{x\in X} \|f(x)\|_Y .$$

If X is a metric space (or a general topological space, see Chapter 10), let $BC(X,Y)$ denote those $f \in \ell^\infty(X,Y)$ which are continuous. The same proof used in Lemma 7.3 shows that $\ell^\infty(X,Y)$ is a Banach space and that $BC(X,Y)$ is a closed subspace of $\ell^\infty(X,Y)$. Similarly, if $1 \le p < \infty$ we may define

$$\ell^p(X,Y) = \left\{ f : X \to Y : \|f\|_p = \left(\sum_{x\in X} \|f(x)\|_Y^p \right)^{1/p} < \infty \right\}.$$

The same proof as in Theorem 7.5 would then show that $\left(\ell^p(X,Y), \|\cdot\|_p\right)$ is a Banach space.

7.2 Bounded Linear Operators Basics

Definition 7.7. *Let X and Y be normed spaces and $T : X \to Y$ be a linear map. Then T is said to be bounded provided there exists $C < \infty$ such that $\|T(x)\| \le C\|x\|_X$ for all $x \in X$. We denote the best constant by $\|T\|$, i.e.*

$$\|T\| = \sup_{x\neq 0} \frac{\|T(x)\|}{\|x\|} = \sup_{x\neq 0} \{\|T(x)\| : \|x\| = 1\}.$$

The number $\|T\|$ is called the operator norm of T.

Proposition 7.8. *Suppose that X and Y are normed spaces and $T : X \to Y$ is a linear map. The the following are equivalent:*

(a) T is continuous.
(b) T is continuous at 0.
(c) T is bounded.

Proof. (a) ⇒ (b) trivial. (b) ⇒ (c) If T continuous at 0 then there exist $\delta > 0$ such that $\|T(x)\| \le 1$ if $\|x\| \le \delta$. Therefore for any $x \in X$, $\|T(\delta x/\|x\|)\| \le 1$

which implies that $\|T(x)\| \leq \frac{1}{\delta}\|x\|$ and hence $\|T\| \leq \frac{1}{\delta} < \infty$. (c) $\Rightarrow$ (a) Let $x \in X$ and $\varepsilon > 0$ be given. Then

$$\|Ty - Tx\| = \|T(y-x)\| \leq \|T\| \; \|y-x\| < \varepsilon$$

provided $\|y - x\| < \varepsilon/\|T\| := \delta$. ■

For the next three exercises, let $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$ and $T : X \to Y$ be a linear transformation so that T is given by matrix multiplication by an $m \times n$ matrix. Let us identify the linear transformation T with this matrix.

Exercise 7.1. Assume the norms on X and Y are the ℓ^1 – norms, i.e. for $x \in \mathbb{R}^n$, $\|x\| = \sum_{j=1}^n |x_j|$. Then the operator norm of T is given by

$$\|T\| = \max_{1 \leq j \leq n} \sum_{i=1}^m |T_{ij}| \,.$$

Exercise 7.2. Suppose that norms on X and Y are the ℓ^∞ – norms, i.e. for $x \in \mathbb{R}^n$, $\|x\| = \max_{1 \leq j \leq n} |x_j|$. Then the operator norm of T is given by

$$\|T\| = \max_{1 \leq i \leq m} \sum_{j=1}^n |T_{ij}| \,.$$

Exercise 7.3. Assume the norms on X and Y are the ℓ^2 – norms, i.e. for $x \in \mathbb{R}^n$, $\|x\|^2 = \sum_{j=1}^n x_j^2$. Show $\|T\|^2$ is the largest eigenvalue of the matrix $T^{tr}T : \mathbb{R}^n \to \mathbb{R}^n$. **Hint:** Use the spectral theorem for orthogonal matrices.

Notation 7.9 *Let $L(X,Y)$ denote the bounded linear operators from X to Y and $L(X) = L(X,X)$. If $Y = \mathbb{F}$ we write X^* for $L(X,\mathbb{F})$ and call X^* the (continuous)* ***dual space*** *to X.*

Lemma 7.10. *Let X, Y be normed spaces, then the operator norm $\|\cdot\|$ on $L(X,Y)$ is a norm. Moreover if Z is another normed space and $T : X \to Y$ and $S : Y \to Z$ are linear maps, then $\|ST\| \leq \|S\|\|T\|$, where $ST := S \circ T$.*

Proof. As usual, the main point in checking the operator norm is a norm is to verify the triangle inequality, the other axioms being easy to check. If $A, B \in L(X,Y)$ then the triangle inequality is verified as follows:

$$\begin{aligned}\|A+B\| &= \sup_{x \neq 0} \frac{\|Ax + Bx\|}{\|x\|} \leq \sup_{x \neq 0} \frac{\|Ax\| + \|Bx\|}{\|x\|} \\ &\leq \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} + \sup_{x \neq 0} \frac{\|Bx\|}{\|x\|} = \|A\| + \|B\| \,.\end{aligned}$$

For the second assertion, we have for $x \in X$, that

$$\|STx\| \leq \|S\|\|Tx\| \leq \|S\|\|T\|\|x\|.$$

From this inequality and the definition of $\|ST\|$, it follows that $\|ST\| \leq \|S\|\|T\|$. ■

The reader is asked to prove the following continuity lemma in Exercise 7.8.

Lemma 7.11. *Let X, Y and Z be normed spaces. Then the maps*

$$(S, x) \in L(X, Y) \times X \longrightarrow Sx \in Y$$

and

$$(S, T) \in L(X, Y) \times L(Y, Z) \longrightarrow ST \in L(X, Z)$$

are continuous relative to the norms

$$\|(S, x)\|_{L(X,Y)\times X} := \|S\|_{L(X,Y)} + \|x\|_X \text{ and}$$
$$\|(S, T)\|_{L(X,Y)\times L(Y,Z)} := \|S\|_{L(X,Y)} + \|T\|_{L(Y,Z)}$$

on $L(X, Y) \times X$ and $L(X, Y) \times L(Y, Z)$ respectively.

Proposition 7.12. *Suppose that X is a normed vector space and Y is a Banach space. Then $(L(X, Y), \|\cdot\|_{op})$ is a Banach space. In particular the dual space X^* is always a Banach space.*

Proof. Let $\{T_n\}_{n=1}^\infty$ be a Cauchy sequence in $L(X, Y)$. Then for each $x \in X$,

$$\|T_n x - T_m x\| \leq \|T_n - T_m\| \|x\| \to 0 \text{ as } m, n \to \infty$$

showing $\{T_n x\}_{n=1}^\infty$ is Cauchy in Y. Using the completeness of Y, there exists an element $Tx \in Y$ such that

$$\lim_{n\to\infty} \|T_n x - Tx\| = 0.$$

The map $T : X \to Y$ is linear map, since for $x, x' \in X$ and $\lambda \in \mathbb{F}$ we have

$$T(x + \lambda x') = \lim_{n\to\infty} T_n(x + \lambda x') = \lim_{n\to\infty} [T_n x + \lambda T_n x'] = Tx + \lambda Tx',$$

wherein we have used the continuity of the vector space operations in the last equality. Moreover,

$$\|Tx - T_n x\| \leq \|Tx - T_m x\| + \|T_m x - T_n x\| \leq \|Tx - T_m x\| + \|T_m - T_n\| \|x\|$$

and therefore

$$\begin{aligned} \|Tx - T_n x\| &\leq \lim\inf_{m\to\infty} (\|Tx - T_m x\| + \|T_m - T_n\| \|x\|) \\ &= \|x\| \cdot \lim\inf_{m\to\infty} \|T_m - T_n\|. \end{aligned}$$

Hence

$$\|T - T_n\| \leq \lim\inf_{m\to\infty} \|T_m - T_n\| \to 0 \text{ as } n \to \infty.$$

Thus we have shown that $T_n \to T$ in $L(X, Y)$ as desired. ■

The following characterization of a Banach space will sometimes be useful in the sequel.

Theorem 7.13. *A normed space $(X, \|\cdot\|)$ is a Banach space iff for every sequence $\{x_n\}_{n=1}^\infty$ such that $\sum_{n=1}^\infty \|x_n\| < \infty$ implies $\lim_{N\to\infty} \sum_{n=1}^N x_n = s$ exists in X (that is to say every absolutely convergent series is a convergent series in X.) As usual we will denote s by $\sum_{n=1}^\infty x_n$.*

Proof. This is very similar to Exercise 6.10. ($\Rightarrow$)If X is complete and $\sum_{n=1}^\infty \|x_n\| < \infty$ then sequence $s_N := \sum_{n=1}^N x_n$ for $N \in \mathbb{N}$ is Cauchy because (for $N > M$)

$$\|s_N - s_M\| \le \sum_{n=M+1}^N \|x_n\| \to 0 \text{ as } M, N \to \infty.$$

Therefore $s = \sum_{n=1}^\infty x_n := \lim_{N\to\infty} \sum_{n=1}^N x_n$ exists in X. ($\Longleftarrow$) Suppose that $\{x_n\}_{n=1}^\infty$ is a Cauchy sequence and let $\{y_k = x_{n_k}\}_{k=1}^\infty$ be a subsequence of $\{x_n\}_{n=1}^\infty$ such that $\sum_{n=1}^\infty \|y_{n+1} - y_n\| < \infty$. By assumption

$$y_{N+1} - y_1 = \sum_{n=1}^N (y_{n+1} - y_n) \to s = \sum_{n=1}^\infty (y_{n+1} - y_n) \in X \text{ as } N \to \infty.$$

This shows that $\lim_{N\to\infty} y_N$ exists and is equal to $x := y_1 + s$. Since $\{x_n\}_{n=1}^\infty$ is Cauchy,

$$\|x - x_n\| \le \|x - y_k\| + \|y_k - x_n\| \to 0 \text{ as } k, n \to \infty$$

showing that $\lim_{n\to\infty} x_n$ exists and is equal to x. ∎

Example 7.14. Here is another proof of Theorem 7.12 which makes use of Proposition 7.12. Suppose that $T_n \in L(X, Y)$ is a sequence of operators such that $\sum_{n=1}^\infty \|T_n\| < \infty$. Then

$$\sum_{n=1}^\infty \|T_n x\| \le \sum_{n=1}^\infty \|T_n\| \|x\| < \infty$$

and therefore by the completeness of Y, $Sx := \sum_{n=1}^\infty T_n x = \lim_{N\to\infty} S_N x$ exists in Y, where $S_N := \sum_{n=1}^N T_n$. The reader should check that $S : X \to Y$ so defined is linear. Since,

$$\|Sx\| = \lim_{N\to\infty} \|S_N x\| \le \lim_{N\to\infty} \sum_{n=1}^N \|T_n x\| \le \sum_{n=1}^\infty \|T_n\| \|x\|,$$

S is bounded and

$$\|S\| \leq \sum_{n=1}^{\infty} \|T_n\|. \tag{7.3}$$

Similarly,

$$\begin{aligned}\|Sx - S_M x\| &= \lim_{N\to\infty} \|S_N x - S_M x\| \\ &\leq \lim_{N\to\infty} \sum_{n=M+1}^{N} \|T_n\| \, \|x\| = \sum_{n=M+1}^{\infty} \|T_n\| \, \|x\|\end{aligned}$$

and therefore,

$$\|S - S_M\| \leq \sum_{n=M}^{\infty} \|T_n\| \to 0 \text{ as } M \to \infty.$$

7.3 General Sums in Banach Spaces

Definition 7.15. *Suppose X is a normed space.*

1. *Suppose that $\{x_n\}_{n=1}^{\infty}$ is a sequence in X, then we say $\sum_{n=1}^{\infty} x_n$ **converges** in X and $\sum_{n=1}^{\infty} x_n = s$ if*

$$\lim_{N\to\infty} \sum_{n=1}^{N} x_n = s \text{ in } X.$$

2. *Suppose that $\{x_\alpha : \alpha \in A\}$ is a given collection of vectors in X. We say the sum $\sum_{\alpha\in A} x_\alpha$ **converges** in X and write $s = \sum_{\alpha\in A} x_\alpha \in X$ if for all $\varepsilon > 0$ there exists a finite set $\Gamma_\varepsilon \subset A$ such that $\left\|s - \sum_{\alpha\in\Lambda} x_\alpha\right\| < \varepsilon$ for any $\Lambda \subset\subset A$ such that $\Gamma_\varepsilon \subset \Lambda$.*

Warning: As usual if $\sum_{\alpha\in A} \|x_\alpha\| < \infty$ then $\sum_{\alpha\in A} x_\alpha$ exists in X, see Exercise 7.12. However, unlike the case of real valued sums the existence of $\sum_{\alpha\in A} x_\alpha$ does **not** imply $\sum_{\alpha\in A} \|x_\alpha\| < \infty$. See Proposition 14.19 below, from which one may manufacture counter-examples to this false premise.

Lemma 7.16. *Suppose that $\{x_\alpha \in X : \alpha \in A\}$ is a given collection of vectors in a normed space, X.*

1. *If $s = \sum_{\alpha\in A} x_\alpha \in X$ exists and $T : X \to Y$ is a bounded linear map between normed spaces, then $\sum_{\alpha\in A} Tx_\alpha$ exists in Y and*

$$Ts = T \sum_{\alpha\in A} x_\alpha = \sum_{\alpha\in A} Tx_\alpha.$$

2. *If $s = \sum_{\alpha \in A} x_\alpha$ exists in X then for every $\varepsilon > 0$ there exists $\Gamma_\varepsilon \subset\subset A$ such that $\left\| \sum_{\alpha \in \Lambda} x_\alpha \right\| < \varepsilon$ for all $\Lambda \subset\subset A \setminus \Gamma_\varepsilon$.*
3. *If $s = \sum_{\alpha \in A} x_\alpha$ exists in X, the set $\Gamma := \{\alpha \in A : x_a \neq 0\}$ is at most countable. Moreover if Γ is infinite and $\{\alpha_n\}_{n=1}^\infty$ is an enumeration of Γ, then*
$$s = \sum_{n=1}^\infty x_{\alpha_n} := \lim_{N \to \infty} \sum_{n=1}^N x_{\alpha_n}. \tag{7.4}$$
4. *If we further assume that X is a Banach space and suppose for all $\varepsilon > 0$ there exists $\Gamma_\varepsilon \subset\subset A$ such that $\left\| \sum_{\alpha \in \Lambda} x_\alpha \right\| < \varepsilon$ whenever $\Lambda \subset\subset A \setminus \Gamma_\varepsilon$, then $\sum_{\alpha \in A} x_\alpha$ exists in X.*

Proof.

1. Let Γ_ε be as in Definition 7.15 and $\Lambda \subset\subset A$ such that $\Gamma_\varepsilon \subset \Lambda$. Then
$$\left\| Ts - \sum_{\alpha \in \Lambda} T x_\alpha \right\| \le \|T\| \left\| s - \sum_{\alpha \in \Lambda} x_\alpha \right\| < \|T\| \, \varepsilon$$
which shows that $\sum_{\alpha \in \Lambda} T x_\alpha$ exists and is equal to Ts.
2. Suppose that $s = \sum_{\alpha \in A} x_\alpha$ exists and $\varepsilon > 0$. Let $\Gamma_\varepsilon \subset\subset A$ be as in Definition 7.15. Then for $\Lambda \subset\subset A \setminus \Gamma_\varepsilon$,
$$\begin{aligned} \left\| \sum_{\alpha \in \Lambda} x_\alpha \right\| &= \left\| \sum_{\alpha \in \Gamma_\varepsilon \cup \Lambda} x_\alpha - \sum_{\alpha \in \Gamma_\varepsilon} x_\alpha \right\| \\ &\le \left\| \sum_{\alpha \in \Gamma_\varepsilon \cup \Lambda} x_\alpha - s \right\| + \left\| \sum_{\alpha \in \Gamma_\varepsilon} x_\alpha - s \right\| < 2\varepsilon. \end{aligned}$$
3. If $s = \sum_{\alpha \in A} x_\alpha$ exists in X, for each $n \in \mathbb{N}$ there exists a finite subset $\Gamma_n \subset A$ such that $\left\| \sum_{\alpha \in \Lambda} x_\alpha \right\| < \frac{1}{n}$ for all $\Lambda \subset\subset A \setminus \Gamma_n$. Without loss of generality we may assume $x_\alpha \neq 0$ for all $\alpha \in \Gamma_n$. Let $\Gamma_\infty := \cup_{n=1}^\infty \Gamma_n$ – a countable subset of A. Then for any $\beta \notin \Gamma_\infty$, we have $\{\beta\} \cap \Gamma_n = \emptyset$ and therefore
$$\|x_\beta\| = \left\| \sum_{\alpha \in \{\beta\}} x_\alpha \right\| \le \frac{1}{n} \to 0 \text{ as } n \to \infty.$$
Let $\{\alpha_n\}_{n=1}^\infty$ be an enumeration of Γ and define $\gamma_N := \{\alpha_n : 1 \le n \le N\}$. Since for any $M \in \mathbb{N}$, γ_N will eventually contain Γ_M for N sufficiently large, we have
$$\limsup_{N \to \infty} \left\| s - \sum_{n=1}^N x_{\alpha_n} \right\| \le \frac{1}{M} \to 0 \text{ as } M \to \infty.$$
Therefore Eq. (7.4) holds.

4. For $n \in \mathbb{N}$, let $\Gamma_n \subset\subset A$ such that $\left\|\sum_{\alpha\in\Lambda} x_\alpha\right\| < \frac{1}{n}$ for all $\Lambda \subset\subset A \setminus \Gamma_n$. Define $\gamma_n := \cup_{k=1}^n \Gamma_k \subset A$ and $s_n := \sum_{\alpha\in\gamma_n} x_\alpha$. Then for $m > n$,

$$\|s_m - s_n\| = \left\|\sum_{\alpha\in\gamma_m\setminus\gamma_n} x_\alpha\right\| \le 1/n \to 0 \text{ as } m, n \to \infty.$$

Therefore $\{s_n\}_{n=1}^\infty$ is Cauchy and hence convergent in X, because X is a Banach space. Let $s := \lim_{n\to\infty} s_n$. Then for $\Lambda \subset\subset A$ such that $\gamma_n \subset \Lambda$, we have

$$\left\|s - \sum_{\alpha\in\Lambda} x_\alpha\right\| \le \|s - s_n\| + \left\|\sum_{\alpha\in\Lambda\setminus\gamma_n} x_\alpha\right\| \le \|s - s_n\| + \frac{1}{n}.$$

Since the right side of this equation goes to zero as $n \to \infty$, it follows that $\sum_{\alpha\in A} x_\alpha$ exists and is equal to s.

■

7.4 Inverting Elements in $L(X)$

Definition 7.17. *A linear map $T : X \to Y$ is an **isometry** if $\|Tx\|_Y = \|x\|_X$ for all $x \in X$. T is said to be **invertible** if T is a bijection and T^{-1} is bounded.*

Notation 7.18 *We will write $GL(X,Y)$ for those $T \in L(X,Y)$ which are invertible. If $X = Y$ we simply write $L(X)$ and $GL(X)$ for $L(X,X)$ and $GL(X,X)$ respectively.*

Proposition 7.19. *Suppose X is a Banach space and $\Lambda \in L(X) := L(X,X)$ satisfies $\sum_{n=0}^\infty \|\Lambda^n\| < \infty$. Then $I - \Lambda$ is invertible and*

$$(I-\Lambda)^{-1} = \text{“}\frac{1}{I-\Lambda}\text{”} = \sum_{n=0}^\infty \Lambda^n \text{ and } \left\|(I-\Lambda)^{-1}\right\| \le \sum_{n=0}^\infty \|\Lambda^n\|.$$

In particular if $\|\Lambda\| < 1$ then the above formula holds and

$$\left\|(I-\Lambda)^{-1}\right\| \le \frac{1}{1-\|\Lambda\|}.$$

Proof. Since $L(X)$ is a Banach space and $\sum_{n=0}^\infty \|\Lambda^n\| < \infty$, it follows from Theorem 7.13 that

$$S := \lim_{N\to\infty} S_N := \lim_{N\to\infty} \sum_{n=0}^N \Lambda^n$$

exists in $L(X)$. Moreover, by Lemma 7.11,

$$\begin{aligned}(I-\Lambda)\,S &= (I-\Lambda)\lim_{N\to\infty} S_N = \lim_{N\to\infty}(I-\Lambda)\,S_N \\ &= \lim_{N\to\infty}(I-\Lambda)\sum_{n=0}^{N}\Lambda^n = \lim_{N\to\infty}(I-\Lambda^{N+1}) = I\end{aligned}$$

and similarly $S\,(I-\Lambda)=I$. This shows that $(I-\Lambda)^{-1}$ exists and is equal to S. Moreover, $(I-\Lambda)^{-1}$ is bounded because

$$\left\|(I-\Lambda)^{-1}\right\| = \|S\| \le \sum_{n=0}^{\infty}\|\Lambda^n\|.$$

If we further assume $\|\Lambda\|<1$, then $\|\Lambda^n\|\le\|\Lambda\|^n$ and

$$\sum_{n=0}^{\infty}\|\Lambda^n\| \le \sum_{n=0}^{\infty}\|\Lambda\|^n = \frac{1}{1-\|\Lambda\|} < \infty.$$

■

Corollary 7.20. *Let X and Y be Banach spaces. Then $GL(X,Y)$ is an open (possibly empty) subset of $L(X,Y)$. More specifically, if $A\in GL(X,Y)$ and $B\in L(X,Y)$ satisfies*

$$\|B-A\| < \|A^{-1}\|^{-1} \tag{7.5}$$

then $B\in GL(X,Y)$

$$B^{-1} = \sum_{n=0}^{\infty}\left[I_X - A^{-1}B\right]^n A^{-1} \in L(Y,X), \tag{7.6}$$

$$\left\|B^{-1}\right\| \le \|A^{-1}\|\frac{1}{1-\|A^{-1}\|\,\|A-B\|} \tag{7.7}$$

and

$$\left\|B^{-1}-A^{-1}\right\| \le \frac{\|A^{-1}\|^2\,\|A-B\|}{1-\|A^{-1}\|\,\|A-B\|}. \tag{7.8}$$

In particular the map

$$A\in GL(X,Y)\to A^{-1}\in GL(Y,X) \tag{7.9}$$

is continuous.

Proof. Let A and B be as above, then

$$B = A-(A-B) = A\left[I_X - A^{-1}(A-B))\right] = A(I_X-\Lambda)$$

where $\Lambda : X\to X$ is given by

$$\Lambda := A^{-1}(A - B) = I_X - A^{-1}B.$$

Now

$$\|\Lambda\| = \left\|A^{-1}(A - B))\right\| \le \|A^{-1}\| \, \|A - B\| < \|A^{-1}\| \|A^{-1}\|^{-1} = 1.$$

Therefore $I - \Lambda$ is invertible and hence so is B (being the product of invertible elements) with

$$B^{-1} = (I_X - \Lambda)^{-1} A^{-1} = \left[I_X - A^{-1}(A - B))\right]^{-1} A^{-1}.$$

Taking norms of the previous equation gives

$$\begin{aligned} \left\|B^{-1}\right\| &\le \left\|(I_X - \Lambda)^{-1}\right\| \left\|A^{-1}\right\| \le \|A^{-1}\| \frac{1}{1 - \|\Lambda\|} \\ &\le \frac{\|A^{-1}\|}{1 - \|A^{-1}\| \, \|A - B\|} \end{aligned}$$

which is the bound in Eq. (7.7). The bound in Eq. (7.8) holds because

$$\begin{aligned} \left\|B^{-1} - A^{-1}\right\| &= \left\|B^{-1}(A - B) A^{-1}\right\| \le \left\|B^{-1}\right\| \left\|A^{-1}\right\| \|A - B\| \\ &\le \frac{\|A^{-1}\|^2 \, \|A - B\|}{1 - \|A^{-1}\| \, \|A - B\|}. \end{aligned}$$

■

For an application of these results to linear ordinary differential equations, see Section 8.3.

7.5 Exercises

Exercise 7.4. Let $(X, \|\cdot\|)$ be a normed space over $\mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$). Show the map

$$(\lambda, x, y) \in \mathbb{F} \times X \times X \to x + \lambda y \in X$$

is continuous relative to the norm on $\mathbb{F} \times X \times X$ defined by

$$\|(\lambda, x, y)\|_{\mathbb{F}\times X\times X} := |\lambda| + \|x\| + \|y\| \, .$$

(See Exercise 10.25 for more on the metric associated to this norm.) Also show that $\|\cdot\| : X \to [0, \infty)$ is continuous.

Exercise 7.5. Let $X = \mathbb{N}$ and for $p, q \in [1, \infty)$ let $\|\cdot\|_p$ denote the $\ell^p(\mathbb{N})$ – norm. Show $\|\cdot\|_p$ and $\|\cdot\|_q$ are inequivalent norms for $p \neq q$ by showing

$$\sup_{f \neq 0} \frac{\|f\|_p}{\|f\|_q} = \infty \text{ if } p < q.$$

Exercise 7.6. Suppose that $(X, \|\cdot\|)$ is a normed space and $S \subset X$ is a linear subspace.

1. Show the closure $\bar{S}$ of S is also a linear subspace.
2. Now suppose that X is a Banach space. Show that S with the inherited norm from X is a Banach space iff S is closed.

Exercise 7.7. Folland Problem 5.9. Showing $C^k([0,1])$ is a Banach space.

Exercise 7.8. Suppose that X, Y and Z are Banach spaces and $Q : X \times Y \to Z$ is a bilinear form, i.e. we are assuming $x \in X \to Q(x, y) \in Z$ is linear for each $y \in Y$ and $y \in Y \to Q(x, y) \in Z$ is linear for each $x \in X$. Show Q is continuous relative to the product norm, $\|(x, y)\|_{X \times Y} := \|x\|_X + \|y\|_Y$, on $X \times Y$ iff there is a constant $M < \infty$ such that

$$\|Q(x, y)\|_Z \leq M \|x\|_X \cdot \|y\|_Y \text{ for all } (x, y) \in X \times Y. \tag{7.10}$$

Then apply this result to prove Lemma 7.11.

Exercise 7.9. Let $d : C(\mathbb{R}) \times C(\mathbb{R}) \to [0, \infty)$ be defined by

$$d(f, g) = \sum_{n=1}^{\infty} 2^{-n} \frac{\|f - g\|_n}{1 + \|f - g\|_n},$$

where $\|f\|_n := \sup\{|f(x)| : |x| \leq n\} = \max\{|f(x)| : |x| \leq n\}$.

1. Show that d is a metric on $C(\mathbb{R})$.
2. Show that a sequence $\{f_n\}_{n=1}^{\infty} \subset C(\mathbb{R})$ converges to $f \in C(\mathbb{R})$ as $n \to \infty$ iff f_n converges to f uniformly on bounded subsets of $\mathbb{R}$.
3. Show that $(C(\mathbb{R}), d)$ is a complete metric space.

Exercise 7.10. Let $X = C([0, 1], \mathbb{R})$ and for $f \in X$, let

$$\|f\|_1 := \int_0^1 |f(t)|\, dt.$$

Show that $(X, \|\cdot\|_1)$ is normed space and show by example that this space is **not** complete. Hint: For the last assertion find a sequence of $\{f_n\}_{n=1}^{\infty} \subset X$ which is "trying" to converge to the function $f = 1_{[\frac{1}{2}, 1]} \notin X$.

Exercise 7.11. Let $(X, \|\cdot\|_1)$ be the normed space in Exercise 7.10. Compute the closure of A when

1. $A = \{f \in X : f(1/2) = 0\}$.
2. $A = \left\{f \in X : \sup_{t \in [0,1]} f(t) \leq 5\right\}$.
3. $A = \left\{f \in X : \int_0^{1/2} f(t)\, dt = 0\right\}$.

Exercise 7.12. Suppose $\{x_\alpha \in X : \alpha \in A\}$ is a given collection of vectors in a Banach space X. Show $\sum_{\alpha \in A} x_\alpha$ exists in X and

$$\left\| \sum_{\alpha \in A} x_\alpha \right\| \le \sum_{\alpha \in A} \|x_\alpha\|$$

if $\sum_{\alpha \in A} \|x_\alpha\| < \infty$. That is to say **"absolute convergence"** implies convergence in a Banach space.

Exercise 7.13. Suppose X is a Banach space and $\{f_n : n \in \mathbb{N}\}$ is a sequence in X such that $\lim_{n\to\infty} f_n = f \in X$. Show $s_N := \frac{1}{N} \sum_{n=1}^{N} f_n$ for $N \in \mathbb{N}$ is still a convergent sequence and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f_n = \lim_{N\to\infty} s_N = f.$$

Exercise 7.14 (Dominated Convergence Theorem Again). Let X be a Banach space, A be a set and suppose $f_n : A \to X$ is a sequence of functions such that $f(\alpha) := \lim_{n\to\infty} f_n(\alpha)$ exists for all $\alpha \in A$. Further assume there exists a summable function $g : A \to [0, \infty)$ such that $\|f_n(\alpha)\| \le g(\alpha)$ for all $\alpha \in A$. Show $\sum_{\alpha \in A} f(\alpha)$ exists in X and

$$\lim_{n\to\infty} \sum_{\alpha \in A} f_n(\alpha) = \sum_{\alpha \in A} f(\alpha).$$

8

The Riemann Integral

In this Chapter, the Riemann integral for Banach space valued functions is defined and developed. Our exposition will be brief, since the Lebesgue integral and the Bochner Lebesgue integral will subsume the content of this chapter. In Definition 11.1 below, we will give a general notion of a compact subset of a "topological" space. However, by Corollary 11.9 below, when we are working with subsets of $\mathbb{R}^d$ this definition is equivalent to the following definition.

Definition 8.1. *A subset $A \subset \mathbb{R}^d$ is said to be* ***compact*** *if A is closed and* ***bounded.***

Theorem 8.2. *Suppose that $K \subset \mathbb{R}^d$ is a compact set and $f \in C(K, X)$. Then*

1. *Every sequence $\{u_n\}_{n=1}^{\infty} \subset K$ has a convergent subsequence.*
2. *The function f is uniformly continuous on K, namely for every $\varepsilon > 0$ there exists a $\delta > 0$ only depending on ε such that $\|f(u) - f(v)\| < \varepsilon$ whenever $u, v \in K$ and $|u - v| < \delta$ where $|\cdot|$ is the standard Euclidean norm on $\mathbb{R}^d$.*

Proof.

1. (This is a special case of Theorem 11.7 and Corollary 11.9 below.) Since K is bounded, $K \subset [-R, R]^d$ for some sufficiently large d. Let t_n be the first component of u_n so that $t_n \in [-R, R]$ for all n. Let $J_1 = [0, R]$ if $t_n \in J_1$ for infinitely many n otherwise let $J_1 = [-R, 0]$. Similarly split J_1 in half and let $J_2 \subset J_1$ be one of the halves such that $t_n \in J_2$ for infinitely many n. Continue this way inductively to find a nested sequence of intervals $J_1 \supset J_2 \supset J_3 \supset J_4 \supset \dots$ such that the length of J_k is $2^{-(k-1)}R$ and for each k, $t_n \in J_k$ for infinitely many n. We may now choose a subsequence, $\{n_k\}_{k=1}^{\infty}$ of $\{n\}_{n=1}^{\infty}$ such that $\tau_k := t_{n_k} \in J_k$ for all k. The sequence $\{\tau_k\}_{k=1}^{\infty}$ is Cauchy and hence convergent. Thus by replacing $\{u_n\}_{n=1}^{\infty}$ by a subsequence if necessary we may assume the first component of $\{u_n\}_{n=1}^{\infty}$ is

convergent. Repeating this argument for the second, then the third and all the way through the d^{th} – components of $\{u_n\}_{n=1}^{\infty}$, we may, by passing to further subsequences, assume all of the components of u_n are convergent. But this implies $\lim u_n = u$ exists and since K is closed, $u \in K$.

2. (This is a special case of Exercise 11.6 below.) If f were not uniformly continuous on K, there would exists an $\varepsilon > 0$ and sequences $\{u_n\}_{n=1}^{\infty}$ and $\{v_n\}_{n=1}^{\infty}$ in K such that

$$\|f(u_n) - f(v_n)\| \geq \varepsilon \text{ while } \lim_{n\to\infty} |u_n - v_n| = 0.$$

By passing to subsequences if necessary we may assume that $\lim_{n\to\infty} u_n$ and $\lim_{n\to\infty} v_n$ exists. Since $\lim_{n\to\infty} |u_n - v_n| = 0$, we must have

$$\lim_{n\to\infty} u_n = u = \lim_{n\to\infty} v_n$$

for some $u \in K$. Since f is continuous, vector addition is continuous and the norm is continuous, we may now conclude that

$$\varepsilon \leq \lim_{n\to\infty} \|f(u_n) - f(v_n)\| = \|f(u) - f(u)\| = 0$$

which is a contradiction.

■

For the remainder of the chapter, let $[a, b]$ be a fixed compact interval and X be a Banach space. The collection $\mathcal{S} = \mathcal{S}([a, b], X)$ of **step functions,** $f : [a, b] \to X$, consists of those functions f which may be written in the form

$$f(t) = x_0 1_{[a,t_1]}(t) + \sum_{i=1}^{n-1} x_i 1_{(t_i,t_{i+1}]}(t), \tag{8.1}$$

where $\pi := \{a = t_0 < t_1 < \cdots < t_n = b\}$ is a partition of $[a, b]$ and $x_i \in X$. For f as in Eq. (8.1), let

$$I(f) := \sum_{i=0}^{n-1} (t_{i+1} - t_i) x_i \in X. \tag{8.2}$$

Exercise 8.1. Show that $I(f)$ is well defined, independent of how f is represented as a step function. (**Hint:** show that adding a point to a partition π of $[a, b]$ does not change the right side of Eq. (8.2).) Also verify that $I : \mathcal{S} \to X$ is a linear operator.

Notation 8.3 *Let $\bar{\mathcal{S}}$ denote the closure of $\mathcal{S}$ inside the Banach space, $\ell^\infty([a, b], X)$ as defined in Remark 7.6.*

The following simple "Bounded Linear Transformation" theorem will often be used in the sequel to define linear transformations.

Theorem 8.4 (B. L. T. Theorem). *Suppose that Z is a normed space, X is a Banach space, and $\mathcal{S} \subset Z$ is a dense linear subspace of Z. If $T : \mathcal{S} \to X$ is a bounded linear transformation (i.e. there exists $C < \infty$ such that $\|Tz\| \leq C\|z\|$ for all $z \in \mathcal{S}$), then T has a unique extension to an element $\bar{T} \in L(Z, X)$ and this extension still satisfies*

$$\left\|\bar{T}z\right\| \leq C\|z\| \text{ for all } z \in \bar{\mathcal{S}}.$$

Exercise 8.2. Prove Theorem 8.4.

Proposition 8.5 (Riemann Integral). *The linear function $I : \mathcal{S} \to X$ extends uniquely to a continuous linear operator $\bar{I}$ from $\bar{\mathcal{S}}$ to X and this operator satisfies,*

$$\|\bar{I}(f)\| \leq (b-a)\|f\|_\infty \text{ for all } f \in \bar{\mathcal{S}}. \tag{8.3}$$

Furthermore, $C([a,b], X) \subset \bar{\mathcal{S}} \subset \ell^\infty([a,b], X)$ and for $f \in$, $\bar{I}(f)$ may be computed as

$$\bar{I}(f) = \lim_{|\pi| \to 0} \sum_{i=0}^{n-1} f(c_i^\pi)(t_{i+1} - t_i) \tag{8.4}$$

where $\pi := \{a = t_0 < t_1 < \cdots < t_n = b\}$ denotes a partition of $[a,b]$, $|\pi| = \max\{|t_{i+1} - t_i| : i = 0, \ldots, n-1\}$ is the mesh size of π and c_i^π may be chosen arbitrarily inside $[t_i, t_{i+1}]$. See Figure 8.1.

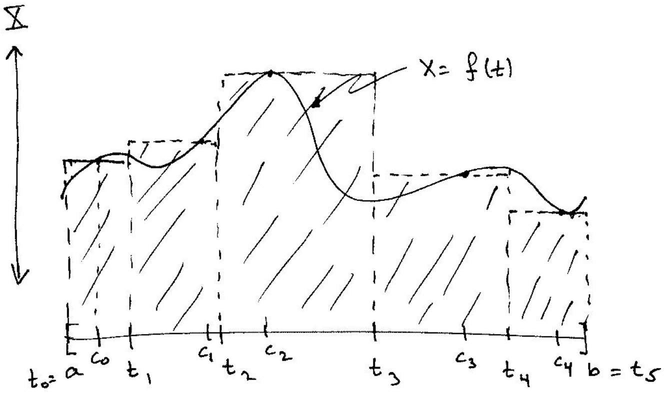


Fig. 8.1. The usual picture associated to the Riemann integral.

Proof. Taking the norm of Eq. (8.2) and using the triangle inequality shows,

$$\|I(f)\| \leq \sum_{i=0}^{n-1}(t_{i+1} - t_i)\|x_i\| \leq \sum_{i=0}^{n-1}(t_{i+1} - t_i)\|f\|_\infty \leq (b-a)\|f\|_\infty. \tag{8.5}$$

The existence of $\bar{I}$ satisfying Eq. (8.3) is a consequence of Theorem 8.4. Given $f \in C([a,b],X)$, $\pi := \{a = t_0 < t_1 < \cdots < t_n = b\}$ a partition of $[a,b]$, and $c_i^\pi \in [t_i, t_{i+1}]$ for $i = 0,1,2\ldots,n-1$, let $f_\pi \in \mathcal{S}$ be defined by

$$f_\pi(t) := f(c_0)_0 1_{[t_0,t_1]}(t) + \sum_{i=1}^{n-1} f(c_i^\pi) 1_{(t_i,t_{i+1}]}(t).$$

Then by the uniform continuity of f on $[a,b]$ (Theorem 8.2), $\lim_{|\pi|\to 0} \|f - f_\pi\|_\infty = 0$ and therefore $f \in \bar{\mathcal{S}}$. Moreover,

$$I(f) = \lim_{|\pi|\to 0} I(f_\pi) = \lim_{|\pi|\to 0} \sum_{i=0}^{n-1} f(c_i^\pi)(t_{i+1} - t_i)$$

which proves Eq. (8.4). ■

If $f_n \in \mathcal{S}$ and $f \in \bar{\mathcal{S}}$ such that $\lim_{n\to\infty} \|f - f_n\|_\infty = 0$, then for $a \le \alpha < \beta \le b$, then $1_{(\alpha,\beta]} f_n \in \mathcal{S}$ and $\lim_{n\to\infty} \left\|1_{(\alpha,\beta]} f - 1_{(\alpha,\beta]} f_n\right\|_\infty = 0$. This shows $1_{(\alpha,\beta]} f \in \bar{\mathcal{S}}$ whenever $f \in \bar{\mathcal{S}}$.

Notation 8.6 *For $f \in \bar{\mathcal{S}}$ and $a \le \alpha \le \beta \le b$ we will write denote $\bar{I}(1_{(\alpha,\beta]} f)$ by $\int_\alpha^\beta f(t)\,dt$ or $\int_{(\alpha,\beta]} f(t)dt$. Also following the usual convention, if $a \le \beta \le \alpha \le b$, we will let*

$$\int_\alpha^\beta f(t)\,dt = -\bar{I}(1_{(\beta,\alpha]} f) = -\int_\beta^\alpha f(t)\,dt.$$

The next Lemma, whose proof is left to the reader contains some of the many familiar properties of the Riemann integral.

Lemma 8.7. *For $f \in \bar{\mathcal{S}}([a,b],X)$ and $\alpha,\beta,\gamma \in [a,b]$, the Riemann integral satisfies:*

1. $\left\|\int_\alpha^\beta f(t)\,dt\right\|_X \le (\beta - \alpha) \sup\{\|f(t)\| : \alpha \le t \le \beta\}$.
2. $\int_\alpha^\gamma f(t)\,dt = \int_\alpha^\beta f(t)\,dt + \int_\beta^\gamma f(t)\,dt$.
3. *The function $G(t) := \int_a^t f(\tau)d\tau$ is continuous on $[a,b]$.*
4. *If Y is another Banach space and $T \in L(X,Y)$, then $Tf \in \bar{\mathcal{S}}([a,b],Y)$ and*
$$T\left(\int_\alpha^\beta f(t)\ dt\right) = \int_\alpha^\beta Tf(t)\ dt.$$
5. *The function $t \to \|f(t)\|_X$ is in $\bar{\mathcal{S}}([a,b],\mathbb{R})$ and*
$$\left\|\int_a^b f(t)\ dt\right\|_X \le \int_a^b \|f(t)\|_X\ dt.$$

6. *If $f, g \in \bar{S}([a,b],\mathbb{R})$ and $f \le g$, then*

$$\int_a^b f(t)\ dt \le \int_a^b g(t)\ dt.$$

Exercise 8.3. Prove Lemma 8.7.

8.1 The Fundamental Theorem of Calculus

Our next goal is to show that our Riemann integral interacts well with differentiation, namely the fundamental theorem of calculus holds. Before doing this we will need a couple of basic definitions and results of differential calculus, more details and the next few results below will be done in greater detail in Chapter 16.

Definition 8.8. *Let $(a,b) \subset \mathbb{R}$. A function $f : (a,b) \to X$ is differentiable at $t \in (a,b)$ iff*

$$L := \lim_{h\to 0} \left(h^{-1}\left[f(t+h) - f(t)\right]\right) = \lim_{h\to 0} \text{“}\frac{f(t+h) - f(t)}{h}\text{”}$$

exists in X. The limit L, if it exists, will be denoted by $\dot{f}(t)$ or $\frac{df}{dt}(t)$. We also say that $f \in C^1((a,b) \to X)$ if f is differentiable at all points $t \in (a,b)$ and $\dot{f} \in C((a,b) \to X)$.

As for the case of real valued functions, the derivative operator $\frac{d}{dt}$ is easily seen to be linear. The next two results have proves very similar to their real valued function analogues.

Lemma 8.9 (Product Rules). *Suppose that $t \to U(t) \in L(X)$, $t \to V(t) \in L(X)$ and $t \to x(t) \in X$ are differentiable at $t = t_0$, then*

1. *$\frac{d}{dt}|_{t_0}\left[U(t)\,x(t)\right] \in X$ exists and*

$$\frac{d}{dt}|_{t_0}\left[U(t)\,x(t)\right] = \left[\dot{U}(t_0)\,x(t_0) + U(t_0)\,\dot{x}(t_0)\right]$$

and

2. *$\frac{d}{dt}|_{t_0}\left[U(t)\,V(t)\right] \in L(X)$ exists and*

$$\frac{d}{dt}|_{t_0}\left[U(t)\,V(t)\right] = \left[\dot{U}(t_0)\,V(t_0) + U(t_0)\,\dot{V}(t_0)\right].$$

3. *If $U(t_0)$ is invertible, then $t \to U(t)^{-1}$ is differentiable at $t = t_0$ and*

$$\frac{d}{dt}|_{t_0}U(t)^{-1} = -U(t_0)^{-1}\,\dot{U}(t_0)\,U(t_0)^{-1}. \tag{8.6}$$

Proof. The reader is asked to supply the proof of the first two items in Exercise 8.10. Before proving item 3., let us assume that $U(t)^{-1}$ is differentiable, then using the product rule we would learn

$$0 = \frac{d}{dt}|_{t_0} I = \frac{d}{dt}|_{t_0} \left[U(t)^{-1} U(t) \right] = \left[\frac{d}{dt}|_{t_0} U(t)^{-1} \right] U(t_0) + U(t_0)^{-1} \dot{U}(t_0).$$

Solving this equation for $\frac{d}{dt}|_{t_0} U(t)^{-1}$ gives the formula in Eq. (8.6). The problem with this argument is that we have not yet shown $t \to U(t)^{-1}$ is invertible at t_0. Here is the formal proof. Since $U(t)$ is differentiable at t_0, $U(t) \to U(t_0)$ as $t \to t_0$ and by Corollary 7.20, $U(t_0 + h)$ is invertible for h near 0 and

$$U(t_0 + h)^{-1} \to U(t_0)^{-1} \text{ as } h \to 0.$$

Therefore, using Lemma 7.11, we may let $h \to 0$ in the identity,

$$\frac{U(t_0 + h)^{-1} - U(t_0)^{-1}}{h} = U(t_0 + h)^{-1} \left(\frac{U(t_0) - U(t_0 + h)}{h} \right) U(t_0)^{-1},$$

to learn

$$\lim_{h \to 0} \frac{U(t_0 + h)^{-1} - U(t_0)^{-1}}{h} = -U(t_0)^{-1} \dot{U}(t_0) U(t_0)^{-1}.$$

■

Proposition 8.10 (Chain Rule). *Suppose $s \to x(s) \in X$ is differentiable at $s = s_0$ and $t \to T(t) \in \mathbb{R}$ is differentiable at $t = t_0$ and $T(t_0) = s_0$, then $t \to x(T(t))$ is differentiable at t_0 and*

$$\frac{d}{dt}|_{t_0} x(T(t)) = x'(T(t_0)) T'(t_0).$$

The proof of the chain rule is essentially the same as the real valued function case, see Exercise 8.11.

Proposition 8.11. *Suppose that $f : [a, b] \to X$ is a continuous function such that $\dot{f}(t)$ exists and is equal to zero for $t \in (a, b)$. Then f is constant.*

Proof. Let $\varepsilon > 0$ and $\alpha \in (a, b)$ be given. (We will later let $\varepsilon \downarrow 0$.) By the definition of the derivative, for all $\tau \in (a, b)$ there exists $\delta_\tau > 0$ such that

$$\|f(t) - f(\tau)\| = \left\| f(t) - f(\tau) - \dot{f}(\tau)(t - \tau) \right\| \leq \varepsilon |t - \tau| \text{ if } |t - \tau| < \delta_\tau. \tag{8.7}$$

Let

$$A = \{t \in [\alpha, b] : \|f(t) - f(\alpha)\| \leq \varepsilon(t - \alpha)\} \tag{8.8}$$

and t_0 be the least upper bound for A. We will now use a standard argument which is referred to as **continuous induction** to show $t_0 = b$. Eq. (8.7) with $\tau = \alpha$ shows $t_0 > \alpha$ and a simple continuity argument shows $t_0 \in A$, i.e.

$$\|f(t_0) - f(\alpha)\| \le \varepsilon(t_0 - \alpha). \tag{8.9}$$

For the sake of contradiction, suppose that $t_0 < b$. By Eqs. (8.7) and (8.9),

$$\begin{aligned}\|f(t) - f(\alpha)\| &\le \|f(t) - f(t_0)\| + \|f(t_0) - f(\alpha)\| \\ &\le \varepsilon(t_0 - \alpha) + \varepsilon(t - t_0) = \varepsilon(t - \alpha)\end{aligned}$$

for $0 \le t - t_0 < \delta_{t_0}$ which violates the definition of t_0 being an upper bound. Thus we have shown $b \in A$ and hence

$$\|f(b) - f(\alpha)\| \le \varepsilon(b - \alpha).$$

Since $\varepsilon > 0$ was arbitrary we may let $\varepsilon \downarrow 0$ in the last equation to conclude $f(b) = f(\alpha)$. Since $\alpha \in (a, b)$ was arbitrary it follows that $f(b) = f(\alpha)$ for all $\alpha \in (a, b]$ and then by continuity for all $\alpha \in [a, b]$, i.e. f is constant. ■

Remark 8.12. The usual real variable proof of Proposition 8.11 makes use Rolle's theorem which in turn uses the extreme value theorem. This latter theorem is not available to vector valued functions. However with the aid of the Hahn Banach Theorem **??** below and Lemma 8.7, it is possible to reduce the proof of Proposition 8.11 and the proof of the Fundamental Theorem of Calculus 8.13 to the real valued case, see Exercise **??**.

Theorem 8.13 (Fundamental Theorem of Calculus). *Suppose that $f \in C([a, b], X)$, Then*

1. $\frac{d}{dt}\int_a^t f(\tau)\,d\tau = f(t)$ *for all* $t \in (a, b)$.
2. *Now assume that $F \in C([a, b], X)$, F is continuously differentiable on (a, b) (i.e. $\dot{F}(t)$ exists and is continuous for $t \in (a, b)$) and $\dot{F}$ extends to a continuous function on $[a, b]$ which is still denoted by $\dot{F}$. Then*

$$\int_a^b \dot{F}(t)\,dt = F(b) - F(a).$$

Proof. Let $h > 0$ be a small number and consider

$$\begin{aligned}\left\|\int_a^{t+h} f(\tau)d\tau - \int_a^t f(\tau)d\tau - f(t)h\right\| &= \left\|\int_t^{t+h} (f(\tau) - f(t))\,d\tau\right\| \\ &\le \int_t^{t+h} \|(f(\tau) - f(t))\|\,d\tau \le h\varepsilon(h),\end{aligned}$$

where $\varepsilon(h) := \max_{\tau\in[t,t+h]} \|(f(\tau) - f(t))\|$. Combining this with a similar computation when $h < 0$ shows, for all $h \in \mathbb{R}$ sufficiently small, that

$$\|\int_a^{t+h} f(\tau)d\tau - \int_a^t f(\tau)d\tau - f(t)h\| \le |h|\varepsilon(h),$$

where now $\varepsilon(h) := \max_{\tau\in[t-|h|,t+|h|]} \|(f(\tau) - f(t))\|$. By continuity of f at t, $\varepsilon(h) \to 0$ and hence $\frac{d}{dt}\int_a^t f(\tau)\,d\tau$ exists and is equal to $f(t)$. For the second item, set $G(t) := \int_a^t \dot{F}(\tau)\,d\tau - F(t)$. Then G is continuous by Lemma 8.7 and $\dot{G}(t) = 0$ for all $t \in (a,b)$ by item 1. An application of Proposition 8.11 shows G is a constant and in particular $G(b) = G(a)$, i.e. $\int_a^b \dot{F}(\tau)\,d\tau - F(b) = -F(a)$. ■

Corollary 8.14 (Mean Value Inequality). *Suppose that $f : [a,b] \to X$ is a continuous function such that $\dot{f}(t)$ exists for $t \in (a,b)$ and $\dot{f}$ extends to a continuous function on $[a,b]$. Then*

$$\|f(b) - f(a)\| \le \int_a^b \|\dot{f}(t)\|dt \le (b-a)\cdot \left\|\dot{f}\right\|_\infty. \tag{8.10}$$

Proof. By the fundamental theorem of calculus, $f(b) - f(a) = \int_a^b \dot{f}(t)dt$ and then by Lemma 8.7,

$$\begin{aligned}\|f(b) - f(a)\| &= \left\|\int_a^b \dot{f}(t)dt\right\| \le \int_a^b \|\dot{f}(t)\|dt \\ &\le \int_a^b \left\|\dot{f}\right\|_\infty dt = (b-a)\cdot\left\|\dot{f}\right\|_\infty.\end{aligned}$$

■

Corollary 8.15 (Change of Variable Formula). *Suppose that $f \in C([a,b],X)$ and $T : [c,d] \to (a,b)$ is a continuous function such that $T(s)$ is continuously differentiable for $s \in (c,d)$ and $T'(s)$ extends to a continuous function on $[c,d]$. Then*

$$\int_c^d f(T(s))\,T'(s)\,ds = \int_{T(c)}^{T(d)} f(t)\,dt.$$

Proof. For $s \in (a,b)$ define $F(t) := \int_{T(c)}^t f(\tau)\,d\tau$. Then $F \in C^1((a,b),X)$ and by the fundamental theorem of calculus and the chain rule,

$$\frac{d}{ds}F(T(s)) = F'(T(s))\,T'(s) = f(T(s))\,T'(s).$$

Integrating this equation on $s \in [c,d]$ and using the chain rule again gives

$$\int_c^d f(T(s))\,T'(s)\,ds = F(T(d)) - F(T(c)) = \int_{T(c)}^{T(d)} f(t)\,dt.$$

■

8.2 Integral Operators as Examples of Bounded Operators

In the examples to follow all integrals are the standard Riemann integrals and we will make use of the following notation.

Notation 8.16 *Given an open set $U \subset \mathbb{R}^d$, let $C_c(U)$ denote the collection of real valued continuous functions f on U such that*

$$\operatorname{supp}(f) := \overline{\{x \in U : f(x) \neq 0\}}$$

is a compact subset of U.

Example 8.17. Suppose that $K : [0,1] \times [0,1] \to \mathbb{C}$ is a continuous function. For $f \in C([0,1])$, let

$$Tf(x) = \int_0^1 K(x,y) f(y) dy.$$

Since

$$\begin{aligned} |Tf(x) - Tf(z)| &\leq \int_0^1 |K(x,y) - K(z,y)| \, |f(y)| \, dy \\ &\leq \|f\|_\infty \max_y |K(x,y) - K(z,y)| \end{aligned} \tag{8.11}$$

and the latter expression tends to 0 as $x \to z$ by uniform continuity of K. Therefore $Tf \in C([0,1])$ and by the linearity of the Riemann integral, $T : C([0,1]) \to C([0,1])$ is a linear map. Moreover,

$$|Tf(x)| \leq \int_0^1 |K(x,y)| \, |f(y)| \, dy \leq \int_0^1 |K(x,y)| \, dy \cdot \|f\|_\infty \leq A \|f\|_\infty$$

where

$$A := \sup_{x \in [0,1]} \int_0^1 |K(x,y)| \, dy < \infty. \tag{8.12}$$

This shows $\|T\| \leq A < \infty$ and therefore T is bounded. We may in fact show $\|T\| = A$. To do this let $x_0 \in [0,1]$ be such that

$$\sup_{x \in [0,1]} \int_0^1 |K(x,y)| \, dy = \int_0^1 |K(x_0,y)| \, dy.$$

Such an x_0 can be found since, using a similar argument to that in Eq. (8.11), $x \to \int_0^1 |K(x,y)| \, dy$ is continuous. Given $\varepsilon > 0$, let

$$f_\varepsilon(y) := \frac{\overline{K(x_0,y)}}{\sqrt{\varepsilon + |K(x_0,y)|^2}}$$

and notice that $\lim_{\varepsilon\downarrow 0}\|f_\varepsilon\|_\infty = 1$ and

$$\|Tf_\varepsilon\|_\infty \geq |Tf_\varepsilon(x_0)| = Tf_\varepsilon(x_0) = \int_0^1 \frac{|K(x_0,y)|^2}{\sqrt{\varepsilon + |K(x_0,y)|^2}} dy.$$

Therefore,

$$\begin{aligned}\|T\| &\geq \lim_{\varepsilon\downarrow 0} \frac{1}{\|f_\varepsilon\|_\infty} \int_0^1 \frac{|K(x_0,y)|^2}{\sqrt{\varepsilon + |K(x_0,y)|^2}} dy \\ &= \lim_{\varepsilon\downarrow 0} \int_0^1 \frac{|K(x_0,y)|^2}{\sqrt{\varepsilon + |K(x_0,y)|^2}} dy = A\end{aligned}$$

since

$$\begin{aligned}0 &\leq |K(x_0,y)| - \frac{|K(x_0,y)|^2}{\sqrt{\varepsilon + |K(x_0,y)|^2}} \\ &= \frac{|K(x_0,y)|}{\sqrt{\varepsilon + |K(x_0,y)|^2}} \left[\sqrt{\varepsilon + |K(x_0,y)|^2} - |K(x_0,y)|\right] \\ &\leq \sqrt{\varepsilon + |K(x_0,y)|^2} - |K(x_0,y)|\end{aligned}$$

and the latter expression tends to zero uniformly in y as $\varepsilon \downarrow 0$.

We may also consider other norms on $C([0,1])$. Let (for now) $L^1([0,1])$ denote $C([0,1])$ with the norm

$$\|f\|_1 = \int_0^1 |f(x)|\, dx,$$

then $T : L^1([0,1], dm) \to C([0,1])$ is bounded as well. Indeed, let $M = \sup\{|K(x,y)| : x, y \in [0,1]\}$, then

$$|(Tf)(x)| \leq \int_0^1 |K(x,y)f(y)|\, dy \leq M\|f\|_1$$

which shows $\|Tf\|_\infty \leq M\|f\|_1$ and hence,

$$\|T\|_{L^1\to C} \leq \max\{|K(x,y)| : x, y \in [0,1]\} < \infty.$$

We can in fact show that $\|T\| = M$ as follows. Let $(x_0, y_0) \in [0,1]^2$ satisfying $|K(x_0,y_0)| = M$. Then given $\varepsilon > 0$, there exists a neighborhood $U = I \times J$ of (x_0,y_0) such that $|K(x,y) - K(x_0,y_0)| < \varepsilon$ for all $(x,y) \in U$. Let $f \in C_c(I, [0,\infty))$ such that $\int_0^1 f(x)dx = 1$. Choose $\alpha \in \mathbb{C}$ such that $|\alpha| = 1$ and $\alpha K(x_0,y_0) = M$, then

$$
\begin{aligned}
|(T\alpha f)(x_0)| &= \left|\int_0^1 K(x_0,y)\alpha f(y)dy\right| = \left|\int_I K(x_0,y)\alpha f(y)dy\right| \\
&\geq \operatorname{Re}\int_I \alpha K(x_0,y)f(y)dy \\
&\geq \int_I (M-\varepsilon)\, f(y)dy = (M-\varepsilon)\left\|\alpha f\right\|_{L^1}
\end{aligned}
$$

and hence

$$\left\|T\alpha f\right\|_C \geq (M-\varepsilon)\left\|\alpha f\right\|_{L^1}$$

showing that $\|T\| \geq M - \varepsilon$. Since $\varepsilon > 0$ is arbitrary, we learn that $\|T\| \geq M$ and hence $\|T\| = M$.

One may also view T as a map from $T : C([0,1]) \to L^1([0,1])$ in which case one may show

$$\left\|T\right\|_{L^1\to C} \leq \int_0^1 \max_y |K(x,y)|\, dx < \infty.$$

8.3 Linear Ordinary Differential Equations

Let X be a Banach space, $J = (a,b) \subset \mathbb{R}$ be an open interval with $0 \in J$, $h \in C(J \to X)$ and $A \in C(J \to L(X))$. In this section we are going to consider the ordinary differential equation,

$$\dot{y}(t) = A(t)y(t) + h\,(t) \text{ where } y(0) = x \in X, \tag{8.13}$$

where y is an unknown function in $C^1(J \to X)$. This equation may be written in its equivalent (as the reader should verify) integral form, namely we are looking for $y \in C(J,X)$ such that

$$y(t) = x + \int_0^t h\,(\tau)\, d\tau + \int_0^t A(\tau)y(\tau)d\tau. \tag{8.14}$$

In what follows, we will abuse notation and use $\|\cdot\|$ to denote the operator norm on $L\,(X)$ associated to then norm, $\|\cdot\|$, on X and let $\|\phi\|_\infty := \max_{t\in J}\|\phi(t)\|$ for $\phi \in BC(J,X)$ or $BC(J, L\,(X))$.

Notation 8.18 *For $t \in \mathbb{R}$ and $n \in \mathbb{N}$, let*

$$\Delta_n(t) = \begin{cases} \{(\tau_1,\dots,\tau_n) \in \mathbb{R}^n : 0 \leq \tau_1 \leq \cdots \leq \tau_n \leq t\} \text{ if } t \geq 0 \\ \{(\tau_1,\dots,\tau_n) \in \mathbb{R}^n : t \leq \tau_n \leq \cdots \leq \tau_1 \leq 0\} \text{ if } t \leq 0 \end{cases}$$

and also write $d\tau = d\tau_1 \dots d\tau_n$ and

$$\int_{\Delta_n(t)} f(\tau_1,\dots\tau_n)d\tau := (-1)^{n\cdot 1_{t<0}} \int_0^t d\tau_n \int_0^{\tau_n} d\tau_{n-1} \dots \int_0^{\tau_2} d\tau_1 f(\tau_1,\dots\tau_n).$$

Lemma 8.19. *Suppose that* $\psi \in C(\mathbb{R}, \mathbb{R})$, *then*

$$(-1)^{n \cdot 1_{t<0}} \int_{\Delta_n(t)} \psi(\tau_1) \dots \psi(\tau_n) d\tau = \frac{1}{n!} \left(\int_0^t \psi(\tau) d\tau \right)^n. \tag{8.15}$$

Proof. Let $\Psi(t) := \int_0^t \psi(\tau) d\tau$. The proof will go by induction on n. The case $n = 1$ is easily verified since

$$(-1)^{1 \cdot 1_{t<0}} \int_{\Delta_1(t)} \psi(\tau_1) d\tau_1 = \int_0^t \psi(\tau) d\tau = \Psi(t).$$

Now assume the truth of Eq. (8.15) for $n-1$ for some $n \geq 2$, then

$$\begin{aligned}
&(-1)^{n \cdot 1_{t<0}} \int_{\Delta_n(t)} \psi(\tau_1) \dots \psi(\tau_n) d\tau \\
&= \int_0^t d\tau_n \int_0^{\tau_n} d\tau_{n-1} \dots \int_0^{\tau_2} d\tau_1 \psi(\tau_1) \dots \psi(\tau_n) \\
&= \int_0^t d\tau_n \frac{\Psi^{n-1}(\tau_n)}{(n-1)!} \psi(\tau_n) = \int_0^t d\tau_n \frac{\Psi^{n-1}(\tau_n)}{(n-1)!} \dot{\Psi}(\tau_n) \\
&= \int_0^{\Psi(t)} \frac{u^{n-1}}{(n-1)!} du = \frac{\Psi^n(t)}{n!},
\end{aligned}$$

wherein we made the change of variables, $u = \Psi(\tau_n)$, in the second to last equality. ∎

Remark 8.20. Eq. (8.15) is equivalent to

$$\int_{\Delta_n(t)} \psi(\tau_1) \dots \psi(\tau_n) d\tau = \frac{1}{n!} \left(\int_{\Delta_1(t)} \psi(\tau) d\tau \right)^n$$

and another way to understand this equality is to view $\int_{\Delta_n(t)} \psi(\tau_1) \dots \psi(\tau_n) d\tau$ as a multiple integral (see Chapter 20 below) rather than an iterated integral. Indeed, taking $t > 0$ for simplicity and letting S_n be the permutation group on $\{1, 2, \dots, n\}$ we have

$$[0, t]^n = \cup_{\sigma \in S_n} \{(\tau_1, \dots, \tau_n) \in \mathbb{R}^n : 0 \leq \tau_{\sigma 1} \leq \dots \leq \tau_{\sigma n} \leq t\}$$

with the union being "essentially" disjoint. Therefore, making a change of variables and using the fact that $\psi(\tau_1) \dots \psi(\tau_n)$ is invariant under permutations, we find

$$\begin{aligned}
\left(\int_0^t \psi(\tau)d\tau\right)^n &= \int_{[0,t]^n} \psi(\tau_1)\dots\psi(\tau_n)d\tau \\
&= \sum_{\sigma\in S_n} \int_{\{(\tau_1,\dots,\tau_n)\in\mathbb{R}^n:0\le\tau_{\sigma 1}\le\dots\le\tau_{\sigma n}\le t\}} \psi(\tau_1)\dots\psi(\tau_n)d\tau \\
&= \sum_{\sigma\in S_n} \int_{\{(s_1,\dots,s_n)\in\mathbb{R}^n:0\le s_1\le\dots\le s_n\le t\}} \psi(s_{\sigma^{-1}1})\dots\psi(s_{\sigma^{-1}n})d\mathbf{s} \\
&= \sum_{\sigma\in S_n} \int_{\{(s_1,\dots,s_n)\in\mathbb{R}^n:0\le s_1\le\dots\le s_n\le t\}} \psi(s_1)\dots\psi(s_n)d\mathbf{s} \\
&= n!\int_{\Delta_n(t)} \psi(\tau_1)\dots\psi(\tau_n)d\tau.
\end{aligned}$$

Theorem 8.21. *Let $\phi \in BC(J,X)$, then the integral equation*

$$y(t) = \phi(t) + \int_0^t A(\tau)y(\tau)d\tau \tag{8.16}$$

has a unique solution given by

$$y(t) = \phi(t) + \sum_{n=1}^{\infty} (-1)^{n\cdot 1_{t<0}} \int_{\Delta_n(t)} A(\tau_n)\dots A(\tau_1)\phi(\tau_1)d\tau \tag{8.17}$$

and this solution satisfies the bound

$$\|y\|_\infty \le \|\phi\|_\infty e^{\int_J \|A(\tau)\|d\tau}.$$

Proof. Define $\Lambda : BC(J,X) \to BC(J,X)$ by

$$(\Lambda y)(t) = \int_0^t A(\tau)y(\tau)d\tau.$$

Then y solves Eq. (8.14) iff $y = \phi + \Lambda y$ or equivalently iff $(I-\Lambda)y = \phi$. An induction argument shows

$$\begin{aligned}
(\Lambda^n\phi)(t) &= \int_0^t d\tau_n A(\tau_n)(\Lambda^{n-1}\phi)(\tau_n) \\
&= \int_0^t d\tau_n \int_0^{\tau_n} d\tau_{n-1} A(\tau_n)A(\tau_{n-1})(\Lambda^{n-2}\phi)(\tau_{n-1}) \\
&\vdots \\
&= \int_0^t d\tau_n \int_0^{\tau_n} d\tau_{n-1}\dots\int_0^{\tau_2} d\tau_1 A(\tau_n)\dots A(\tau_1)\phi(\tau_1) \\
&= (-1)^{n\cdot 1_{t<0}} \int_{\Delta_n(t)} A(\tau_n)\dots A(\tau_1)\phi(\tau_1)d\tau.
\end{aligned}$$

Taking norms of this equation and using the triangle inequality along with Lemma 8.19 gives,

$$\begin{aligned}\|(\Lambda^n\phi)(t)\| &\le \|\phi\|_\infty \cdot \int_{\Delta_n(t)} \|A(\tau_n)\| \dots \|A(\tau_1)\| d\tau \\ &\le \|\phi\|_\infty \cdot \frac{1}{n!}\left(\int_{\Delta_1(t)} \|A(\tau)\| d\tau\right)^n \\ &\le \|\phi\|_\infty \cdot \frac{1}{n!}\left(\int_J \|A(\tau)\| d\tau\right)^n.\end{aligned}$$

Therefore,

$$\|\Lambda^n\|_{op} \le \frac{1}{n!}\left(\int_J \|A(\tau)\| d\tau\right)^n \tag{8.18}$$

and

$$\sum_{n=0}^{\infty} \|\Lambda^n\|_{op} \le e^{\int_J \|A(\tau)\| d\tau} < \infty$$

where $\|\cdot\|_{op}$ denotes the operator norm on $L\left(BC(J,X)\right)$. An application of Proposition 7.19 now shows $(I-\Lambda)^{-1} = \sum_{n=0}^{\infty} \Lambda^n$ exists and

$$\left\|(I-\Lambda)^{-1}\right\|_{op} \le e^{\int_J \|A(\tau)\| d\tau}.$$

It is now only a matter of working through the notation to see that these assertions prove the theorem. ■

Corollary 8.22. *Suppose $h \in C(J \to X)$ and $x \in X$, then there exits a unique solution, $y \in C^1(J,X)$, to the linear ordinary differential Eq. (8.13).*

Proof. Let

$$\phi(t) = x + \int_0^t h(\tau)\, d\tau.$$

By applying Theorem 8.21 with and J replaced by any open interval J_0 such that $0 \in J_0$ and $\bar{J}_0$ is a compact subinterval[1] of J, there exists a unique solution y_{J_0} to Eq. (8.13) which is valid for $t \in J_0$. By uniqueness of solutions, if J_1 is a subinterval of J such that $J_0 \subset J_1$ and $\bar{J}_1$ is a compact subinterval of J, we have $y_{J_1} = y_{J_0}$ on J_0. Because of this observation, we may construct a solution y to Eq. (8.13) which is defined on the full interval J by setting $y(t) = y_{J_0}(t)$ for any J_0 as above which also contains $t \in J$. ■

Corollary 8.23. *Suppose that $A \in L(X)$ is independent of time, then the solution to*

$$\dot{y}(t) = Ay(t) \text{ with } y(0) = x$$

[1] We do this so that $\phi|_{J_0}$ will be bounded.

is given by $y(t) = e^{tA}x$ *where*

$$e^{tA} = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n. \tag{8.19}$$

Moreover,

$$e^{(t+s)A} = e^{tA}e^{sA} \text{ for all } s, t \in \mathbb{R}. \tag{8.20}$$

Proof. The first assertion is a simple consequence of Eq. 8.17 and Lemma 8.19 with $\psi = 1$. The assertion in Eq. (8.20) may be proved by explicit computation but the following proof is more instructive. Given $x \in X$, let $y(t) := e^{(t+s)A}x$. By the chain rule,

$$\begin{aligned}\frac{d}{dt}y(t) &= \frac{d}{d\tau}|_{\tau=t+s}e^{\tau A}x = Ae^{\tau A}x|_{\tau=t+s} \\ &= Ae^{(t+s)A}x = Ay(t) \text{ with } y(0) = e^{sA}x.\end{aligned}$$

The unique solution to this equation is given by

$$y(t) = e^{tA}x(0) = e^{tA}e^{sA}x.$$

This completes the proof since, by definition, $y(t) = e^{(t+s)A}x$. ■

We also have the following converse to this corollary whose proof is outlined in Exercise 8.21 below.

Theorem 8.24. *Suppose that* $T_t \in L(X)$ *for* $t \geq 0$ *satisfies*

1. *(Semi-group property.)* $T_0 = Id_X$ *and* $T_tT_s = T_{t+s}$ *for all* $s, t \geq 0$.
2. *(Norm Continuity)* $t \to T_t$ *is continuous at* 0, *i.e.* $\|T_t - I\|_{L(X)} \to 0$ *as* $t \downarrow 0$.

Then there exists $A \in L(X)$ *such that* $T_t = e^{tA}$ *where* e^{tA} *is defined in Eq. (8.19).*

8.4 Classical Weierstrass Approximation Theorem

Definition 8.25 (Support). *Let* $f : X \to Z$ *be a function from a metric space* (X, ρ) *to a vector space* Z. *The support of* f *is the closed subset,* $\text{supp}(f)$, *of* X *defined by*

$$\text{supp}(f) := \overline{\{x \in X : f(x) \neq 0\}}.$$

Example 8.26. For example if $f : \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = \sin(x)1_{[0,4\pi]}(x) \in \mathbb{R}$, then

$$\{f \neq 0\} = (0, 4\pi) \setminus \{\pi, 2\pi, 3\pi\}$$

and therefore $\text{supp}(f) = [0, 4\pi]$.

For the remainder of this section, Z will be used to denote a Banach space.

Definition 8.27 (Convolution). *For $f, g \in C(\mathbb{R})$ with either f or g having compact support, we define the* ***convolution*** *of f and g by*

$$f * g(x) = \int_{\mathbb{R}} f(x-y)g(y)dy = \int_{\mathbb{R}} f(y)g(x-y)dy.$$

We will also use this definition when one of the functions, either f or g, takes values in a Banach space Z.

Lemma 8.28 (Approximate δ − sequences). *Suppose that $\{q_n\}_{n=1}^{\infty}$ is a sequence non-negative continuous real valued functions on $\mathbb{R}$ with compact support that satisfy*

$$\int_{\mathbb{R}} q_n(x)\ dx = 1 \text{ and} \tag{8.21}$$

$$\lim_{n\to\infty} \int\limits_{|x|\geq\varepsilon} q_n(x)dx = 0 \text{ for all } \varepsilon > 0. \tag{8.22}$$

If $f \in BC(\mathbb{R}, Z)$, then

$$q_n * f(x) := \int_{\mathbb{R}} q_n(y) f(x-y)dy$$

converges to f uniformly on compact subsets of $\mathbb{R}\times W \subset \mathbb{R}^{d+1}$.

Proof. Let $x \in \mathbb{R}$, then because of Eq. (8.21),

$$\begin{aligned}\|q_n * f(x) - f(x)\| &= \left\|\int_{\mathbb{R}} q_n(y)\left(f(x-y) - f(x)\right) dy\right\| \\ &\leq \int_{\mathbb{R}} q_n(y) \left\|f(x-y) - f(x)\right\| dy.\end{aligned}$$

Let $M = \sup\{\|f(x)\| : x \in \mathbb{R}\}$. Then for any $\varepsilon > 0$, using Eq. (8.21),

$$\begin{aligned}\|q_n * f(x) - f(x)\| &\leq \int_{|y|\leq\varepsilon} q_n(y) \|f(x-y) - f(x)\| dy \\ &\quad + \int_{|y|>\varepsilon} q_n(y) \|f(x-y) - f(x)\| dy \\ &\leq \sup_{|w|\leq\varepsilon} \|f(x+w) - f(x)\| + 2M \int_{|y|>\varepsilon} q_n(y)dy.\end{aligned}$$

So if K is a compact subset of $\mathbb{R}$ (for example a large interval) we have

$$\begin{aligned}&\sup_{(x)\in K} \|q_n * f(x) - f(x)\| \\ &\qquad \leq \sup_{|w|\leq\varepsilon,\ x\in K} \|f(x+w) - f(x)\| + 2M \int_{\|y\|>\varepsilon} q_n(y)dy\end{aligned}$$

and hence by Eq. (8.22),

$$\limsup_{n\to\infty} \sup_{x\in K} \|q_n * f(x) - f(x)\| \\ \leq \sup_{|w|\leq\varepsilon,\ x\in K} \|f(x+w) - f(x)\|.$$

This finishes the proof since the right member of this equation tends to 0 as $\varepsilon \downarrow 0$ by uniform continuity of f on compact subsets of $\mathbb{R}$. ■

Let $q_n : \mathbb{R} \to [0,\infty)$ be defined by

$$q_n(x) := \frac{1}{c_n}(1-x^2)^n 1_{|x|\leq 1} \text{ where } c_n := \int_{-1}^{1} (1-x^2)^n dx. \tag{8.23}$$

Figure 8.2 displays the key features of the functions q_n.

Fig. 8.2. A plot of q_1, q_{50}, and q_{100}. The most peaked curve is q_{100} and the least is q_1. The total area under each of these curves is one.

Lemma 8.29. *The sequence $\{q_n\}_{n=1}^{\infty}$ is an approximate δ – sequence, i.e. they satisfy Eqs. (8.21) and (8.22).*

Proof. By construction, $q_n \in C_c(\mathbb{R}, [0,\infty))$ for each n and Eq. 8.21 holds. Since

$$\int_{|x|\geq\varepsilon} q_n(x)dx = \frac{2\int_\varepsilon^1 (1-x^2)^n dx}{2\int_0^\varepsilon (1-x^2)^n dx + 2\int_\varepsilon^1 (1-x^2)^n dx} \\ \leq \frac{\int_\varepsilon^1 \frac{x}{\varepsilon}(1-x^2)^n dx}{\int_0^\varepsilon \frac{x}{\varepsilon}(1-x^2)^n dx} = \frac{(1-x^2)^{n+1}|_\varepsilon^1}{(1-x^2)^{n+1}|_0^\varepsilon} \\ = \frac{(1-\varepsilon^2)^{n+1}}{1-(1-\varepsilon^2)^{n+1}} \to 0 \text{ as } n\to\infty,$$

the proof is complete. ■

Notation 8.30 *Let* $\mathbb{Z}_+ := \mathbb{N} \cup \{0\}$ *and for* $x \in \mathbb{R}^d$ *and* $\alpha \in \mathbb{Z}_+^d$ *let* $x^\alpha = \prod_{i=1}^d x_i^{\alpha_i}$ *and* $|\alpha| = \sum_{i=1}^d \alpha_i$*. A polynomial on* $\mathbb{R}^d$ *with values in* Z *is a function* $p : \mathbb{R}^d \to Z$ *of the form*

$$p(x) = \sum_{\alpha:|\alpha|\le N} p_\alpha x^\alpha \text{ with } p_\alpha \in Z \text{ and } N \in \mathbb{Z}_+.$$

If $p_\alpha \neq 0$ *for some* α *such that* $|\alpha| = N$*, then we define* $\deg(p) := N$ *to be the degree of* p*. If* Z *is a complex Banach space, the function* p *has a natural extension to* $z \in \mathbb{C}^d$*, namely* $p(z) = \sum_{\alpha:|\alpha|\le N} p_\alpha z^\alpha$ *where* $z^\alpha = \prod_{i=1}^d z_i^{\alpha_i}$.

Given a compact subset $K \subset \mathbb{R}^d$ and $f \in C(K, \mathbb{C})$[2], we are going to show, in the Weierstrass approximation Theorem 8.34 below, that f may be uniformly approximated by polynomial functions on K. The next theorem addresses this question when K is a compact subinterval of $\mathbb{R}$.

Theorem 8.31 (Weierstrass Approximation Theorem). *Suppose* $-\infty < a < b < \infty$, $J = [a, b]$ *and* $f \in C(J, Z)$*. Then there exists polynomials* p_n *on* $\mathbb{R}$ *such that* $p_n \to f$ *uniformly on* J*.*

Proof. By replacing f by F where

$$F(t) := f(a + t(b-a)) - [f(a) + t(f(b) - f(a))] \text{ for } t \in [0,1],$$

it suffices to assume $a = 0$, $b = 1$ and $f(0) = f(1) = 0$. Furthermore we may now extend f to a continuous function on all $\mathbb{R}$ by setting $f \equiv 0$ on $\mathbb{R} \setminus [0,1]$.

With q_n defined as in Eq. (8.23), let $f_n(x) := (q_n * f)(x)$ and recall from Lemma 8.28 that $f_n(x) \to f(x)$ as $n \to \infty$ with the convergence being uniform in $x \in [0,1]$. This completes the proof since f_n is equal to a polynomial function on $[0,1]$. Indeed, there are polynomials, $a_k(y)$, such that

$$(1 - (x-y)^2)^n = \sum_{k=0}^{2n} a_k(y)\, x^k,$$

and therefore, for $x \in [0,1]$,

---

[2] Note that f is automatically bounded because if not there would exist $u_n \in K$ such that $\lim_{n\to\infty} |f(u_n)| = \infty$. Using Theorem 8.2 we may, by passing to a subsequence if necessary, assume $u_n \to u \in K$ as $n \to \infty$. Now the continuity of f would then imply

$$\infty = \lim_{n\to\infty} |f(u_n)| = |f(u)|$$

which is absurd since f takes values in $\mathbb{C}$.

$$\begin{aligned}
f_n(x) &= \int_{\mathbb{R}} q_n(x-y) f(y) dy \\
&= \frac{1}{c_n} \int_{[0,1]} f(y) \left[(1-(x-y)^2)^n 1_{|x-y| \le 1} \right] dy \\
&= \frac{1}{c_n} \int_{[0,1]} f(y) (1-(x-y)^2)^n dy \\
&= \frac{1}{c_n} \int_{[0,1]} f(y) \sum_{k=0}^{2n} a_k(y) x^k dy = \sum_{k=0}^{2n} A_k x^k
\end{aligned}$$

where

$$A_k = \int_{[0,1]} f(y) a_k(y) dy.$$

■

Lemma 8.32. *Suppose $J = [a, b]$ is a compact subinterval of $\mathbb{R}$ and K is a compact subset of $\mathbb{R}^{d-1}$, then the linear mapping $R : C(J \times K, Z) \to C(J, C(K, Z))$ defined by $(Rf)(t) = f(t, \cdot) \in C(K, Z)$ for $t \in J$ is an isometric isomorphism of Banach spaces.*

Proof. By uniform continuity of f on $J \times K$ (see Theorem 8.2),

$$\|(Rf)(t) - (Rf)(s)\|_{C(K,Z)} = \max_{y \in K} \|f(t, y) - f(s, y)\|_Z \to 0 \text{ as } s \to t$$

which shows that Rf is indeed in $C(J \to C(K, Z))$. Moreover,

$$\begin{aligned}
\|Rf\|_{C(J \to C(K,Z))} &= \max_{t \in J} \|(Rf)(t)\|_{C(K,Z)} \\
&= \max_{t \in J} \max_{y \in K} \|f(t, y)\|_Z = \|f\|_{C(J \times K, Z)},
\end{aligned}$$

showing R is isometric and therefore injective.

To see that R is surjective, let $F \in C(J \to C(K, Z))$ and define $f(t, y) := F(t)(y)$. Since

$$\begin{aligned}
\|f(t, y) - f(s, y')\|_Z &\le \|f(t, y) - f(s, y)\|_Z + \|f(s, y) - f(s, y')\|_Z \\
&\le \|F(t) - F(s)\|_{C(K,Z)} + \|F(s)(y) - F(s)(y')\|_Z,
\end{aligned}$$

it follows by the continuity of $t \to F(t)$ and $y \to F(s)(y)$ that

$$\|f(t, y) - f(s, y')\|_Z \to 0 \text{ as } (t, y) \to (s, y').$$

This shows $f \in C(J \times K, Z)$ and thus completes the proof because $Rf = F$ by construction. ■

Corollary 8.33 (Weierstrass Approximation Theorem). *Let $d \in \mathbb{N}$, $J_i = [a_i, b_i]$ be compact subintervals of $\mathbb{R}$ for $i = 1, 2, \ldots, d$, $J := J_1 \times \cdots \times J_d$ and $f \in C(J, Z)$. Then there exists polynomials p_n on $\mathbb{R}^d$ such that $p_n \to f$ uniformly on J.*

Proof. The proof will be by induction on d with the case $d = 1$ being the content of Theorem 8.31. Now suppose that $d > 1$ and the theorem holds with d replaced by $d-1$. Let $K := J_2 \times \cdots \times J_d$, $Z_0 = C(K, Z)$, $R : C(J_1 \times K, Z) \to C(J_1, Z_0)$ be as in Lemma 8.32 and $F := Rf$. By Theorem 8.31, for any $\varepsilon > 0$ there exists a polynomial function

$$p(t) = \sum_{k=0}^{n} c_k t^k$$

with $c_k \in Z_0 = C(K, Z)$ such that $\|F - p\|_{C(J_1, Z_0)} < \varepsilon$. By the induction hypothesis, there exists polynomial functions $q_k : K \to Z$ such that

$$\|c_k - q_k\|_{Z_0} < \frac{\varepsilon}{n(|a| + |b|)^k}.$$

It is now easily verified (you check) that the polynomial function,

$$\rho(x) := \sum_{k=0}^{n} x_1^k q_k(x_2, \ldots, x_d) \text{ for } x \in J$$

satisfies $\|f - \rho\|_{C(J,Z)} < 2\varepsilon$ and this completes the induction argument and hence the proof. ∎

The reader is referred to Chapter 20 for a two more alternative proofs of this corollary.

Theorem 8.34 (Weierstrass Approximation Theorem). *Suppose that $K \subset \mathbb{R}^d$ is a compact subset and $f \in C(K, \mathbb{C})$. Then there exists polynomials p_n on $\mathbb{R}^d$ such that $p_n \to f$ uniformly on K.*

Proof. Choose $\lambda > 0$ and $b \in \mathbb{R}^d$ such that

$$K_0 := \lambda K - b := \{\lambda x - b : x \in K\} \subset B_d$$

where $B_d := (0,1)^d$. The function $F(y) := f\left(\lambda^{-1}(y + b)\right)$ for $y \in K_0$ is in $C(K_0, \mathbb{C})$ and if $\hat{p}_n(y)$ are polynomials on $\mathbb{R}^d$ such that $\hat{p}_n \to F$ uniformly on K_0 then $p_n(x) := \hat{p}_n(\lambda x - b)$ are polynomials on $\mathbb{R}^d$ such that $p_n \to f$ uniformly on K. Hence we may now assume that K is a compact subset of B_d. Let $g \in C(K \cup B_d^c)$ be defined by

$$g(x) = \begin{cases} f(x) & \text{if } x \in K \\ 0 & \text{if } x \in B_d^c \end{cases}$$

and then use the Tietze extension Theorem 7.4 (applied to the real and imaginary parts of F) to find a continuous function $F \in C(\mathbb{R}^d, \mathbb{C})$ such that $F = g|_{K \cup B_d^c}$. If p_n are polynomials on $\mathbb{R}^d$ such that $p_n \to F$ uniformly on $[0,1]^d$ then p_n also converges to f uniformly on K. Hence, by replacing f by F, we may now assume that $f \in C(\mathbb{R}^d, \mathbb{C})$, $K = \bar{B}_d = [0,1]^d$, and $f \equiv 0$ on B_d^c. The result now follows by an application of Corollary 8.33 with $Z = \mathbb{C}$. ∎

Remark 8.35. The mapping $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \to z = x + iy \in \mathbb{C}^d$ is an isomorphism of vector spaces. Letting $\bar{z} = x - iy$ as usual, we have $x = \frac{z+\bar{z}}{2}$ and $y = \frac{z-\bar{z}}{2i}$. Therefore under this identification any polynomial $p(x, y)$ on $\mathbb{R}^d \times \mathbb{R}^d$ may be written as a polynomial q in $(z, \bar{z})$, namely

$$q(z, \bar{z}) = p(\frac{z + \bar{z}}{2}, \frac{z - \bar{z}}{2i}).$$

Conversely a polynomial q in $(z, \bar{z})$ may be thought of as a polynomial p in (x, y), namely $p(x, y) = q(x + iy, x - iy)$.

Corollary 8.36 (Complex Weierstrass Approximation Theorem). *Suppose that $K \subset \mathbb{C}^d$ is a compact set and $f \in C(K, \mathbb{C})$. Then there exists polynomials $p_n(z, \bar{z})$ for $z \in \mathbb{C}^d$ such that $\sup_{z \in K} |p_n(z, \bar{z}) - f(z)| \to 0$ as $n \to \infty$.*

Proof. This is an immediate consequence of Theorem 8.34 and Remark 8.35. ■

Example 8.37. Let $K = S^1 = \{z \in \mathbb{C} : |z| = 1\}$ and $\mathcal{A}$ be the set of polynomials in $(z, \bar{z})$ restricted to S^1. Then $\mathcal{A}$ is dense in $C(S^1)$.[3] Since $\bar{z} = z^{-1}$ on S^1, we have shown polynomials in z and z^{-1} are dense in $C(S^1)$. This example generalizes in an obvious way to $K = \left(S^1\right)^d \subset \mathbb{C}^d$.

Exercise 8.4. Suppose $-\infty < a < b < \infty$ and $f \in C\left([a, b], \mathbb{C}\right)$ satisfies

$$\int_a^b f\left(t\right) t^n dt = 0 \text{ for } n = 0, 1, 2 \ldots.$$

Show $f \equiv 0$.

Exercise 8.5. Suppose $f \in C\left(\mathbb{R}, \mathbb{C}\right)$ is a 2π – periodic function (i.e. $f\left(x + 2\pi\right) = f\left(x\right)$ for all $x \in \mathbb{R}$) and

$$\int_0^{2\pi} f\left(x\right) e^{inx} dx = 0 \text{ for all } n \in \mathbb{Z},$$

show again that $f \equiv 0$. **Hint**: Use Example 8.37 to show that any 2π – periodic continuous function g on $\mathbb{R}$ is the uniform limit of trigonometric polynomials of the form

$$p\left(x\right) = \sum_{k=-n}^{n} p_k e^{ikx} \text{ with } p_k \in \mathbb{C} \text{ for all } k.$$

[3] Note that it is easy to extend $f \in C(S^1)$ to a function $F \in C(\mathbb{C})$ by setting $F(z) = z f(\frac{z}{|z|})$ for $z \neq 0$ and $F(0) = 0$. So this special case does not require the Tietze extension theorem.

8.5 Iterated Integrals

Theorem 8.38 (Baby Fubini Theorem). *Let $a_i, b_i \in \mathbb{R}$ with $a_i \neq b_i$ for $i = 1, 2 \ldots, n$, $f(t_1, t_2, \ldots, t_n) \in Z$ be a continuous function of $(t_1, t_2, \ldots, t_n)$ where t_i between a_i and b_i for each i and for any given permutation, σ, of $\{1, 2 \ldots, n\}$ let*

$$I_\sigma(f) := \int_{a_{\sigma 1}}^{b_{\sigma 1}} dt_{\sigma 1} \ldots \int_{a_{\sigma n}}^{b_{\sigma n}} dt_{\sigma n} f(t_1, t_2, \ldots, t_n). \tag{8.24}$$

Then $I_\sigma(f)$ is well defined and independent of σ, i.e. the order of iterated integrals is irrelevant under these hypothesis.

Proof. Let $J_i := [\min(a_i, b_i), \max(a_i, b_i)]$, $J := J_1 \times \cdots \times J_n$ and $|J_i| := \max(a_i, b_i) - \min(a_i, b_i)$. Using the uniform continuity of f (Theorem 8.2) and the continuity of the Riemann integral, it is easy to prove (compare with the proof of Lemma 8.32) that the map

$$(t_1, \ldots \hat{t}_{\sigma n}, \ldots, t_n) \in (J_1 \times \cdots \times \hat{J}_{\sigma n} \times \cdots \times J_n) \to \int_{a_{\sigma n}}^{b_{\sigma n}} dt_{\sigma n} f(t_1, t_2, \ldots, t_n)$$

is continuous, where the hat is used to denote a missing element from a list. From this remark, it follows that each of the integrals in Eq. (8.24) are well defined and hence so is $I_\sigma(f)$. Moreover by an induction argument using Lemma 8.32 and the boundedness of the Riemann integral, we have the estimate,

$$\|I_\sigma(f)\|_Z \leq \left(\prod_{i=1}^n |J_i|\right) \|f\|_{C(J,Z)}. \tag{8.25}$$

Now suppose τ is another permutation. Because of Eq. (8.25), I_σ and I_τ are bounded operators on $C(J, Z)$ and so to shows $I_\sigma = I_\tau$ is suffices to shows there are equal on the dense set of polynomial functions (see Corollary 8.33) in $C(J, Z)$. Moreover by linearity, it suffices to show $I_\sigma(f) = I_\tau(f)$ when f has the form

$$f(t_1, t_2, \ldots, t_n) = t_1^{k_1} \ldots t_n^{k_n} z$$

for some $k_i \in \mathbb{N}_0$ and $z \in Z$. However for this function, explicit computations show

$$I_\sigma(f) = I_\tau(f) = \left(\prod_{i=1}^n \frac{b_i^{k_i+1} - a_i^{k_i+1}}{k_i + 1}\right) \cdot z.$$

■

Proposition 8.39 (Equality of Mixed Partial Derivatives). *Let $Q = (a, b) \times (c, d)$ be an open rectangle in $\mathbb{R}^2$ and $f \in C(Q, Z)$. Assume that $\frac{\partial}{\partial t} f(s, t)$, $\frac{\partial}{\partial s} f(s, t)$ and $\frac{\partial}{\partial t}\frac{\partial}{\partial s} f(s, t)$ exists and are continuous for $(s, t) \in Q$, then $\frac{\partial}{\partial s}\frac{\partial}{\partial t} f(s, t)$ exists for $(s, t) \in Q$ and*

$$\frac{\partial}{\partial s}\frac{\partial}{\partial t} f(s, t) = \frac{\partial}{\partial t}\frac{\partial}{\partial s} f(s, t) \text{ for } (s, t) \in Q. \tag{8.26}$$

Proof. Fix $(s_0, t_0) \in Q$. By two applications of Theorem 8.13,

$$\begin{aligned} f(s,t) &= f(s_{t_0}, t) + \int_{s_0}^{s} \frac{\partial}{\partial \sigma} f(\sigma, t) d\sigma \\ &= f(s_0, t) + \int_{s_0}^{s} \frac{\partial}{\partial \sigma} f(\sigma, t_0) d\sigma + \int_{s_0}^{s} d\sigma \int_{t_0}^{t} d\tau \frac{\partial}{\partial \tau} \frac{\partial}{\partial \sigma} f(\sigma, \tau) \end{aligned} \tag{8.27}$$

and then by Fubini's Theorem 8.38 we learn

$$f(s,t) = f(s_0, t) + \int_{s_0}^{s} \frac{\partial}{\partial \sigma} f(\sigma, t_0) d\sigma + \int_{t_0}^{t} d\tau \int_{s_0}^{s} d\sigma \frac{\partial}{\partial \tau} \frac{\partial}{\partial \sigma} f(\sigma, \tau).$$

Differentiating this equation in t and then in s (again using two more applications of Theorem 8.13) shows Eq. (8.26) holds. ■

8.6 Exercises

Throughout these problems, $(X, \|\cdot\|)$ is a Banach space.

Exercise 8.6. Show $f = (f_1, \dots, f_n) \in \bar{\mathcal{S}}([a,b], \mathbb{R}^n)$ iff $f_i \in \bar{\mathcal{S}}([a,b], \mathbb{R})$ for $i = 1, 2, \dots, n$ and

$$\int_a^b f(t) dt = \left(\int_a^b f_1(t) dt, \dots, \int_a^b f_n(t) dt \right).$$

Here $\mathbb{R}^n$ is to be equipped with the usual Euclidean norm. **Hint:** Use Lemma 8.7 to prove the forward implication.

Exercise 8.7. Prove Theorem 8.38 using the following strategy.

1. Use the results from the proof in the text of Theorem 8.38 that
$$s \to \int_c^d f(s,t) dt \text{ and } t \to \int_a^b f(s,t) ds$$
are continuous maps.
2. For the moment take $X = \mathbb{R}$ and prove Eq. (8.24) holds by first proving it holds when $f(s,t) = s^m t^n$ with $m, n \in \mathbb{N}_0$. Then use this result along with Theorem 8.34 to show Eq. (8.24) holds for all $f \in C([a,b] \times [c,d], \mathbb{R})$.
3. For the general case, use the special case proved in item 2. along with Hahn - Banach theorem.

Exercise 8.8. Give another proof of Proposition 8.39 which does not use Fubini's Theorem 8.38 as follows.

1. By a simple translation argument we may assume $(0,0) \in Q$ and we are trying to prove Eq. (8.26) holds at $(s,t) = (0,0)$.

2. Let $h(s,t) := \frac{\partial}{\partial t}\frac{\partial}{\partial s} f(s,t)$ and

$$G(s,t) := \int_0^s d\sigma \int_0^t d\tau h(\sigma,\tau)$$

so that Eq. (8.27) states

$$f(s,t) = f(0,t) + \int_0^s \frac{\partial}{\partial \sigma} f(\sigma, t_0) d\sigma + G(s,t)$$

and differentiating this equation at $t = 0$ shows

$$\frac{\partial}{\partial t} f(s,0) = \frac{\partial}{\partial t} f(0,0) + \frac{\partial}{\partial t} G(s,0). \tag{8.28}$$

Now show using the definition of the derivative that

$$\frac{\partial}{\partial t} G(s,0) = \int_0^s d\sigma h(\sigma,0). \tag{8.29}$$

Hint: Consider

$$G(s,t) - t \int_0^s d\sigma h(\sigma,0) = \int_0^s d\sigma \int_0^t d\tau \left[h(\sigma,\tau) - h(\sigma,0)\right].$$

3. Now differentiate Eq. (8.28) in s using Theorem 8.13 to finish the proof.

Exercise 8.9. Give another proof of Eq. (8.24) in Theorem 8.38 based on Proposition 8.39. To do this let $t_0 \in (c,d)$ and $s_0 \in (a,b)$ and define

$$G(s,t) := \int_{t_0}^t d\tau \int_{s_0}^s d\sigma f(\sigma,\tau)$$

Show G satisfies the hypothesis of Proposition 8.39 which combined with two applications of the fundamental theorem of calculus implies

$$\frac{\partial}{\partial t}\frac{\partial}{\partial s} G(s,t) = \frac{\partial}{\partial s}\frac{\partial}{\partial t} G(s,t) = f(s,t).$$

Use two more applications of the fundamental theorem of calculus along with the observation that $G = 0$ if $t = t_0$ or $s = s_0$ to conclude

$$G(s,t) = \int_{s_0}^s d\sigma \int_{t_0}^t d\tau \frac{\partial}{\partial \tau}\frac{\partial}{\partial \sigma} G(\sigma,\tau) = \int_{s_0}^s d\sigma \int_{t_0}^t d\tau \frac{\partial}{\partial \tau} f(\sigma,\tau). \tag{8.30}$$

Finally let $s = b$ and $t = d$ in Eq. (8.30) and then let $s_0 \downarrow a$ and $t_0 \downarrow c$ to prove Eq. (8.24).

Exercise 8.10 (Product Rule). Prove items 1. and 2. of Lemma 8.9. This can be modeled on the standard proof for real valued functions.

Exercise 8.11 (Chain Rule). Prove the chain rule in Proposition 8.10. Again this may be modeled on the on the standard proof for real valued functions.

Exercise 8.12. To each $A \in L(X)$, we may define $L_A, R_A : L(X) \to L(X)$ by

$$L_A B = AB \text{ and } R_A B = BA \text{ for all } B \in L(X).$$

Show $L_A, R_A \in L(L(X))$ and that

$$\|L_A\|_{L(L(X))} = \|A\|_{L(X)} = \|R_A\|_{L(L(X))}.$$

Exercise 8.13. Suppose that $A : \mathbb{R} \to L(X)$ is a continuous function and $U, V : \mathbb{R} \to L(X)$ are the unique solution to the linear differential equations

$$\dot{V}(t) = A(t)V(t) \text{ with } V(0) = I \tag{8.31}$$

and

$$\dot{U}(t) = -U(t)A(t) \text{ with } U(0) = I. \tag{8.32}$$

Prove that $V(t)$ is invertible and that $V^{-1}(t) = U(t)$[4], where by abuse of notation I am writing $V^{-1}(t)$ for $[V(t)]^{-1}$. **Hints:** 1) show $\frac{d}{dt}[U(t)V(t)] = 0$ (which is sufficient if $\dim(X) < \infty$) and 2) show compute $y(t) := V(t)U(t)$ solves a linear differential ordinary differential equation that has $y \equiv Id$ as an obvious solution. (The results of Exercise 8.12 may be useful here.) Then use the uniqueness of solutions to linear ODEs.

Exercise 8.14. Suppose that $(X, \|\cdot\|)$ is a Banach space, $J = (a, b)$ with $-\infty \leq a < b \leq \infty$ and $f_n : \mathbb{R} \to X$ are continuously differentiable functions such that there exists a summable sequence $\{a_n\}_{n=1}^{\infty}$ satisfying

$$\|f_n(t)\| + \left\|\dot{f}_n(t)\right\| \leq a_n \text{ for all } t \in J \text{ and } n \in \mathbb{N}.$$

Show:

1. $\sup\left\{\left\|\frac{f_n(t+h) - f_n(t)}{h}\right\| : (t, h) \in J \times \mathbb{R} \ni t + h \in J \text{ and } h \neq 0\right\} \leq a_n.$
2. The function $F : \mathbb{R} \to X$ defined by

$$F(t) := \sum_{n=1}^{\infty} f_n(t) \text{ for all } t \in J$$

is differentiable and for $t \in J$,

$$\dot{F}(t) = \sum_{n=1}^{\infty} \dot{f}_n(t).$$

[4] The fact that $U(t)$ must be defined as in Eq. (8.32) follows from Lemma 8.9.

Exercise 8.15. Suppose that $A \in L(X)$. Show directly that:

1. e^{tA} define in Eq. (8.19) is convergent in $L(X)$ when equipped with the operator norm.
2. e^{tA} is differentiable in t and that $\frac{d}{dt}e^{tA} = Ae^{tA}$.

Exercise 8.16. Suppose that $A \in L(X)$ and $v \in X$ is an eigenvector of A with eigenvalue λ, i.e. that $Av = \lambda v$. Show $e^{tA}v = e^{t\lambda}v$. Also show that if $X = \mathbb{R}^n$ and A is a diagonalizable $n \times n$ matrix with

$$A = SDS^{-1} \text{ with } D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$$

then $e^{tA} = Se^{tD}S^{-1}$ where $e^{tD} = \operatorname{diag}(e^{t\lambda_1}, \dots, e^{t\lambda_n})$. Here $\operatorname{diag}(\lambda_1, \dots, \lambda_n)$ denotes the diagonal matrix Λ such that $\Lambda_{ii} = \lambda_i$ for $i = 1, 2, \dots, n$.

Exercise 8.17. Suppose that $A, B \in L(X)$ and $[A, B] := AB - BA = 0$. Show that $e^{(A+B)} = e^A e^B$.

Exercise 8.18. Suppose $A \in C(\mathbb{R}, L(X))$ satisfies $[A(t), A(s)] = 0$ for all $s, t \in \mathbb{R}$. Show

$$y(t) := e^{\left(\int_0^t A(\tau)d\tau\right)}x$$

is the unique solution to $\dot{y}(t) = A(t)y(t)$ with $y(0) = x$.

Exercise 8.19. Compute e^{tA} when

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

and use the result to prove the formula

$$\cos(s + t) = \cos s \cos t - \sin s \sin t.$$

Hint: Sum the series and use $e^{tA}e^{sA} = e^{(t+s)A}$.

Exercise 8.20. Compute e^{tA} when

$$A = \begin{pmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{pmatrix}$$

with $a, b, c \in \mathbb{R}$. Use your result to compute $e^{t(\lambda I + A)}$ where $\lambda \in \mathbb{R}$ and I is the 3×3 identity matrix. **Hint:** Sum the series.

Exercise 8.21. Prove Theorem 8.24 using the following outline.

1. Using the right continuity at 0 and the semi-group property for T_t, show there are constants M and C such that $\|T_t\|_{L(X)} \le MC^t$ for all $t > 0$.
2. Show $t \in [0, \infty) \to T_t \in L(X)$ is continuous.

3. For $\varepsilon > 0$, let $S_\varepsilon := \frac{1}{\varepsilon}\int_0^\varepsilon T_\tau d\tau \in L(X)$. Show $S_\varepsilon \to I$ as $\varepsilon \downarrow 0$ and conclude from this that S_ε is invertible when $\varepsilon > 0$ is sufficiently small. For the remainder of the proof fix such a small $\varepsilon > 0$.
4. Show
$$T_t S_\varepsilon = \frac{1}{\varepsilon}\int_t^{t+\varepsilon} T_\tau d\tau$$
and conclude from this that
$$\lim_{t\downarrow 0}\left(\frac{T_t - I}{t}\right) S_\varepsilon = \frac{1}{\varepsilon}\left(T_\varepsilon - Id_X\right).$$
5. Using the fact that S_ε is invertible, conclude $A = \lim_{t\downarrow 0} t^{-1}\left(T_t - I\right)$ exists in $L(X)$ and that
$$A = \frac{1}{\varepsilon}\left(T_\varepsilon - I\right) S_\varepsilon^{-1}.$$
6. Now show, using the semigroup property and step 4., that $\frac{d}{dt}T_t = AT_t$ for all $t > 0$.
7. Using step 5, show $\frac{d}{dt}e^{-tA}T_t = 0$ for all $t > 0$ and therefore $e^{-tA}T_t = e^{-0A}T_0 = I$.

Exercise 8.22 (Duhamel' s Principle I). Suppose that $A : \mathbb{R} \to L(X)$ is a continuous function and $V : \mathbb{R} \to L(X)$ is the unique solution to the linear differential equation in Eq. (8.31). Let $x \in X$ and $h \in C(\mathbb{R}, X)$ be given. Show that the unique solution to the differential equation:

$$\dot{y}(t) = A(t)y(t) + h(t) \text{ with } y(0) = x \tag{8.33}$$

is given by

$$y(t) = V(t)x + V(t)\int_0^t V(\tau)^{-1}h(\tau)\,d\tau. \tag{8.34}$$

Hint: compute $\frac{d}{dt}[V^{-1}(t)y(t)]$ (see Exercise 8.13) when y solves Eq. (8.33).

Exercise 8.23 (Duhamel' s Principle II). Suppose that $A : \mathbb{R} \to L(X)$ is a continuous function and $V : \mathbb{R} \to L(X)$ is the unique solution to the linear differential equation in Eq. (8.31). Let $W_0 \in L(X)$ and $H \in C(\mathbb{R}, L(X))$ be given. Show that the unique solution to the differential equation:

$$\dot{W}(t) = A(t)W(t) + H(t) \text{ with } W(0) = W_0 \tag{8.35}$$

is given by

$$W(t) = V(t)W_0 + V(t)\int_0^t V(\tau)^{-1}H(\tau)\,d\tau. \tag{8.36}$$

9

Hölder Spaces as Banach Spaces

Notation 9.1 *Let Ω be an open subset of $\mathbb{R}^d$, $BC(\Omega)$ and $BC(\bar{\Omega})$ be the bounded continuous functions on Ω and $\bar{\Omega}$ respectively. By identifying $f \in BC(\bar{\Omega})$ with $f|_{\Omega} \in BC(\Omega)$, we will consider $BC(\bar{\Omega})$ as a subset of $BC(\Omega)$. For $u \in BC(\Omega)$ and $0 < \beta \leq 1$ let*

$$\|u\|_u := \sup_{x\in\Omega} |u(x)| \text{ and } [u]_\beta := \sup_{\substack{x,y\in\Omega \\ x\neq y}} \left\{ \frac{|u(x) - u(y)|}{|x-y|^\beta} \right\}.$$

*If $[u]_\beta < \infty$, then u is **Hölder continuous** with holder exponent*[1] *β. The collection of β – Hölder continuous function on Ω will be denoted by*

$$C^{0,\beta}(\Omega) := \{u \in BC(\Omega) : [u]_\beta < \infty\}$$

and for $u \in C^{0,\beta}(\Omega)$ let

$$\|u\|_{C^{0,\beta}(\Omega)} := \|u\|_u + [u]_\beta. \tag{9.1}$$

Remark 9.2. If $u : \Omega \to \mathbb{C}$ and $[u]_\beta < \infty$ for some $\beta > 1$, then u is constant on each connected component of Ω. Indeed, if $x \in \Omega$ and $h \in \mathbb{R}^d$ then

$$\left| \frac{u(x+th) - u(x)}{t} \right| \leq [u]_\beta t^\beta / t \to 0 \text{ as } t \to 0$$

which shows $\partial_h u(x) = 0$ for all $x \in \Omega$. If $y \in \Omega$ is in the same connected component as x, then by Exercise 22.8 below there exists a smooth curve $\sigma : [0,1] \to \Omega$ such that $\sigma(0) = x$ and $\sigma(1) = y$. So by the fundamental theorem of calculus and the chain rule,

$$u(y) - u(x) = \int_0^1 \frac{d}{dt} u(\sigma(t)) dt = \int_0^1 0 \; dt = 0.$$

This is why we do not talk about Hölder spaces with Hölder exponents larger than 1.

[1] If $\beta = 1$, u is is said to be Lipschitz continuous.

Lemma 9.3. *Suppose* $u \in C^1(\Omega) \cap BC(\Omega)$ *and* $\partial_i u \in BC(\Omega)$ *for* $i = 1, 2, \ldots, d$, *then* $u \in C^{0,1}(\Omega)$, *i.e.* $[u]_1 < \infty$.

The proof of this lemma is left to the reader as Exercise 9.1.

Theorem 9.4. *Let* Ω *be an open subset of* $\mathbb{R}^d$. *Then*

1. *Under the identification of* $u \in BC\left(\bar{\Omega}\right)$ *with* $u|_\Omega \in BC\left(\Omega\right)$, $BC(\bar{\Omega})$ *is a closed subspace of* $BC(\Omega)$.
2. *Every element* $u \in C^{0,\beta}(\Omega)$ *has a unique extension to a continuous function (still denoted by* u*) on* $\bar{\Omega}$. *Therefore we may identify* $C^{0,\beta}(\Omega)$ *with* $C^{0,\beta}(\bar{\Omega}) \subset BC(\bar{\Omega})$. *(In particular we may consider* $C^{0,\beta}(\Omega)$ *and* $C^{0,\beta}(\bar{\Omega})$ *to be the same when* $\beta > 0$.)
3. *The function* $u \in C^{0,\beta}(\Omega) \to \|u\|_{C^{0,\beta}(\Omega)} \in [0, \infty)$ *is a norm on* $C^{0,\beta}(\Omega)$ *which make* $C^{0,\beta}(\Omega)$ *into a Banach space.*

Proof. 1. The first item is trivial since for $u \in BC(\bar{\Omega})$, the sup-norm of u on $\bar{\Omega}$ agrees with the sup-norm on Ω and $BC(\bar{\Omega})$ is complete in this norm. **2.** Suppose that $[u]_\beta < \infty$ and $x_0 \in \mathrm{bd}(\Omega)$. Let $\{x_n\}_{n=1}^\infty \subset \Omega$ be a sequence such that $x_0 = \lim_{n\to\infty} x_n$. Then

$$|u(x_n) - u(x_m)| \le [u]_\beta \left|x_n - x_m\right|^\beta \to 0 \text{ as } m, n \to \infty$$

showing $\{u(x_n)\}_{n=1}^\infty$ is Cauchy so that $\bar{u}(x_0) := \lim_{n\to\infty} u(x_n)$ exists. If $\{y_n\}_{n=1}^\infty \subset \Omega$ is another sequence converging to x_0, then

$$|u(x_n) - u(y_n)| \le [u]_\beta \left|x_n - y_n\right|^\beta \to 0 \text{ as } n \to \infty,$$

showing $\bar{u}(x_0)$ is well defined. In this way we define $\bar{u}(x)$ for all $x \in \mathrm{bd}(\Omega)$ and let $\bar{u}(x) = u(x)$ for $x \in \Omega$. Since a similar limiting argument shows

$$|\bar{u}(x) - \bar{u}(y)| \le [u]_\beta \left|x - y\right|^\beta \text{ for all } x, y \in \bar{\Omega}$$

it follows that $\bar{u}$ is still continuous and $[\bar{u}]_\beta = [u]_\beta$. In the sequel we will abuse notation and simply denote $\bar{u}$ by u. **3.** For $u, v \in C^{0,\beta}(\Omega)$,

$$\begin{aligned}
[v+u]_\beta &= \sup_{\substack{x,y\in\Omega \\ x\neq y}} \left\{ \frac{|v(y) + u(y) - v(x) - u(x)|}{|x-y|^\beta} \right\} \\
&\le \sup_{\substack{x,y\in\Omega \\ x\neq y}} \left\{ \frac{|v(y) - v(x)| + |u(y) - u(x)|}{|x-y|^\beta} \right\} \le [v]_\beta + [u]_\beta
\end{aligned}$$

and for $\lambda \in \mathbb{C}$ it is easily seen that $[\lambda u]_\beta = |\lambda|\,[u]_\beta$. This shows $[\cdot]_\beta$ is a semi-norm on $C^{0,\beta}(\Omega)$ and therefore $\|\cdot\|_{C^{0,\beta}(\Omega)}$ defined in Eq. (9.1) is a norm. To see that $C^{0,\beta}(\Omega)$ is complete, let $\{u_n\}_{n=1}^\infty$ be a $C^{0,\beta}(\Omega)$–Cauchy sequence. Since $BC(\bar{\Omega})$ is complete, there exists $u \in BC(\bar{\Omega})$ such that $\|u - u_n\|_\infty \to 0$ as $n \to \infty$. For $x, y \in \Omega$ with $x \neq y$,

$$\frac{|u(x)-u(y)|}{|x-y|^{\beta}} = \lim_{n\to\infty} \frac{|u_n(x)-u_n(y)|}{|x-y|^{\beta}}$$
$$\le \limsup_{n\to\infty} [u_n]_\beta \le \lim_{n\to\infty} \|u_n\|_{C^{0,\beta}(\Omega)} < \infty,$$

and so we see that $u \in C^{0,\beta}(\Omega)$. Similarly,

$$\frac{|u(x)-u_n(x)-(u(y)-u_n(y))|}{|x-y|^{\beta}} = \lim_{m\to\infty} \frac{|(u_m-u_n)(x)-(u_m-u_n)(y)|}{|x-y|^{\beta}}$$
$$\le \limsup_{m\to\infty} [u_m-u_n]_\beta \to 0 \text{ as } n\to\infty,$$

showing $[u-u_n]_\beta \to 0$ as $n \to \infty$ and therefore $\lim_{n\to\infty} \|u-u_n\|_{C^{0,\beta}(\Omega)} = 0$. ∎

Notation 9.5 *Since Ω and $\bar{\Omega}$ are locally compact Hausdorff spaces, we may define $C_0(\Omega)$ and $C_0(\bar{\Omega})$ as in Definition 12.22. We will also let*

$$C_0^{0,\beta}(\Omega) := C^{0,\beta}(\Omega) \cap C_0(\Omega) \text{ and } C_0^{0,\beta}(\bar{\Omega}) := C^{0,\beta}(\Omega) \cap C_0(\bar{\Omega}).$$

It has already been shown in Proposition 12.23 that $C_0(\Omega)$ and $C_0(\bar{\Omega})$ are closed subspaces of $BC(\Omega)$ and $BC(\bar{\Omega})$ respectively. The next proposition describes the relation between $C_0(\Omega)$ and $C_0(\bar{\Omega})$.

Proposition 9.6. *Each $u \in C_0(\Omega)$ has a unique extension to a continuous function on $\bar{\Omega}$ given by $\bar{u} = u$ on Ω and $\bar{u} = 0$ on $\mathrm{bd}(\Omega)$ and the extension $\bar{u}$ is in $C_0(\bar{\Omega})$. Conversely if $u \in C_0(\bar{\Omega})$ and $u|_{\mathrm{bd}(\Omega)} = 0$, then $u|_\Omega \in C_0(\Omega)$. In this way we may identify $C_0(\Omega)$ with those $u \in C_0(\bar{\Omega})$ such that $u|_{\mathrm{bd}(\Omega)} = 0$.*

Proof. Any extension $u \in C_0(\Omega)$ to an element $\bar{u} \in C(\bar{\Omega})$ is necessarily unique, since Ω is dense inside $\bar{\Omega}$. So define $\bar{u} = u$ on Ω and $\bar{u} = 0$ on $\mathrm{bd}(\Omega)$. We must show $\bar{u}$ is continuous on $\bar{\Omega}$ and $\bar{u} \in C_0(\bar{\Omega})$. For the continuity assertion it is enough to show $\bar{u}$ is continuous at all points in $\mathrm{bd}(\Omega)$. For any $\varepsilon > 0$, by assumption, the set $K_\varepsilon := \{x \in \Omega : |u(x)| \ge \varepsilon\}$ is a compact subset of Ω. Since $\mathrm{bd}(\Omega) = \bar{\Omega} \setminus \Omega$, $\mathrm{bd}(\Omega) \cap K_\varepsilon = \emptyset$ and therefore the distance, $\delta := d(K_\varepsilon, \mathrm{bd}(\Omega))$, between K_ε and $\mathrm{bd}(\Omega)$ is positive. So if $x \in \mathrm{bd}(\Omega)$ and $y \in \bar{\Omega}$ and $|y-x| < \delta$, then $|\bar{u}(x) - \bar{u}(y)| = |u(y)| < \varepsilon$ which shows $\bar{u} : \bar{\Omega} \to \mathbb{C}$ is continuous. This also shows $\{|\bar{u}| \ge \varepsilon\} = \{|u| \ge \varepsilon\} = K_\varepsilon$ is compact in Ω and hence also in $\bar{\Omega}$. Since $\varepsilon > 0$ was arbitrary, this shows $\bar{u} \in C_0(\bar{\Omega})$. Conversely if $u \in C_0(\bar{\Omega})$ such that $u|_{\mathrm{bd}(\Omega)} = 0$ and $\varepsilon > 0$, then $K_\varepsilon := \left\{x \in \bar{\Omega} : |u(x)| \ge \varepsilon\right\}$ is a compact subset of $\bar{\Omega}$ which is contained in Ω since $\mathrm{bd}(\Omega) \cap K_\varepsilon = \emptyset$. Therefore K_ε is a compact subset of Ω showing $u|_\Omega \in C_0(\bar{\Omega})$. ∎

Definition 9.7. *Let Ω be an open subset of $\mathbb{R}^d$, $k \in \mathbb{N} \cup \{0\}$ and $\beta \in (0,1]$. Let $BC^k(\Omega)$ $(BC^k(\bar{\Omega}))$ denote the set of k – times continuously differentiable functions u on Ω such that $\partial^\alpha u \in BC(\Omega)$ $(\partial^\alpha u \in BC(\bar{\Omega}))^2$ for all $|\alpha| \le k$.*

[2] To say $\partial^\alpha u \in BC(\bar{\Omega})$ means that $\partial^\alpha u \in BC(\Omega)$ and $\partial^\alpha u$ extends to a continuous function on $\bar{\Omega}$.

Similarly, let $BC^{k,\beta}(\Omega)$ *denote those* $u \in BC^k(\Omega)$ *such that* $[\partial^\alpha u]_\beta < \infty$ *for all* $|\alpha| = k$. *For* $u \in BC^k(\Omega)$ *let*

$$\|u\|_{C^k(\Omega)} = \sum_{|\alpha|\le k} \|\partial^\alpha u\|_u \text{ and}$$

$$\|u\|_{C^{k,\beta}(\overline{\Omega})} = \sum_{|\alpha|\le k} \|\partial^\alpha u\|_u + \sum_{|\alpha|=k} [\partial^\alpha u]_\beta.$$

Theorem 9.8. *The spaces* $BC^k(\Omega)$ *and* $BC^{k,\beta}(\Omega)$ *equipped with* $\|\cdot\|_{C^k(\Omega)}$ *and* $\|\cdot\|_{C^{k,\beta}(\overline{\Omega})}$ *respectively are Banach spaces and* $BC^k(\bar{\Omega})$ *is a closed subspace of* $BC^k(\Omega)$ *and* $BC^{k,\beta}(\Omega) \subset BC^k(\bar{\Omega})$. *Also*

$$C_0^{k,\beta}(\Omega) = C_0^{k,\beta}(\bar{\Omega}) = \{u \in BC^{k,\beta}(\Omega) : \partial^\alpha u \in C_0(\Omega) \ \forall \ |\alpha| \le k\}$$

is a closed subspace of $BC^{k,\beta}(\Omega)$.

Proof. Suppose that $\{u_n\}_{n=1}^\infty \subset BC^k(\Omega)$ is a Cauchy sequence, then $\{\partial^\alpha u_n\}_{n=1}^\infty$ is a Cauchy sequence in $BC(\Omega)$ for $|\alpha| \le k$. Since $BC(\Omega)$ is complete, there exists $g_\alpha \in BC(\Omega)$ such that $\lim_{n\to\infty} \|\partial^\alpha u_n - g_\alpha\|_\infty = 0$ for all $|\alpha| \le k$. Letting $u := g_0$, we must show $u \in C^k(\Omega)$ and $\partial^\alpha u = g_\alpha$ for all $|\alpha| \le k$. This will be done by induction on $|\alpha|$. If $|\alpha| = 0$ there is nothing to prove. Suppose that we have verified $u \in C^l(\Omega)$ and $\partial^\alpha u = g_\alpha$ for all $|\alpha| \le l$ for some $l < k$. Then for $x \in \Omega$, $i \in \{1, 2, \dots, d\}$ and $t \in \mathbb{R}$ sufficiently small,

$$\partial^a u_n(x + te_i) = \partial^a u_n(x) + \int_0^t \partial_i \partial^a u_n(x + \tau e_i) d\tau.$$

Letting $n \to \infty$ in this equation gives

$$\partial^a u(x + te_i) = \partial^a u(x) + \int_0^t g_{\alpha+e_i}(x + \tau e_i) d\tau$$

from which it follows that $\partial_i \partial^\alpha u(x)$ exists for all $x \in \Omega$ and $\partial_i \partial^\alpha u = g_{\alpha+e_i}$. This completes the induction argument and also the proof that $BC^k(\Omega)$ is complete. It is easy to check that $BC^k(\bar{\Omega})$ is a closed subspace of $BC^k(\Omega)$ and by using Exercise 9.1 and Theorem 9.4 that that $BC^{k,\beta}(\Omega)$ is a subspace of $BC^k(\bar{\Omega})$. The fact that $C_0^{k,\beta}(\Omega)$ is a closed subspace of $BC^{k,\beta}(\Omega)$ is a consequence of Proposition 12.23. To prove $BC^{k,\beta}(\Omega)$ is complete, let $\{u_n\}_{n=1}^\infty \subset BC^{k,\beta}(\Omega)$ be a $\|\cdot\|_{C^{k,\beta}(\overline{\Omega})}$ – Cauchy sequence. By the completeness of $BC^k(\Omega)$ just proved, there exists $u \in BC^k(\Omega)$ such that $\lim_{n\to\infty} \|u - u_n\|_{C^k(\Omega)} = 0$. An application of Theorem 9.4 then shows $\lim_{n\to\infty} \|\partial^\alpha u_n - \partial^\alpha u\|_{C^{0,\beta}(\Omega)} = 0$ for $|\alpha| = k$ and therefore $\lim_{n\to\infty} \|u - u_n\|_{C^{k,\beta}(\overline{\Omega})} = 0$. ∎

The reader is asked to supply the proof of the following lemma.

Lemma 9.9. *The following inclusions hold. For any* $\beta \in [0, 1]$

$$BC^{k+1,0}(\Omega) \subset BC^{k,1}(\Omega) \subset BC^{k,\beta}(\Omega)$$

$$BC^{k+1,0}(\bar{\Omega}) \subset BC^{k,1}(\bar{\Omega}) \subset BC^{k,\beta}(\Omega).$$

9.1 Exercises

Exercise 9.1. Prove Lemma 9.3.

Part III

Topological Spaces

10

Topological Space Basics

Using the metric space results above as motivation we will axiomatize the notion of being an open set to more general settings.

Definition 10.1. *A collection of subsets τ of X is a **topology** if*

1. $\emptyset, X \in \tau$
2. *τ is closed under arbitrary unions, i.e. if $V_\alpha \in \tau$, for $\alpha \in I$ then $\bigcup_{\alpha \in I} V_\alpha \in \tau$.*
3. *τ is closed under finite intersections, i.e. if $V_1, \dots, V_n \in \tau$ then $V_1 \cap \dots \cap V_n \in \tau$.*
   *A pair (X, τ) where τ is a topology on X will be called a **topological space**.*

Notation 10.2 *Let (X, τ) be a topological space.*

1. *The elements, $V \in \tau$, are called **open** sets. We will often write $V \subset_o X$ to indicate V is an open subset of X.*
2. *A subset $F \subset X$ is **closed** if F^c is open and we will write $F \sqsubset X$ if F is a closed subset of X.*
3. *An **open neighborhood** of a point $x \in X$ is an open set $V \subset X$ such that $x \in V$. Let $\tau_x = \{V \in \tau : x \in V\}$ denote the collection of open neighborhoods of x.*
4. *A subset $W \subset X$ is a **neighborhood** of x if there exists $V \in \tau_x$ such that $V \subset W$.*
5. *A collection $\eta \subset \tau_x$ is called a **neighborhood base** at $x \in X$ if for all $V \in \tau_x$ there exists $W \in \eta$ such that $W \subset V$.*

The notation τ_x should not be confused with

$$\tau_{\{x\}} := i_{\{x\}}^{-1}(\tau) = \{\{x\} \cap V : V \in \tau\} = \{\emptyset, \{x\}\}.$$

Example 10.3. 1. Let (X, d) be a metric space, we write τ_d for the collection of d – open sets in X. We have already seen that τ_d is a topology, see Exercise 6.2. The collection of sets $\eta = \{B_x(\varepsilon) : \varepsilon \in \mathbb{D}\}$ where $\mathbb{D}$ is any dense subset of $(0, 1]$ is a neighborhood base at x.

2. Let X be any set, then $\tau = 2^X$ is a topology. In this topology all subsets of X are both open and closed. At the opposite extreme we have the **trivial** topology, $\tau = \{\emptyset, X\}$. In this topology only the empty set and X are open (closed).
3. Let $X = \{1, 2, 3\}$, then $\tau = \{\emptyset, X, \{2, 3\}\}$ is a topology on X which does not come from a metric.
4. Again let $X = \{1, 2, 3\}$. Then $\tau = \{\{1\}, \{2, 3\}, \emptyset, X\}$. is a topology, and the sets X, $\{1\}$, $\{2, 3\}, \emptyset$ are open and closed. The sets $\{1, 2\}$ and $\{1, 3\}$ are neither open nor closed.

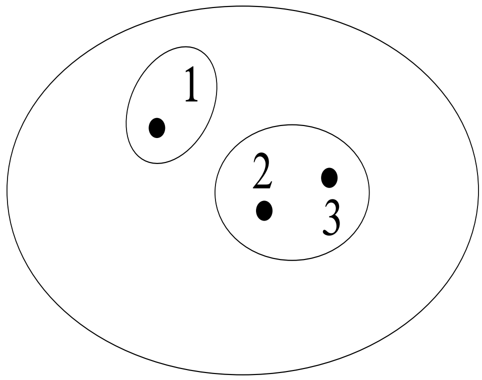


Fig. 10.1. A topology.

Definition 10.4. *Let (X, τ_X) and (Y, τ_Y) be topological spaces. A function $f : X \to Y$ is **continuous** if*

$$f^{-1}(\tau_Y) := \left\{f^{-1}(V) : V \in \tau_Y\right\} \subset \tau_X.$$

We will also say that f is τ_X/τ_Y –continuous or (τ_X, τ_Y) – continuous. Let $C(X, Y)$ denote the set of continuous functions from X to Y.

Exercise 10.1. Show $f : X \to Y$ is continuous iff $f^{-1}(C)$ is closed in X for all closed subsets C of Y.

Definition 10.5. *A map $f : X \to Y$ between topological spaces is called a **homeomorphism** provided that f is bijective, f is continuous and $f^{-1} : Y \to X$ is continuous. If there exists $f : X \to Y$ which is a homeomorphism, we say that X and Y are homeomorphic. (As topological spaces X and Y are essentially the same.)*

10.1 Constructing Topologies and Checking Continuity

Proposition 10.6. *Let $\mathcal{E}$ be any collection of subsets of X. Then there exists a unique smallest topology $\tau(\mathcal{E})$ which contains $\mathcal{E}$.*

Proof. Since 2^X is a topology and $\mathcal{E} \subset 2^X$, $\mathcal{E}$ is always a subset of a topology. It is now easily seen that

$$\tau(\mathcal{E}) := \bigcap\{\tau : \tau \text{ is a topology and } \mathcal{E} \subset \tau\}$$

is a topology which is clearly the smallest possible topology containing $\mathcal{E}$. ∎

The following proposition gives an explicit descriptions of $\tau(\mathcal{E})$.

Proposition 10.7. *Let X be a set and $\mathcal{E} \subset 2^X$. For simplicity of notation, assume that $X, \emptyset \in \mathcal{E}$. (If this is not the case simply replace $\mathcal{E}$ by $\mathcal{E} \cup \{X, \emptyset\}$.) Then*

$$\tau(\mathcal{E}) := \{\textit{arbitrary unions of finite intersections of elements from } \mathcal{E}\}. \tag{10.1}$$

Proof. Let τ be given as in the right side of Eq. (10.1). From the definition of a topology any topology containing $\mathcal{E}$ must contain τ and hence $\mathcal{E} \subset \tau \subset \tau(\mathcal{E})$. The proof will be completed by showing τ is a topology. The validation of τ being a topology is routine except for showing that τ is closed under taking finite intersections. Let $V, W \in \tau$ which by definition may be expressed as

$$V = \cup_{\alpha \in A} V_\alpha \text{ and } W = \cup_{\beta \in B} W_\beta,$$

where V_α and W_β are sets which are finite intersection of elements from $\mathcal{E}$. Then

$$V \cap W = (\cup_{\alpha \in A} V_\alpha) \cap (\cup_{\beta \in B} W_\beta) = \bigcup_{(\alpha,\beta) \in A \times B} V_\alpha \cap W_\beta.$$

Since for each $(\alpha, \beta) \in A \times B$, $V_\alpha \cap W_\beta$ is still a finite intersection of elements from $\mathcal{E}$, $V \cap W \in \tau$ showing τ is closed under taking finite intersections. ∎

Definition 10.8. *Let (X, τ) be a topological space. We say that $\mathcal{S} \subset \tau$ is a **sub-base** for the topology τ iff $\tau = \tau(\mathcal{S})$ and $X = \cup\mathcal{S} := \cup_{V \in \mathcal{S}} V$. We say $\mathcal{V} \subset \tau$ is a **base** for the topology τ iff $\mathcal{V}$ is a sub-base with the property that every element $V \in \tau$ may be written as*

$$V = \cup\{B \in \mathcal{V} : B \subset V\}.$$

Exercise 10.2. Suppose that $\mathcal{S}$ is a sub-base for a topology τ on a set X.

1. Show $\mathcal{V} := \mathcal{S}_f$ ($\mathcal{S}_f$ is the collection of finite intersections of elements from $\mathcal{S}$) is a base for τ.
2. Show $\mathcal{S}$ is itself a base for τ iff

$$V_1 \cap V_2 = \cup\{S \in \mathcal{S} : S \subset V_1 \cap V_2\}.$$

for every pair of sets $V_1, V_2 \in \mathcal{S}$.

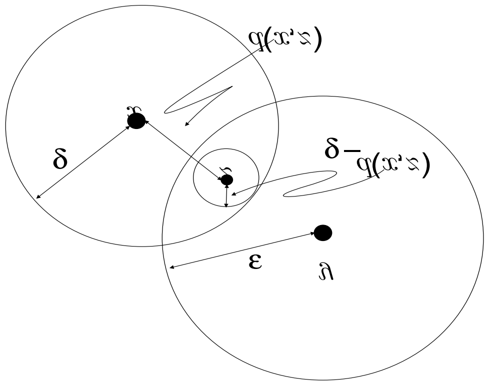


Fig. 10.2. Fitting balls in the intersection.

Remark 10.9. Let (X,d) be a metric space, then $\mathcal{E} = \{B_x(\delta) : x \in X$ and $\delta > 0\}$ is a base for τ_d – the topology associated to the metric d. This is the content of Exercise 6.3.

Let us check directly that $\mathcal{E}$ is a base for a topology. Suppose that $x, y \in X$ and $\varepsilon, \delta > 0$. If $z \in B(x,\delta) \cap B(y,\varepsilon)$, then

$$B(z,\alpha) \subset B(x,\delta) \cap B(y,\varepsilon) \tag{10.2}$$

where $\alpha = \min\{\delta - d(x,z), \varepsilon - d(y,z)\}$, see Figure 10.2. This is a formal consequence of the triangle inequality. For example let us show that $B(z,\alpha) \subset B(x,\delta)$. By the definition of α, we have that $\alpha \le \delta - d(x,z)$ or that $d(x,z) \le \delta - \alpha$. Hence if $w \in B(z,\alpha)$, then

$$d(x,w) \le d(x,z) + d(z,w) \le \delta - \alpha + d(z,w) < \delta - \alpha + \alpha = \delta$$

which shows that $w \in B(x,\delta)$. Similarly we show that $w \in B(y,\varepsilon)$ as well.

Owing to Exercise 10.2, this shows $\mathcal{E}$ is a base for a topology. We do not need to use Exercise 10.2 here since in fact Equation (10.2) may be generalized to finite intersection of balls. Namely if $x_i \in X$, $\delta_i > 0$ and $z \in \cap_{i=1}^n B(x_i,\delta_i)$, then

$$B(z,\alpha) \subset \cap_{i=1}^n B(x_i,\delta_i) \tag{10.3}$$

where now $\alpha := \min\{\delta_i - d(x_i,z) : i = 1,2,\dots,n\}$. By Eq. (10.3) it follows that any finite intersection of open balls may be written as a union of open balls.

Exercise 10.3. Suppose $f : X \to Y$ is a function and τ_X and τ_Y are topologies on X and Y respectively. Show

$$f^{-1}\tau_Y := \left\{f^{-1}(V) \subset X : V \in \tau_Y\right\} \text{ and } f_*\tau_X := \left\{V \subset Y : f^{-1}(V) \in \tau_X\right\}$$

(as in Notation 2.7) are also topologies on X and Y respectively.

Remark 10.10. Let $f : X \to Y$ be a function. Given a topology $\tau_Y \subset 2^Y$, the topology $\tau_X := f^{-1}(\tau_Y)$ is the smallest topology on X such that f is (τ_X, τ_Y) - continuous. Similarly, if τ_X is a topology on X then $\tau_Y = f_* \tau_X$ is the largest topology on Y such that f is (τ_X, τ_Y) - continuous.

Definition 10.11. *Let (X, τ) be a topological space and A subset of X. The **relative topology** or **induced topology** on A is the collection of sets*

$$\tau_A = i_A^{-1}(\tau) = \{A \cap V : V \in \tau\},$$

where $i_A : A \to X$ be the inclusion map as in Definition 2.8.

Lemma 10.12. *The relative topology, τ_A, is a topology on A. Moreover a subset $B \subset A$ is τ_A – closed iff there is a τ – closed subset, C, of X such that $B = C \cap A$.*

Proof. The first assertion is a consequence of Exercise 10.3. For the second, $B \subset A$ is τ_A – closed iff $A \setminus B = A \cap V$ for some $V \in \tau$ which is equivalent to $B = A \setminus (A \cap V) = A \cap V^c$ for some $V \in \tau$. ∎

Exercise 10.4. Show if (X, d) is a metric space and $\tau = \tau_d$ is the topology coming from d, then $(\tau_d)_A$ is the topology induced by making A into a metric space using the metric $d|_{A \times A}$.

Lemma 10.13. *Suppose that (X, τ_X), (Y, τ_Y) and (Z, τ_Z) are topological spaces. If $f : (X, \tau_X) \to (Y, \tau_Y)$ and $g : (Y, \tau_Y) \to (Z, \tau_Z)$ are continuous functions then $g \circ f : (X, \tau_X) \to (Z, \tau_Z)$ is continuous as well.*

Proof. This is easy since by assumption $g^{-1}(\tau_Z) \subset \tau_Y$ and $f^{-1}(\tau_Y) \subset \tau_X$ so that

$$(g \circ f)^{-1}(\tau_Z) = f^{-1}\left(g^{-1}(\tau_Z)\right) \subset f^{-1}(\tau_Y) \subset \tau_X.$$

∎

The following elementary lemma turns out to be extremely useful because it may be used to greatly simplify the verification that a given function is continuous.

Lemma 10.14. *Suppose that $f : X \to Y$ is a function, $\mathcal{E} \subset 2^Y$ and $A \subset Y$, then*

$$\tau\left(f^{-1}(\mathcal{E})\right) = f^{-1}(\tau(\mathcal{E})) \text{ and} \tag{10.4}$$

$$\tau(\mathcal{E}_A) = (\tau(\mathcal{E}))_A. \tag{10.5}$$

Moreover, if $\tau_Y = \tau(\mathcal{E})$ and τ_X is a topology on X, then f is (τ_X, τ_Y) – continuous iff $f^{-1}(\mathcal{E}) \subset \tau_X$.

Proof. We will give two proof of Eq. (10.4). The first proof is more constructive than the second, but the second proof will work in the context of σ – algebras to be developed later. **First Proof.** There is no harm (as the reader should verify) in replacing $\mathcal{E}$ by $\mathcal{E} \cup \{\emptyset, Y\}$ if necessary so that we may assume that $\emptyset, Y \in \mathcal{E}$. By Proposition 10.7, the general element V of $\tau(\mathcal{E})$ is an arbitrary unions of finite intersections of elements from $\mathcal{E}$. Since f^{-1} preserves all of the set operations, it follows that $f^{-1}\tau(\mathcal{E})$ consists of sets which are arbitrary unions of finite intersections of elements from $f^{-1}\mathcal{E}$, which is precisely $\tau\left(f^{-1}(\mathcal{E})\right)$ by another application of Proposition 10.7. **Second Proof.** By Exercise 10.3, $f^{-1}(\tau(\mathcal{E}))$ is a topology and since $\mathcal{E} \subset \tau(\mathcal{E})$, $f^{-1}(\mathcal{E}) \subset f^{-1}(\tau(\mathcal{E}))$. It now follows that $\tau(f^{-1}(\mathcal{E})) \subset f^{-1}(\tau(\mathcal{E}))$. For the reverse inclusion notice that

$$f_*\tau\left(f^{-1}(\mathcal{E})\right) = \left\{B \subset Y : f^{-1}(B) \in \tau\left(f^{-1}(\mathcal{E})\right)\right\}$$

is a topology which contains $\mathcal{E}$ and thus $\tau(\mathcal{E}) \subset f_*\tau\left(f^{-1}(\mathcal{E})\right)$. Hence if $B \in \tau(\mathcal{E})$ we know that $f^{-1}(B) \in \tau\left(f^{-1}(\mathcal{E})\right)$, i.e. $f^{-1}(\tau(\mathcal{E})) \subset \tau\left(f^{-1}(\mathcal{E})\right)$ and Eq. (10.4) has been proved. Applying Eq. (10.4) with $X = A$ and $f = i_A$ being the inclusion map implies

$$(\tau(\mathcal{E}))_A = i_A^{-1}(\tau(\mathcal{E})) = \tau(i_A^{-1}(\mathcal{E})) = \tau(\mathcal{E}_A).$$

Lastly if $f^{-1}\mathcal{E} \subset \tau_X$, then $f^{-1}\tau(\mathcal{E}) = \tau\left(f^{-1}\mathcal{E}\right) \subset \tau_X$ which shows f is (τ_X, τ_Y) – continuous. ∎

Corollary 10.15. *If (X, τ) is a topological space and $f : X \to \mathbb{R}$ is a function then the following are equivalent:*

1. *f is $(\tau, \tau_{\mathbb{R}})$ - continuous,*
2. *$f^{-1}((a,b)) \in \tau$ for all $-\infty < a < b < \infty$,*
3. *$f^{-1}((a,\infty)) \in \tau$ and $f^{-1}((-\infty,b)) \in \tau$ for all $a, b \in \mathbb{Q}$.*

(We are using $\tau_{\mathbb{R}}$ to denote the standard topology on $\mathbb{R}$ induced by the metric $d(x,y) = |x - y|$.)

Proof. Apply Lemma 10.14 with appropriate choices of $\mathcal{E}$. ∎

Definition 10.16. *Let (X, τ_X) and (Y, τ_Y) be topological spaces. A function $f : X \to Y$ is **continuous at a point** $x \in X$ if for every open neighborhood V of $f(x)$ there is an open neighborhood U of x such that $U \subset f^{-1}(V)$. See Figure 10.3.*

Exercise 10.5. Show $f : X \to Y$ is continuous (Definition 10.16) iff f is continuous at all points $x \in X$.

Definition 10.17. *Given topological spaces (X, τ) and (Y, τ') and a subset $A \subset X$. We say a function $f : A \to Y$ is **continuous** iff f is τ_A/τ' – continuous.*

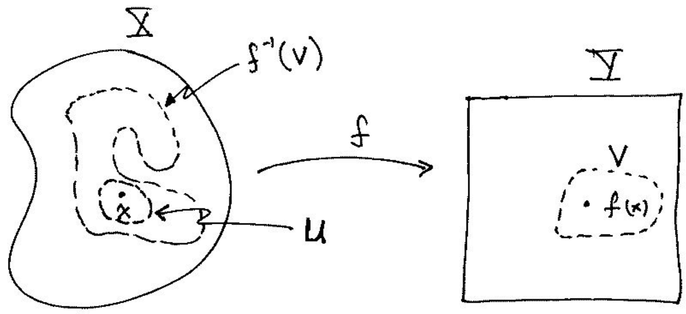


Fig. 10.3. Checking that a function is continuous at $x \in X$.

Definition 10.18. *Let (X, τ) be a topological space and $A \subset X$. A collection of subsets $\mathcal{U} \subset \tau$ is an* ***open cover*** *of A if $A \subset \bigcup \mathcal{U} := \bigcup_{U \in \mathcal{U}} U$.*

Proposition 10.19 (Localizing Continuity). *Let (X, τ) and (Y, τ') be topological spaces and $f : X \to Y$ be a function.*

1. *If f is continuous and $A \subset X$ then $f|_A : A \to Y$ is continuous.*
2. *Suppose there exist an open cover, $\mathcal{U} \subset \tau$, of X such that $f|_A$ is continuous for all $A \in \mathcal{U}$, then f is continuous.*

Proof. 1. If $f : X \to Y$ is a continuous, $f^{-1}(V) \in \tau$ for all $V \in \tau'$ and therefore

$$f|_A^{-1}(V) = A \cap f^{-1}(V) \in \tau_A \text{ for all } V \in \tau'.$$

2. Let $V \in \tau'$, then

$$f^{-1}(V) = \cup_{A \in \mathcal{U}} \left(f^{-1}(V) \cap A \right) = \cup_{A \in \mathcal{U}} f|_A^{-1}(V). \tag{10.6}$$

Since each $A \in \mathcal{U}$ is open, $\tau_A \subset \tau$ and by assumption, $f|_A^{-1}(V) \in \tau_A \subset \tau$. Hence Eq. (10.6) shows $f^{-1}(V)$ is a union of τ – open sets and hence is also τ – open. ■

Exercise 10.6 (A Baby Extension Theorem). Suppose $V \in \tau$ and $f : V \to \mathbb{C}$ is a continuous function. Further assume there is a closed subset C such that $\{x \in V : f(x) \neq 0\} \subset C \subset V$, then $F : X \to \mathbb{C}$ defined by

$$F(x) = \begin{cases} f(x) & \text{if } x \in V \\ 0 & \text{if } x \notin V \end{cases}$$

is continuous.

Exercise 10.7 (Building Continuous Functions). Prove the following variant of item 2. of Proposition 10.19. Namely, suppose there exists a **finite** collection $\mathcal{F}$ of closed subsets of X such that $X = \cup_{A \in \mathcal{F}} A$ and $f|_A$ is continuous for all $A \in \mathcal{F}$, then f is continuous. Given an example showing that the assumption that $\mathcal{F}$ is finite can not be eliminated. **Hint:** consider $f^{-1}(C)$ where C is a closed subset of Y.

10.2 Product Spaces I

Definition 10.20. *Let X be a set and suppose there is a collection of topological spaces $\{(Y_\alpha, \tau_\alpha) : \alpha \in A\}$ and functions $f_\alpha : X \to Y_\alpha$ for all $\alpha \in A$. Let $\tau(f_\alpha : \alpha \in A)$ denote the smallest topology on X such that each f_α is continuous, i.e.*

$$\tau(f_\alpha : \alpha \in A) = \tau(\cup_\alpha f_\alpha^{-1}(\tau_\alpha)).$$

Proposition 10.21 (Topologies Generated by Functions). *Assuming the notation in Definition 10.20 and additionally let (Z, τ_Z) be a topological space and $g : Z \to X$ be a function. Then g is $(\tau_Z, \tau(f_\alpha : \alpha \in A))$ – continuous iff $f_\alpha \circ g$ is (τ_Z, τ_α)–continuous for all $\alpha \in A$.*

Proof. ($\Rightarrow$) If g is $(\tau_Z, \tau(f_\alpha : \alpha \in A))$ – continuous, then the composition $f_\alpha \circ g$ is (τ_Z, τ_α) – continuous by Lemma 10.13. ($\Leftarrow$) Let

$$\tau_X = \tau(f_\alpha : \alpha \in A) = \tau\left(\cup_{\alpha \in A} f_\alpha^{-1}(\tau_\alpha)\right).$$

If $f_\alpha \circ g$ is (τ_Z, τ_α) – continuous for all α, then

$$g^{-1} f_\alpha^{-1}(\tau_\alpha) \subset \tau_Z \, \forall \, \alpha \in A$$

and therefore

$$g^{-1}\left(\cup_{\alpha \in A} f_\alpha^{-1}(\tau_\alpha)\right) = \cup_{\alpha \in A} g^{-1} f_\alpha^{-1}(\tau_\alpha) \subset \tau_Z$$

Hence

$$g^{-1}(\tau_X) = g^{-1}\left(\tau\left(\cup_{\alpha \in A} f_\alpha^{-1}(\tau_\alpha)\right)\right) = \tau(g^{-1}\left(\cup_{\alpha \in A} f_\alpha^{-1}(\tau_\alpha)\right) \subset \tau_Z$$

which shows that g is (τ_Z, τ_X) – continuous. ■

Let $\{(X_\alpha, \tau_\alpha)\}_{\alpha \in A}$ be a collection of topological spaces, $X = X_A = \prod_{\alpha \in A} X_\alpha$ and $\pi_\alpha : X_A \to X_\alpha$ be the canonical projection map as in Notation 2.2.

Definition 10.22. *The* ***product topology*** $\tau = \otimes_{\alpha \in A} \tau_\alpha$ *is the smallest topology on X_A such that each projection π_α is continuous. Explicitly, τ is the topology generated by the collection of sets,*

$$\mathcal{E} = \{\pi_\alpha^{-1}(V_\alpha) : \alpha \in A, V_\alpha \in \tau_\alpha\} = \cup_{\alpha \in A} \pi^{-1} \tau_\alpha. \tag{10.7}$$

Applying Proposition 10.21 in this setting implies the following proposition.

Proposition 10.23. *Suppose Y is a topological space and $f : Y \to X_A$ is a map. Then f is continuous iff $\pi_\alpha \circ f : Y \to X_\alpha$ is continuous for all $\alpha \in A$. In particular if $A = \{1, 2, \ldots, n\}$ so that $X_A = X_1 \times X_2 \times \cdots \times X_n$ and $f(y) = (f_1(y), f_2(y), \ldots, f_n(y)) \in X_1 \times X_2 \times \cdots \times X_n$, then $f : Y \to X_A$ is continuous iff $f_i : Y \to X_i$ is continuous for all i.*

Proposition 10.24. *Suppose that (X, τ) is a topological space and $\{f_n\} \subset X^A$ (see Notation 2.2) is a sequence. Then $f_n \to f$ in the product topology of X^A iff $f_n(\alpha) \to f(\alpha)$ for all $\alpha \in A$.*

Proof. Since π_α is continuous, if $f_n \to f$ then $f_n(\alpha) = \pi_\alpha(f_n) \to \pi_\alpha(f) = f(\alpha)$ for all $\alpha \in A$. Conversely, $f_n(\alpha) \to f(\alpha)$ for all $\alpha \in A$ iff $\pi_\alpha(f_n) \to \pi_\alpha(f)$ for all $\alpha \in A$. Therefore if $V = \pi_\alpha^{-1}(V_\alpha) \in \mathcal{E}$ (with $\mathcal{E}$ as in Eq. (10.7)) and $f \in V$, then $\pi_\alpha(f) \in V_\alpha$ and $\pi_\alpha(f_n) \in V_\alpha$ for a.a. n and hence $f_n \in V$ for a.a. n. This shows that $f_n \to f$ as $n \to \infty$. ■

Proposition 10.25. *Suppose that $(X_\alpha, \tau_\alpha)_{\alpha \in A}$ is a collection of topological spaces and $\otimes_{\alpha \in A} \tau_\alpha$ is the product topology on $X := \prod_{\alpha \in A} X_\alpha$.*

1. *If $\mathcal{E}_\alpha \subset \tau_\alpha$ generates τ_α for each $\alpha \in A$, then*
$$\otimes_{\alpha \in A} \tau_\alpha = \tau\left(\cup_{\alpha \in A} \pi_\alpha^{-1}(\mathcal{E}_\alpha)\right) \tag{10.8}$$
2. *If $\mathcal{B}_\alpha \subset \tau_\alpha$ is a base for τ_α for each α, then the collection of sets, $\mathcal{V}$, of the form*
$$V = \cap_{\alpha \in \Lambda} \pi_\alpha^{-1} V_\alpha = \prod_{\alpha \in \Lambda} V_\alpha \times \prod_{\alpha \notin \Lambda} X_\alpha =: V_\Lambda \times X_{A \setminus \Lambda}, \tag{10.9}$$
where $\Lambda \subset\subset A$ and $V_\alpha \in \mathcal{B}_\alpha$ for all $\alpha \in \Lambda$ is base for $\otimes_{\alpha \in A} \tau_\alpha$.

Proof. 1. Since
$$\begin{aligned}\cup_\alpha \pi_\alpha^{-1} \mathcal{E}_\alpha \subset \cup_\alpha \pi_\alpha^{-1} \tau_\alpha &= \cup_\alpha \pi_\alpha^{-1}(\tau(\mathcal{E}_\alpha)) \\ &= \cup_\alpha \tau(\pi_\alpha^{-1} \mathcal{E}_\alpha) \subset \tau\left(\cup_\alpha \pi_\alpha^{-1} \mathcal{E}_\alpha\right),\end{aligned}$$
it follows that
$$\tau\left(\cup_\alpha \pi_\alpha^{-1} \mathcal{E}_\alpha\right) \subset \otimes_\alpha \tau_\alpha \subset \tau\left(\cup_\alpha \pi_\alpha^{-1} \mathcal{E}_\alpha\right).$$
2. Now let $\mathcal{U} = \left[\cup_\alpha \pi_\alpha^{-1} \tau_\alpha\right]_f$ denote the collection of sets consisting of finite intersections of elements from $\cup_\alpha \pi_\alpha^{-1} \tau_\alpha$. Notice that $\mathcal{U}$ may be described as those sets in Eq. (10.9) where $V_\alpha \in \tau_\alpha$ for all $\alpha \in \Lambda$. By Exercise 10.2, $\mathcal{U}$ is a base for the product topology, $\otimes_{\alpha \in A} \tau_\alpha$. Hence for $W \in \otimes_{\alpha \in A} \tau_\alpha$ and $x \in W$, there exists a $V \in \mathcal{U}$ of the form in Eq. (10.9) such that $x \in V \subset W$. Since $\mathcal{B}_\alpha$ is a base for τ_α, there exists $U_\alpha \in \mathcal{B}_\alpha$ such that $x_\alpha \in U_\alpha \subset V_\alpha$ for each $\alpha \in \Lambda$. With this notation, the set $U_\Lambda \times X_{A \setminus \Lambda} \in \mathcal{V}$ and $x \in U_\Lambda \times X_{A \setminus \Lambda} \subset V \subset W$. This shows that every open set in X may be written as a union of elements from $\mathcal{V}$, i.e. $\mathcal{V}$ is a base for the product topology. ■

Notation 10.26 *Let $\mathcal{E}_i \subset 2^{X_i}$ be a collection of subsets of a set X_i for each $i = 1, 2, \ldots, n$. We will write, by abuse of notation, $\mathcal{E}_1 \times \mathcal{E}_2 \times \cdots \times \mathcal{E}_n$ for the collection of subsets of $X_1 \times \cdots \times X_n$ of the form $A_1 \times A_2 \times \cdots \times A_n$ with $A_i \in \mathcal{E}_i$ for all i. That is we are identifying $(A_1, A_2, \ldots, A_n)$ with $A_1 \times A_2 \times \cdots \times A_n$.*

Corollary 10.27. *Suppose* $A = \{1, 2, \ldots, n\}$ *so* $X = X_1 \times X_2 \times \cdots \times X_n$.

1. *If* $\mathcal{E}_i \subset 2^{X_i}$, $\tau_i = \tau(\mathcal{E}_i)$ *and* $X_i \in \mathcal{E}_i$ *for each* i, *then*
$$\tau_1 \otimes \tau_2 \otimes \cdots \otimes \tau_n = \tau(\mathcal{E}_1 \times \mathcal{E}_2 \times \cdots \times \mathcal{E}_n) \tag{10.10}$$
and in particular
$$\tau_1 \otimes \tau_2 \otimes \cdots \otimes \tau_n = \tau(\tau_1 \times \cdots \times \tau_n). \tag{10.11}$$
2. *Furthermore if* $\mathcal{B}_i \subset \tau_i$ *is a base for the topology* τ_i *for each* i, *then* $\mathcal{B}_1 \times \cdots \times \mathcal{B}_n$ *is a base for the product topology,* $\tau_1 \otimes \tau_2 \otimes \cdots \otimes \tau_n$.

Proof. (The proof is a minor variation on the proof of Proposition 10.25.) 1. Let $\left[\cup_{i \in A} \pi_i^{-1}(\mathcal{E}_i)\right]_f$ denotes the collection of sets which are finite intersections from $\cup_{i \in A} \pi_i^{-1}(\mathcal{E}_i)$, then, using $X_i \in \mathcal{E}_i$ for all i,
$$\cup_{i \in A} \pi_i^{-1}(\mathcal{E}_i) \subset \mathcal{E}_1 \times \mathcal{E}_2 \times \cdots \times \mathcal{E}_n \subset \left[\cup_{i \in A} \pi_i^{-1}(\mathcal{E}_i)\right]_f.$$
Therefore
$$\tau = \tau\left(\cup_{i \in A} \pi_i^{-1}(\mathcal{E}_i)\right) \subset \tau\left(\mathcal{E}_1 \times \mathcal{E}_2 \times \cdots \times \mathcal{E}_n\right) \subset \tau\left(\left[\cup_{i \in A} \pi_i^{-1}(\mathcal{E}_i)\right]_f\right) = \tau.$$
2. Observe that $\tau_1 \times \cdots \times \tau_n$ is closed under finite intersections and generates $\tau_1 \otimes \tau_2 \otimes \cdots \otimes \tau_n$, therefore $\tau_1 \times \cdots \times \tau_n$ is a base for the product topology. The proof that $\mathcal{B}_1 \times \cdots \times \mathcal{B}_n$ is also a base for $\tau_1 \otimes \tau_2 \otimes \cdots \otimes \tau_n$ follows the same method used to prove item 2. in Proposition 10.25. ■

Lemma 10.28. *Let* (X_i, d_i) *for* $i = 1, \ldots, n$ *be metric spaces,* $X := X_1 \times \cdots \times X_n$ *and for* $x = (x_1, x_2, \ldots, x_n)$ *and* $y = (y_1, y_2, \ldots, y_n)$ *in* X *let*
$$d(x, y) = \sum_{i=1}^n d_i(x_i, y_i). \tag{10.12}$$
Then the topology, τ_d, *associated to the metric* d *is the product topology on* X, *i.e.*
$$\tau_d = \tau_{d_1} \otimes \tau_{d_2} \otimes \cdots \otimes \tau_{d_n}.$$

Proof. Let $\rho(x, y) = \max\{d_i(x_i, y_i) : i = 1, 2, \ldots, n\}$. Then ρ is equivalent to d and hence $\tau_\rho = \tau_d$. Moreover if $\varepsilon > 0$ and $x = (x_1, x_2, \ldots, x_n) \in X$, then
$$B_x^\rho(\varepsilon) = B_{x_1}^{d_1}(\varepsilon) \times \cdots \times B_{x_n}^{d_n}(\varepsilon).$$
By Remark 10.9,
$$\mathcal{E} := \{B_x^\rho(\varepsilon) : x \in X \text{ and } \varepsilon > 0\}$$
is a base for τ_ρ and by Proposition 10.25 $\mathcal{E}$ is also a base for $\tau_{d_1} \otimes \tau_{d_2} \otimes \cdots \otimes \tau_{d_n}$. Therefore,
$$\tau_{d_1} \otimes \tau_{d_2} \otimes \cdots \otimes \tau_{d_n} = \tau(\mathcal{E}) = \tau_\rho = \tau_d.$$
■

10.3 Closure operations

Definition 10.29. *Let (X, τ) be a topological space and A be a subset of X.*

1. *The* ***closure*** *of A is the smallest closed set $\bar{A}$ containing A, i.e.*
$$\bar{A} := \cap \{F : A \subset F \sqsubset X\}.$$
(Because of Exercise 6.4 this is consistent with Definition 6.10 for the closure of a set in a metric space.)
2. *The* ***interior*** *of A is the largest open set A^o contained in A, i.e.*
$$A^o = \cup \{V \in \tau : V \subset A\}.$$
(With this notation the definition of a neighborhood of $x \in X$ may be stated as: $A \subset X$ is a neighborhood of a point $x \in X$ if $x \in A^o$.)
3. *The* ***accumulation points*** *of A is the set*
$$\mathrm{acc}(A) = \{x \in X : V \cap A \setminus \{x\} \neq \emptyset \text{ for all } V \in \tau_x\}.$$
4. *The* ***boundary*** *of A is the set $\mathrm{bd}(A) := \bar{A} \setminus A^o$.*

Remark 10.30. The relationships between the interior and the closure of a set are:
$$(A^o)^c = \bigcap \{V^c : V \in \tau \text{ and } V \subset A\} = \bigcap \{C : C \text{ is closed } C \supset A^c\} = \overline{A^c}$$
and similarly, $(\bar{A})^c = (A^c)^o$. Hence the boundary of A may be written as
$$\mathrm{bd}(A) := \bar{A} \setminus A^o = \bar{A} \cap (A^o)^c = \bar{A} \cap \overline{A^c}, \tag{10.13}$$
which is to say $\mathrm{bd}(A)$ consists of the points in both the closure of A and A^c.

Proposition 10.31. *Let $A \subset X$ and $x \in X$.*

1. *If $V \subset_o X$ and $A \cap V = \emptyset$ then $\bar{A} \cap V = \emptyset$.*
2. *$x \in \bar{A}$ iff $V \cap A \neq \emptyset$ for all $V \in \tau_x$.*
3. *$x \in \mathrm{bd}(A)$ iff $V \cap A \neq \emptyset$ and $V \cap A^c \neq \emptyset$ for all $V \in \tau_x$.*
4. *$\bar{A} = A \cup \mathrm{acc}(A)$.*

Proof. 1. Since $A \cap V = \emptyset$, $A \subset V^c$ and since V^c is closed, $\bar{A} \subset V^c$. That is to say $\bar{A} \cap V = \emptyset$. 2. By Remark 10.30[1], $\bar{A} = ((A^c)^o)^c$ so $x \in \bar{A}$ iff $x \notin (A^c)^o$ which happens iff $V \not\subset A^c$ for all $V \in \tau_x$, i.e. iff $V \cap A \neq \emptyset$ for all $V \in \tau_x$. 3. This assertion easily follows from the Item 2. and Eq. (10.13). 4. Item 4. is an easy consequence of the definition of $\mathrm{acc}(A)$ and item 2. ■

[1] Here is another direct proof of item 2. which goes by showing $x \notin \bar{A}$ iff there exists $V \in \tau_x$ such that $V \cap A = \emptyset$. If $x \notin \bar{A}$ then $V = \left(\bar{A}\right)^c \in \tau_x$ and $V \cap A \subset V \cap \bar{A} = \emptyset$. Conversely if there exists $V \in \tau_x$ such that $A \cap V = \emptyset$ then by Item 1. $\bar{A} \cap V = \emptyset$.

Lemma 10.32. *Let $A \subset Y \subset X$, $\bar{A}^Y$ denote the closure of A in Y with its relative topology and $\bar{A} = \bar{A}^X$ be the closure of A in X, then $\bar{A}^Y = \bar{A}^X \cap Y$.*

Proof. Using Lemma 10.12,

$$\begin{aligned}\bar{A}^Y &= \cap\{B \sqsubset Y : A \subset B\} = \cap\{C \cap Y : A \subset C \sqsubset X\} \\ &= Y \cap (\cap\{C : A \subset C \sqsubset X\}) = Y \cap \bar{A}^X.\end{aligned}$$

Alternative proof. Let $x \in Y$ then $x \in \bar{A}^Y$ iff $V \cap A \neq \emptyset$ for all $V \in \tau_Y$ such that $x \in V$. This happens iff for all $U \in \tau_x$, $U \cap Y \cap A = U \cap A \neq \emptyset$ which happens iff $x \in \bar{A}^X$. That is to say $\bar{A}^Y = \bar{A}^X \cap Y$. ■

The support of a function may now be defined as in Definition 8.25 above.

Definition 10.33 (Support). *Let $f : X \to Y$ be a function from a topological space (X, τ_X) to a vector space Y. Then we define the support of f by*

$$\operatorname{supp}(f) := \overline{\{x \in X : f(x) \neq 0\}},$$

a closed subset of X.

The next result is included for completeness but will not be used in the sequel so may be omitted.

Lemma 10.34. *Suppose that $f : X \to Y$ is a map between topological spaces. Then the following are equivalent:*

1. *f is continuous.*
2. *$f(\bar{A}) \subset \overline{f(A)}$ for all $A \subset X$*
3. *$\overline{f^{-1}(B)} \subset f^{-1}(\bar{B})$ for all $B \subset X$.*

Proof. If f is continuous, then $f^{-1}\left(\overline{f(A)}\right)$ is closed and since $A \subset f^{-1}(f(A)) \subset f^{-1}\left(\overline{f(A)}\right)$ it follows that $\bar{A} \subset f^{-1}\left(\overline{f(A)}\right)$. From this equation we learn that $f(\bar{A}) \subset \overline{f(A)}$ so that (1) implies (2) Now assume (2), then for $B \subset Y$ (taking $A = f^{-1}(\bar{B})$) we have

$$f(\overline{f^{-1}(B)}) \subset f(\overline{f^{-1}(\bar{B})}) \subset \overline{f(f^{-1}(\bar{B}))} \subset \bar{B}$$

and therefore

$$\overline{f^{-1}(B)} \subset f^{-1}(\bar{B}). \tag{10.14}$$

This shows that (2) implies (3) Finally if Eq. (10.14) holds for all B, then when B is closed this shows that

$$\overline{f^{-1}(B)} \subset f^{-1}(\bar{B}) = f^{-1}(B) \subset \overline{f^{-1}(B)}$$

which shows that

$$f^{-1}(B) = \overline{f^{-1}(B)}.$$

Therefore $f^{-1}(B)$ is closed whenever B is closed which implies that f is continuous. ■

10.4 Countability Axioms

Definition 10.35. *Let (X, τ) be a topological space. A sequence $\{x_n\}_{n=1}^{\infty} \subset X$* ***converges*** *to a point $x \in X$ if for all $V \in \tau_x$, $x_n \in V$ almost always (abbreviated a.a.), i.e. $\#(\{n : x_n \notin V\}) < \infty$. We will write $x_n \to x$ as $n \to \infty$ or $\lim_{n\to\infty} x_n = x$ when x_n converges to x.*

Example 10.36. Let $X = \{1, 2, 3\}$ and $\tau = \{X, \emptyset, \{1, 2\}, \{2, 3\}, \{2\}\}$ and $x_n = 2$ for all n. Then $x_n \to x$ for every $x \in X$. So limits need not be unique!

Definition 10.37 (First Countable). *A topological space, (X, τ), is* ***first countable*** *iff every point $x \in X$ has a countable neighborhood base as defined in Notation 10.2*

Example 10.38. All metric spaces, (X, d), are first countable. Indeed, if $x \in X$ then $\left\{B\left(x, \frac{1}{n}\right) : n \in \mathbb{N}\right\}$ is a countable neigborhood base at $x \in X$.

Exercise 10.8. Suppose X is an uncountable set and let $V \in \tau$ iff V^c is finite or countable of $V = \emptyset$. Show τ is a topology on X which is closed under countable intersections and that (X, τ) is **not** first countable.

Exercise 10.9. Let $\{0, 1\}$ be equipped with the discrete topology and $X = \{0, 1\}^{\mathbb{R}}$ be equipped with the product topology, τ. Show (X, τ) is not first countable.

The spaces described in Exercises 10.8 and 10.9 are examples of topological spaces which are not metrizable, i.e. the topology is not induced by any metric on X. Like for metric spaces, when τ is first countable, we may formulate many topological notions in terms of sequences.

Proposition 10.39. *If $f : X \to Y$ is continuous at $x \in X$ and $\lim_{n\to\infty} x_n = x \in X$, then $\lim_{n\to\infty} f(x_n) = f(x) \in Y$. Moreover, if there exists a countable neighborhood base η of $x \in X$, then f is continuous at x iff $\lim\limits_{n\to\infty} f(x_n) = f(x)$ for all sequences $\{x_n\}_{n=1}^{\infty} \subset X$ such that $x_n \to x$ as $n \to \infty$.*

Proof. If $f : X \to Y$ is continuous and $W \in \tau_Y$ is a neighborhood of $f(x) \in Y$, then there exists a neighborhood V of $x \in X$ such that $f(V) \subset W$. Since $x_n \to x$, $x_n \in V$ a.a. and therefore $f(x_n) \in f(V) \subset W$ a.a., i.e. $f(x_n) \to f(x)$ as $n \to \infty$. Conversely suppose that $\eta := \{W_n\}_{n=1}^{\infty}$ is a countable neighborhood base at x and $\lim\limits_{n\to\infty} f(x_n) = f(x)$ for all sequences $\{x_n\}_{n=1}^{\infty} \subset X$ such that $x_n \to x$. By replacing W_n by $W_1 \cap \cdots \cap W_n$ if necessary, we may assume that $\{W_n\}_{n=1}^{\infty}$ is a decreasing sequence of sets. If f were **not** continuous at x then there exists $V \in \tau_{f(x)}$ such that $x \notin \left[f^{-1}(V)\right]^o$. Therefore, W_n is not a subset of $f^{-1}(V)$ for all n. Hence for each n, we may choose $x_n \in W_n \setminus f^{-1}(V)$. This sequence then has the property that $x_n \to x$ as $n \to \infty$ while $f(x_n) \notin V$ for all n and hence $\lim_{n\to\infty} f(x_n) \neq f(x)$. ■

Lemma 10.40. *Suppose there exists $\{x_n\}_{n=1}^{\infty} \subset A$ such that $x_n \to x$, then $x \in \bar{A}$. Conversely if (X, τ) is a first countable space (like a metric space) then if $x \in \bar{A}$ there exists $\{x_n\}_{n=1}^{\infty} \subset A$ such that $x_n \to x$.*

Proof. Suppose $\{x_n\}_{n=1}^{\infty} \subset A$ and $x_n \to x \in X$. Since $\bar{A}^c$ is an open set, if $x \in \bar{A}^c$ then $x_n \in \bar{A}^c \subset A^c$ a.a. contradicting the assumption that $\{x_n\}_{n=1}^{\infty} \subset A$. Hence $x \in \bar{A}$. For the converse we now assume that (X, τ) is first countable and that $\{V_n\}_{n=1}^{\infty}$ is a countable neighborhood base at x such that $V_1 \supset V_2 \supset V_3 \supset \dots$. By Proposition 10.31, $x \in \bar{A}$ iff $V \cap A \neq \emptyset$ for all $V \in \tau_x$. Hence $x \in \bar{A}$ implies there exists $x_n \in V_n \cap A$ for all n. It is now easily seen that $x_n \to x$ as $n \to \infty$. ∎

Definition 10.41. *A topological space, (X, τ), is **second countable** if there exists a countable base $\mathcal{V}$ for τ, i.e. $\mathcal{V} \subset \tau$ is a countable set such that for every $W \in \tau$,*

$$W = \cup\{V : V \in \mathcal{V} \ni V \subset W\}.$$

Definition 10.42. *A subset D of a topological space X is **dense** if $\bar{D} = X$. A topological space is said to be **separable** if it contains a countable dense subset, D.*

Example 10.43. The following are examples of countable dense sets.

1. The rational number $\mathbb{Q}$ are dense in $\mathbb{R}$ equipped with the usual topology.
2. More generally, $\mathbb{Q}^d$ is a countable dense subset of $\mathbb{R}^d$ for any $d \in \mathbb{N}$.
3. Even more generally, for any function $\mu : \mathbb{N} \to (0, \infty)$, $\ell^p(\mu)$ is separable for all $1 \leq p < \infty$. For example, let $\Gamma \subset \mathbb{F}$ be a countable dense set, then
$$D := \{x \in \ell^p(\mu) : x_i \in \Gamma \text{ for all } i \text{ and } \#\{j : x_j \neq 0\} < \infty\}.$$
The set Γ can be taken to be $\mathbb{Q}$ if $\mathbb{F} = \mathbb{R}$ or $\mathbb{Q} + i\mathbb{Q}$ if $\mathbb{F} = \mathbb{C}$.
4. If (X, d) is a metric space which is separable then every subset $Y \subset X$ is also separable in the induced topology.

To prove 4. above, let $A = \{x_n\}_{n=1}^{\infty} \subset X$ be a countable dense subset of X. Let $d_Y(x) = \inf\{d(x, y) : y \in Y\}$ be the distance from x to Y and recall that $d_Y : X \to [0, \infty)$ is continuous. Let $\varepsilon_n = \max\left\{d_Y(x_n), \frac{1}{n}\right\} \geq 0$ and for each n let $y_n \in B_{x_n}(2\varepsilon_n)$. Then if $y \in Y$ and $\varepsilon > 0$ we may choose $n \in \mathbb{N}$ such that $d(y, x_n) \leq \varepsilon_n < \varepsilon/3$. Then $d(y_n, x_n) \leq 2\varepsilon_n < 2\varepsilon/3$ and therefore

$$d(y, y_n) \leq d(y, x_n) + d(x_n, y_n) < \varepsilon.$$

This shows that $B := \{y_n\}_{n=1}^{\infty}$ is a countable dense subset of Y.

Exercise 10.10. Show $\ell^{\infty}(\mathbb{N})$ is not separable.

Exercise 10.11. Show every second countable topological space (X, τ) is separable. Show the converse is not true by showing $X := \mathbb{R}$ with $\tau = \{\emptyset\} \cup \{V \subset \mathbb{R} : 0 \in V\}$ is a separable, first countable but not a second countable topological space.

Exercise 10.12. Every separable metric space, (X,d) is second countable.

Exercise 10.13. Suppose $\mathcal{E} \subset 2^X$ is a countable collection of subsets of X, then $\tau = \tau(\mathcal{E})$ is a second countable topology on X.

10.5 Connectedness

Definition 10.44. *(X,τ) is **disconnected** if there exists non-empty open sets U and V of X such that $U \cap V = \emptyset$ and $X = U \cup V$. We say $\{U,V\}$ is a **disconnection** of X. The topological space (X,τ) is called **connected** if it is not disconnected, i.e. if there is no disconnection of X. If $A \subset X$ we say A is connected iff (A,τ_A) is connected where τ_A is the relative topology on A. Explicitly, A is disconnected in (X,τ) iff there exists $U,V \in \tau$ such that $U \cap A \neq \emptyset$, $U \cap A \neq \emptyset$, $A \cap U \cap V = \emptyset$ and $A \subset U \cup V$.*

The reader should check that the following statement is an equivalent definition of connectivity. A topological space (X,τ) is connected iff the only sets $A \subset X$ which are both open and closed are the sets X and $\emptyset$. This version of the definition is often used in practice.

Remark 10.45. Let $A \subset Y \subset X$. Then A is connected in X iff A is connected in Y.

Proof. Since

$$\tau_A := \{V \cap A : V \subset X\} = \{V \cap A \cap Y : V \subset X\} = \{U \cap A : U \subset_o Y\},$$

the relative topology on A inherited from X is the same as the relative topology on A inherited from Y. Since connectivity is a statement about the relative topologies on A, A is connected in X iff A is connected in Y. ■

The following elementary but important lemma is left as an exercise to the reader.

Lemma 10.46. *Suppose that $f : X \to Y$ is a continuous map between topological spaces. Then $f(X) \subset Y$ is connected if X is connected.*

Here is a typical way these connectedness ideas are used.

Example 10.47. Suppose that $f : X \to Y$ is a continuous map between two topological spaces, the space X is connected and the space Y is "T_1," i.e. $\{y\}$ is a closed set for all $y \in Y$ as in Definition 12.35 below. Further assume f is locally constant, i.e. for all $x \in X$ there exists an open neighborhood V of x in X such that $f|_V$ is constant. Then f is constant, i.e. $f(X) = \{y_0\}$ for some $y_0 \in Y$. To prove this, let $y_0 \in f(X)$ and let $W := f^{-1}(\{y_0\})$. Since $\{y_0\} \subset Y$ is a closed set and since f is continuous $W \subset X$ is also closed. Since f is locally constant, W is open as well and since X is connected it follows that $W = X$, i.e. $f(X) = \{y_0\}$.

As a concrete application of this result, suppose that X is a connected open subset of $\mathbb{R}^d$ and $f : X \to \mathbb{R}$ is a C^1 – function such that $\nabla f \equiv 0$. If $x \in X$ and $\varepsilon > 0$ such that $B(x, \varepsilon) \subset X$, we have, for any $|v| < \varepsilon$ and $t \in [-1, 1]$, that

$$\frac{d}{dt} f(x + tv) = \nabla f(x + tv) \cdot v = 0.$$

Therefore $f(x + v) = f(x)$ for all $|v| < \varepsilon$ and this shows f is locally constant. Hence, by what we have just proved, f is constant on X.

Theorem 10.48 (Properties of Connected Sets). *Let (X, τ) be a topological space.*

1. *If $B \subset X$ is a connected set and X is the disjoint union of two open sets U and V, then either $B \subset U$ or $B \subset V$.*
2. *If $A \subset X$ is connected,*
   a) *then $\bar{A}$ is connected.*
   b) *More generally, if A is connected and $B \subset \mathrm{acc}(A)$, then $A \cup B$ is connected as well. (Recall that $\mathrm{acc}(A)$ – the set of accumulation points of A was defined in Definition 10.29 above.)*
3. *If $\{E_\alpha\}_{\alpha \in A}$ is a collection of connected sets such that $\bigcap_{\alpha \in A} E_\alpha \neq \emptyset$, then $Y := \bigcup_{\alpha \in A} E_\alpha$ is connected as well.*
4. *Suppose $A, B \subset X$ are non-empty connected subsets of X such that $\bar{A} \cap B \neq \emptyset$, then $A \cup B$ is connected in X.*
5. *Every point $x \in X$ is contained in a unique maximal connected subset C_x of X and this subset is closed. The set C_x is called the **connected component** of x.*

Proof.

1. Since B is the disjoint union of the relatively open sets $B \cap U$ and $B \cap V$, we must have $B \cap U = B$ or $B \cap V = B$ for otherwise $\{B \cap U, B \cap V\}$ would be a disconnection of B.
2. a. Let $Y = \bar{A}$ be equipped with the relative topology from X. Suppose that $U, V \subset_o Y$ form a disconnection of $Y = \bar{A}$. Then by 1. either $A \subset U$ or $A \subset V$. Say that $A \subset U$. Since U is both open an closed in Y, it follows that $Y = \bar{A} \subset U$. Therefore $V = \emptyset$ and we have a contradiction to the assumption that $\{U, V\}$ is a disconnection of $Y = \bar{A}$. Hence we must conclude that $Y = \bar{A}$ is connected as well.
   b. Now let $Y = A \cup B$ with $B \subset \mathrm{acc}(A)$, then
   $$\bar{A}^Y = \bar{A} \cap Y = (A \cup \mathrm{acc}(A)) \cap Y = A \cup B.$$
   Because A is connected in Y, by (2a) $Y = A \cup B = \bar{A}^Y$ is also connected.
3. Let $Y := \bigcup_{\alpha \in A} E_\alpha$. By Remark 10.45, we know that E_α is connected in Y for each $\alpha \in A$. If $\{U, V\}$ were a disconnection of Y, by item (1), either $E_\alpha \subset U$ or $E_\alpha \subset V$ for all α. Let $\Lambda = \{\alpha \in A : E_\alpha \subset U\}$ then

$U = \cup_{\alpha \in \Lambda} E_\alpha$ and $V = \cup_{\alpha \in A \setminus \Lambda} E_\alpha$. (Notice that neither Λ or $A \setminus \Lambda$ can be empty since U and V are not empty.) Since

$$\emptyset = U \cap V = \bigcup\nolimits_{\alpha \in \Lambda, \beta \in \Lambda^c} (E_\alpha \cap E_\beta) \supset \bigcap_{\alpha \in A} E_\alpha \neq \emptyset.$$

we have reached a contradiction and hence no such disconnection exists.

4. (A good example to keep in mind here is $X = \mathbb{R}$, $A = (0,1)$ and $B = [1,2)$.) For sake of contradiction suppose that $\{U, V\}$ were a disconnection of $Y = A \cup B$. By item (1) either $A \subset U$ or $A \subset V$, say $A \subset U$ in which case $B \subset V$. Since $Y = A \cup B$ we must have $A = U$ and $B = V$ and so we may conclude: A and B are disjoint subsets of Y which are both open and closed. This implies
$$A = \bar{A}^Y = \bar{A} \cap Y = \bar{A} \cap (A \cup B) = A \cup \left(\bar{A} \cap B\right)$$
and therefore
$$\emptyset = A \cap B = \left[A \cup \left(\bar{A} \cap B\right)\right] \cap B = \bar{A} \cap B \neq \emptyset$$
which gives us the desired contradiction.
5. Let $\mathcal{C}$ denote the collection of connected subsets $C \subset X$ such that $x \in C$. Then by item 3., the set $C_x := \cup \mathcal{C}$ is also a connected subset of X which contains x and clearly this is the unique maximal connected set containing x. Since $\bar{C}_x$ is also connected by item (2) and C_x is maximal, $C_x = \bar{C}_x$, i.e. C_x is closed.

■

Theorem 10.49 (The Connected Subsets of $\mathbb{R}$). *The connected subsets of $\mathbb{R}$ are intervals.*

Proof. Suppose that $A \subset \mathbb{R}$ is a connected subset and that $a, b \in A$ with $a < b$. If there exists $c \in (a, b)$ such that $c \notin A$, then $U := (-\infty, c) \cap A$ and $V := (c, \infty) \cap A$ would form a disconnection of A. Hence $(a, b) \subset A$. Let $\alpha := \inf(A)$ and $\beta := \sup(A)$ and choose $\alpha_n, \beta_n \in A$ such that $\alpha_n < \beta_n$ and $\alpha_n \downarrow \alpha$ and $\beta_n \uparrow \beta$ as $n \to \infty$. By what we have just shown, $(\alpha_n, \beta_n) \subset A$ for all n and hence $(\alpha, \beta) = \cup_{n=1}^\infty (\alpha_n, \beta_n) \subset A$. From this it follows that $A = (\alpha, \beta)$, $[\alpha, \beta)$, $(\alpha, \beta]$ or $[\alpha, \beta]$, i.e. A is an interval.

Conversely suppose that A is an interval, and for sake of contradiction, suppose that $\{U, V\}$ is a disconnection of A with $a \in U$, $b \in V$. After relabelling U and V if necessary we may assume that $a < b$. Since A is an interval $[a, b] \subset A$. Let $p = \sup\left([a, b] \cap U\right)$, then because U and V are open, $a < p < b$. Now p can not be in U for otherwise $\sup\left([a, b] \cap U\right) > p$ and p can not be in V for otherwise $p < \sup\left([a, b] \cap U\right)$. From this it follows that $p \notin U \cup V$ and hence $A \neq U \cup V$ contradicting the assumption that $\{U, V\}$ is a disconnection. ■

Theorem 10.50 (Intermediate Value Theorem). *Suppose that (X, τ) is a connected topological space and $f : X \to \mathbb{R}$ is a continuous map. Then f satisfies the intermediate value property. Namely, for every pair $x, y \in X$ such that $f(x) < f(y)$ and $c \in (f(x), f(y))$, there exits $z \in X$ such that $f(z) = c$.*

Proof. By Lemma 10.46, $f(X)$ is connected subset of $\mathbb{R}$. So by Theorem 10.49, $f(X)$ is a subinterval of $\mathbb{R}$ and this completes the proof. ∎

Definition 10.51. *A topological space X is **path connected** if to every pair of points $\{x_0, x_1\} \subset X$ there exists a continuous **path,** $\sigma \in C([0,1], X)$, such that $\sigma(0) = x_0$ and $\sigma(1) = x_1$. The space X is said to be **locally path connected** if for each $x \in X$, there is an open neighborhood $V \subset X$ of x which is path connected.*

Proposition 10.52. *Let X be a topological space.*

1. *If X is path connected then X is connected.*
2. *If X is connected and locally path connected, then X is path connected.*
3. *If X is any connected open subset of $\mathbb{R}^n$, then X is path connected.*

Proof. The reader is asked to prove this proposition in Exercises 10.20 – 10.22 below. ∎

Proposition 10.53 (Stability of Connectedness Under Products). *Let (X_α, τ_α) be connected topological spaces. Then the product space $X_A = \prod_{\alpha \in A} X_\alpha$ equipped with the product topology is connected.*

Proof. Let us begin with the case of two factors, namely assume that X and Y are connected topological spaces, then we will show that $X \times Y$ is connected as well. Given $x \in X$, let $f_x : Y \to X \times Y$ be the map $f_x(y) = (x, y)$ and notice that f_x is continuous since $\pi_X \circ f_x(y) = x$ and $\pi_Y \circ f_x(y) = y$ are continuous maps. From this we conclude that $\{x\} \times Y = f_x(Y)$ is connected by Lemma 10.46. A similar argument shows $X \times \{y\}$ is connected for all $y \in Y$.

Let $p = (x_0, y_0) \in X \times Y$ and C_p denote the connected component of p. Since $\{x_0\} \times Y$ is connected and $p \in \{x_0\} \times Y$ it follows that $\{x_0\} \times Y \subset C_p$ and hence C_p is also the connected component (x_0, y) for all $y \in Y$. Similarly, $X \times \{y\} \subset C_{(x_0,y)} = C_p$ is connected, and therefore $X \times \{y\} \subset C_p$. So we have shown $(x, y) \in C_p$ for all $x \in X$ and $y \in Y$, see Figure 10.4. By induction the theorem holds whenever A is a finite set, i.e. for products of a finite number of connected spaces.

For the general case, again choose a point $p \in X_A = X^A$ and again let $C = C_p$ be the connected component of p. Recall that C_p is closed and therefore if C_p is a proper subset of X_A, then $X_A \setminus C_p$ is a non-empty open set. By the definition of the product topology, this would imply that $X_A \setminus C_p$ contains an open set of the form

$$V := \cap_{\alpha \in \Lambda} \pi_\alpha^{-1}(V_\alpha) = V_\Lambda \times X_{A \setminus \Lambda}$$

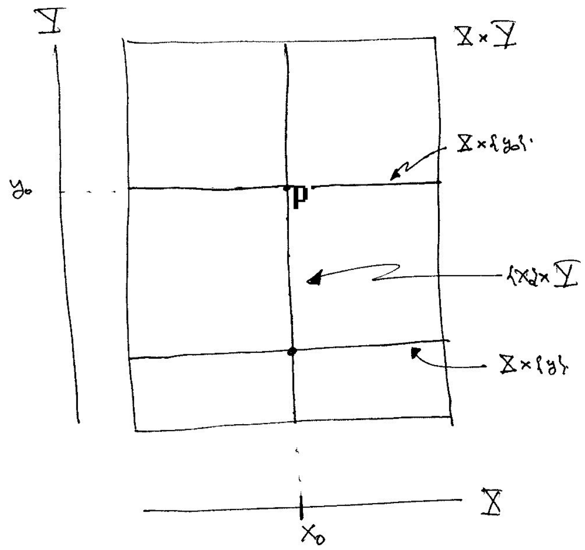


Fig. 10.4. This picture illustrates why the connected component of p in $X \times Y$ must contain all points of $X \times Y$.

where $\Lambda \subset\subset A$ and $V_\alpha \in \tau_\alpha$ for all $\alpha \in \Lambda$. We will now show that no such V can exist and hence $X_A = C_p$, i.e. X_A is connected.

Define $\phi : X_\Lambda \to X_A$ by $\phi(y) = x$ where

$$x_\alpha = \begin{cases} y_\alpha & \text{if } \alpha \in \Lambda \\ p_\alpha & \text{if } \alpha \notin \Lambda. \end{cases}$$

If $\alpha \in \Lambda$, $\pi_\alpha \circ \phi(y) = y_\alpha = \pi_\alpha(y)$ and if $\alpha \in A \backslash \Lambda$ then $\pi_\alpha \circ \phi(y) = p_\alpha$ so that in every case $\pi_\alpha \circ \phi : X_\Lambda \to X_\alpha$ is continuous and therefore ϕ is continuous. Since X_Λ is a product of a finite number of connected spaces and so is connected and thus so is the continuous image, $\phi(X_\Lambda) = X_\Lambda \times \{p_\alpha\}_{\alpha \in A \backslash \Lambda} \subset X_A$. Now $p \in \phi(X_\Lambda)$ and $\phi(X_\Lambda)$ is connected implies that $\phi(X_\Lambda) \subset C$. On the other hand one easily sees that

$$\emptyset \neq V \cap \phi(X_\Lambda) \subset V \cap C$$

contradicting the assumption that $V \subset C^c$. ∎

10.6 Exercises

10.6.1 General Topological Space Problems

Exercise 10.14. Let V be an open subset of $\mathbb{R}$. Show V may be written as a disjoint union of open intervals $J_n = (a_n, b_n)$, where $a_n, b_n \in \mathbb{R} \cup \{\pm\infty\}$ for $n = 1, 2, \cdots < N$ with $N = \infty$ possible.

Exercise 10.15. Let (X, τ) and (Y, τ') be a topological spaces, $f : X \to Y$ be a function, $\mathcal{U}$ be an open cover of X and $\{F_j\}_{j=1}^n$ be a finite cover of X by closed sets.

1. If $A \subset X$ is any set and $f : X \to Y$ is (τ, τ') – continuous then $f|_A : A \to Y$ is (τ_A, τ') – continuous.
2. Show $f : X \to Y$ is (τ, τ') – continuous iff $f|_U : U \to Y$ is (τ_U, τ') – continuous for all $U \in \mathcal{U}$.
3. Show $f : X \to Y$ is (τ, τ') – continuous iff $f|_{F_j} : F_j \to Y$ is (τ_{F_j}, τ') – continuous for all $j = 1, 2, \ldots, n$.

Exercise 10.16. Suppose that X is a set, $\{(Y_\alpha, \tau_\alpha) : \alpha \in A\}$ is a family of topological spaces and $f_\alpha : X \to Y_\alpha$ is a given function for all $\alpha \in A$. Assuming that $\mathcal{S}_\alpha \subset \tau_\alpha$ is a sub-base for the topology τ_α for each $\alpha \in A$, show $\mathcal{S} := \cup_{\alpha \in A} f_\alpha^{-1}(\mathcal{S}_\alpha)$ is a sub-base for the topology $\tau := \tau(f_\alpha : \alpha \in A)$.

10.6.2 Connectedness Problems

Exercise 10.17. Show any non-trivial interval in $\mathbb{Q}$ is disconnected.

Exercise 10.18. Suppose $a < b$ and $f : (a, b) \to \mathbb{R}$ is a non-decreasing function. Show if f satisfies the intermediate value property (see Theorem 10.50), then f is continuous.

Exercise 10.19. Suppose $-\infty < a < b \le \infty$ and $f : [a, b) \to \mathbb{R}$ is a strictly increasing continuous function. By Lemma 10.46, $f([a, b))$ is an interval and since f is strictly increasing it must of the form $[c, d)$ for some $c \in \mathbb{R}$ and $d \in \bar{\mathbb{R}}$ with $c < d$. Show the inverse function $f^{-1} : [c, d) \to [a, b)$ is continuous and is strictly increasing. In particular if $n \in \mathbb{N}$, apply this result to $f(x) = x^n$ for $x \in [0, \infty)$ to construct the positive n^{th} – root of a real number. Compare with Exercise 3.8

Exercise 10.20. Prove item 1. of Proposition 10.52. **Hint:** show X is not connected implies X is not path connected.

Exercise 10.21. Prove item 2. of Proposition 10.52. **Hint:** fix $x_0 \in X$ and let W denote the set of $x \in X$ such that there exists $\sigma \in C([0, 1], X)$ satisfying $\sigma(0) = x_0$ and $\sigma(1) = x$. Then show W is both open and closed.

Exercise 10.22. Prove item 3. of Proposition 10.52.

Exercise 10.23. Let

$$X := \left\{(x, y) \in \mathbb{R}^2 : y = \sin(x^{-1})\right\} \cup \{(0, 0)\}$$

equipped with the relative topology induced from the standard topology on $\mathbb{R}^2$. Show X is connected but not path connected.

10.6.3 Metric Spaces as Topological Spaces

Definition 10.54. *Two metrics d and ρ on a set X are said to be* ***equivalent*** *if there exists a constant $c \in (0, \infty)$ such that $c^{-1}\rho \le d \le c\rho$.*

Exercise 10.24. Suppose that d and ρ are two metrics on X.

1. Show $\tau_d = \tau_\rho$ if d and ρ are equivalent.
2. Show by example that it is possible for $\tau_d = \tau_\rho$ even thought d and ρ are inequivalent.

Exercise 10.25. Let (X_i, d_i) for $i = 1, \ldots, n$ be a finite collection of metric spaces and for $1 \le p \le \infty$ and $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $X := \prod_{i=1}^n X_i$, let

$$\rho_p(x, y) = \begin{cases} \left(\sum_{i=1}^n [d_i(x_i, y_i)]^p\right)^{1/p} & \text{if } p \ne \infty \\ \max_i d_i(x_i, y_i) & \text{if } p = \infty \end{cases}.$$

1. Show (X, ρ_p) is a metric space for $p \in [1, \infty]$. **Hint**: Minkowski's inequality.
2. Show for any $p, q \in [1, \infty]$, the metrics ρ_p and ρ_q are equivalent. **Hint:** This can be done with explicit estimates or you could use Theorem 11.12 below.

Notation 10.55 *Let X be a set and $\mathbf{p} := \{p_n\}_{n=0}^\infty$ be a family of semi-metrics on X, i.e. $p_n : X \times X \to [0, \infty)$ are functions satisfying the assumptions of metric except for the assertion that $p_n(x, y) = 0$ implies $x = y$. Further assume that $p_n(x, y) \le p_{n+1}(x, y)$ for all n and if $p_n(x, y) = 0$ for all $n \in \mathbb{N}$ then $x = y$. Given $n \in \mathbb{N}$ and $x \in X$ let*

$$B_n(x, \varepsilon) := \{y \in X : p_n(x, y) < \varepsilon\}.$$

We will write $\tau(\mathbf{p})$ form the smallest topology on X such that $p_n(x, \cdot) : X \to [0, \infty)$ is continuous for all $n \in \mathbb{N}$ and $x \in X$, i.e. $\tau(\mathbf{p}) := \tau(p_n(x\cdot) : n \in \mathbb{N}$ and $x \in X)$.

Exercise 10.26. Using Notation 10.55, show that collection of balls,

$$\mathcal{B} := \{B_n(x, \varepsilon) : n \in \mathbb{N},\ x \in X \text{ and } \varepsilon > 0\},$$

forms a base for the topology $\tau(\mathbf{p})$. **Hint:** Use Exercise 10.16 to show $\mathcal{B}$ is a sub-base for the topology $\tau(\mathbf{p})$ and then use Exercise 10.2 to show $\mathcal{B}$ is in fact a base for the topology $\tau(\mathbf{p})$.

Exercise 10.27 (A minor variant of Exercise 6.12). Let p_n be as in Notation 10.55 and

$$d(x, y) := \sum_{n=0}^\infty 2^{-n} \frac{p_n(x, y)}{1 + p_n(x, y)}.$$

Show d is a metric on X and $\tau_d = \tau(\mathbf{p})$. Conclude that a sequence $\{x_k\}_{k=1}^\infty \subset X$ converges to $x \in X$ iff

$$\lim_{k\to\infty} p_n(x_k, x) = 0 \text{ for all } n \in \mathbb{N}.$$

Exercise 10.28. Let $\{(X_n, d_n)\}_{n=1}^\infty$ be a sequence of metric spaces, $X := \prod_{n=1}^\infty X_n$, and for $x = (x(n))_{n=1}^\infty$ and $y = (y(n))_{n=1}^\infty$ in X let

$$d(x,y) = \sum_{n=1}^\infty 2^{-n} \frac{d_n(x(n), y(n))}{1 + d_n(x(n), y(n))}.$$

(See Exercise 6.12.) Moreover, let $\pi_n : X \to X_n$ be the projection maps, show

$$\tau_d = \otimes_{n=1}^\infty \tau_{d_n} := \tau(\{\pi_n : n \in \mathbb{N}\}).$$

That is show the d – metric topology is the same as the product topology on X. **Suggestions:** 1) show π_n is τ_d continuous for each n and 2) show for each $x \in X$ that $d(x, \cdot)$ is $\otimes_{n=1}^\infty \tau_{d_n}$ – continuous. For the second assertion notice that $d(x, \cdot) = \sum_{n=1}^\infty f_n$ where $f_n = 2^{-n} \left(\frac{d_n(x(n), \cdot)}{1 + d_n(x(n), \cdot)} \right) \circ \pi_n$.

11

Compactness

Definition 11.1. *The subset A of a topological space $(X\,\tau)$ is said to be* ***compact*** *if every open cover (Definition 10.18) of A has finite a sub-cover, i.e. if $\mathcal{U}$ is an open cover of A there exists $\mathcal{U}_0 \subset\subset \mathcal{U}$ such that $\mathcal{U}_0$ is a cover of A. (We will write $A \sqsubset\sqsubset X$ to denote that $A \subset X$ and A is compact.) A subset $A \subset X$ is* ***precompact*** *if $\bar{A}$ is compact.*

Proposition 11.2. *Suppose that $K \subset X$ is a compact set and $F \subset K$ is a closed subset. Then F is compact. If $\{K_i\}_{i=1}^n$ is a finite collections of compact subsets of X then $K = \cup_{i=1}^n K_i$ is also a compact subset of X.*

Proof. Let $\mathcal{U} \subset \tau$ be an open cover of F, then $\mathcal{U}\cup\{F^c\}$ is an open cover of K. The cover $\mathcal{U}\cup\{F^c\}$ of K has a finite subcover which we denote by $\mathcal{U}_0\cup\{F^c\}$ where $\mathcal{U}_0 \subset\subset \mathcal{U}$. Since $F \cap F^c = \emptyset$, it follows that $\mathcal{U}_0$ is the desired subcover of F. For the second assertion suppose $\mathcal{U} \subset \tau$ is an open cover of K. Then $\mathcal{U}$ covers each compact set K_i and therefore there exists a finite subset $\mathcal{U}_i \subset\subset \mathcal{U}$ for each i such that $K_i \subset \cup\mathcal{U}_i$. Then $\mathcal{U}_0 := \cup_{i=1}^n \mathcal{U}_i$ is a finite cover of K. ∎

Exercise 11.1 (Suggested by Michael Gurvich). Show by example that the intersection of two compact sets need not be compact. (This pathology disappears if one assumes the topology is Hausdorff, see Definition 12.2 below.)

Exercise 11.2. Suppose $f : X \to Y$ is continuous and $K \subset X$ is compact, then $f(K)$ is a compact subset of Y. Give an example of continuous map, $f : X \to Y$, and a compact subset K of Y such that $f^{-1}(K)$ is not compact.

Exercise 11.3 (Dini's Theorem). Let X be a compact topological space and $f_n : X \to [0, \infty)$ be a sequence of continuous functions such that $f_n(x) \downarrow 0$ as $n \to \infty$ for each $x \in X$. Show that in fact $f_n \downarrow 0$ uniformly in x, i.e. $\sup_{x\in X} f_n(x) \downarrow 0$ as $n \to \infty$. **Hint:** Given $\varepsilon > 0$, consider the open sets $V_n := \{x \in X : f_n(x) < \varepsilon\}$.

Definition 11.3. *A collection $\mathcal{F}$ of closed subsets of a topological space (X,τ) has the **finite intersection property** if $\cap\mathcal{F}_0 \neq \emptyset$ for all $\mathcal{F}_0 \subset\subset \mathcal{F}$.*

The notion of compactness may be expressed in terms of closed sets as follows.

Proposition 11.4. *A topological space X is compact iff every family of closed sets $\mathcal{F} \subset 2^X$ having the **finite intersection property** satisfies $\bigcap \mathcal{F} \neq \emptyset$.*

Proof. $(\Rightarrow)$ Suppose that X is compact and $\mathcal{F} \subset 2^X$ is a collection of closed sets such that $\bigcap \mathcal{F} = \emptyset$. Let

$$\mathcal{U} = \mathcal{F}^c := \{C^c : C \in \mathcal{F}\} \subset \tau,$$

then $\mathcal{U}$ is a cover of X and hence has a finite subcover, $\mathcal{U}_0$. Let $\mathcal{F}_0 = \mathcal{U}_0^c \subset\subset \mathcal{F}$, then $\cap\mathcal{F}_0 = \emptyset$ so that $\mathcal{F}$ does not have the finite intersection property. $(\Leftarrow)$ If X is not compact, there exists an open cover $\mathcal{U}$ of X with no finite subcover. Let

$$\mathcal{F} = \mathcal{U}^c := \{U^c : U \in \mathcal{U}\},$$

then $\mathcal{F}$ is a collection of closed sets with the finite intersection property while $\bigcap \mathcal{F} = \emptyset$. ■

Exercise 11.4. Let (X,τ) be a topological space. Show that $A \subset X$ is compact iff (A, τ_A) is a compact topological space.

11.1 Metric Space Compactness Criteria

Let (X,d) be a metric space and for $x \in X$ and $\varepsilon > 0$ let $B'_x(\varepsilon) = B_x(\varepsilon) \setminus \{x\}$ – the deleted ball centered at x of radius $\varepsilon > 0$. Recall from Definition 10.29 that a point $x \in X$ is an accumulation point of a subset $E \subset X$ if $\emptyset \neq E \cap V \setminus \{x\}$ for all open neighborhoods, V, of x. The proof of the following elementary lemma is left to the reader.

Lemma 11.5. *Let $E \subset X$ be a subset of a metric space (X,d). Then the following are equivalent:*

1. *$x \in X$ is an accumulation point of E.*
2. *$B'_x(\varepsilon) \cap E \neq \emptyset$ for all $\varepsilon > 0$.*
3. *$B_x(\varepsilon) \cap E$ is an infinite set for all $\varepsilon > 0$.*
4. *There exists $\{x_n\}_{n=1}^\infty \subset E \setminus \{x\}$ with $\lim_{n\to\infty} x_n = x$.*

Definition 11.6. *A metric space (X,d) is $\varepsilon-$ **bounded** $(\varepsilon > 0)$ if there exists a finite cover of X by balls of radius ε and it is **totally bounded** if it is ε – bounded for all $\varepsilon > 0$.*

Theorem 11.7. *Let (X,d) be a metric space. The following are equivalent.*

(a) X *is compact.*
(b) Every infinite subset of X *has an accumulation point.*
(c) Every sequence $\{x_n\}_{n=1}^{\infty} \subset X$ *has a convergent subsequence.*
(d) X *is totally bounded and complete.*

Proof. The proof will consist of showing that $a \Rightarrow b \Rightarrow c \Rightarrow d \Rightarrow a$.

($a \Rightarrow b$) We will show that **not** $b \Rightarrow$ **not** a. Suppose there exists an infinite subset $E \subset X$ which has no accumulation points. Then for all $x \in X$ there exists $\delta_x > 0$ such that $V_x := B_x(\delta_x)$ satisfies $(V_x \setminus \{x\}) \cap E = \emptyset$. Clearly $\mathcal{V} = \{V_x\}_{x \in X}$ is a cover of X, yet $\mathcal{V}$ has no finite sub cover. Indeed, for each $x \in X$, $V_x \cap E \subset \{x\}$ and hence if $\Lambda \subset\subset X$, $\cup_{x \in \Lambda} V_x$ can only contain a finite number of points from E (namely $\Lambda \cap E$). Thus for any $\Lambda \subset\subset X$, $E \nsubseteq \cup_{x \in \Lambda} V_x$ and in particular $X \neq \cup_{x \in \Lambda} V_x$. (See Figure 11.1.)

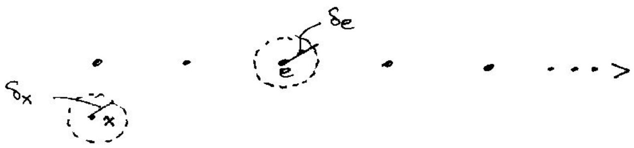


Fig. 11.1. The construction of an open cover with no finite sub-cover.

($b \Rightarrow c$) Let $\{x_n\}_{n=1}^{\infty} \subset X$ be a sequence and $E := \{x_n : n \in \mathbb{N}\}$. If $\#(E) < \infty$, then $\{x_n\}_{n=1}^{\infty}$ has a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ which is constant and hence convergent. On the other hand if $\#(E) = \infty$ then by assumption E has an accumulation point and hence by Lemma 11.5, $\{x_n\}_{n=1}^{\infty}$ has a convergent subsequence.

($c \Rightarrow d$) Suppose $\{x_n\}_{n=1}^{\infty} \subset X$ is a Cauchy sequence. By assumption there exists a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ which is convergent to some point $x \in X$. Since $\{x_n\}_{n=1}^{\infty}$ is Cauchy it follows that $x_n \to x$ as $n \to \infty$ showing X is complete. We now show that X is totally bounded. Let $\varepsilon > 0$ be given and choose an arbitrary point $x_1 \in X$. If possible choose $x_2 \in X$ such that $d(x_2, x_1) \geq \varepsilon$, then if possible choose $x_3 \in X$ such that $d_{\{x_1,x_2\}}(x_3) \geq \varepsilon$ and continue inductively choosing points $\{x_j\}_{j=1}^{n} \subset X$ such that $d_{\{x_1,\dots,x_{n-1}\}}(x_n) \geq \varepsilon$. (See Figure 11.2.) This process must terminate, for otherwise we would produce a sequence $\{x_n\}_{n=1}^{\infty} \subset X$ which can have no convergent subsequences. Indeed, the x_n have been chosen so that $d(x_n, x_m) \geq \varepsilon > 0$ for every $m \neq n$ and hence no subsequence of $\{x_n\}_{n=1}^{\infty}$ can be Cauchy.

($d \Rightarrow a$) For sake of contradiction, assume there exists an open cover $\mathcal{V} = \{V_\alpha\}_{\alpha \in A}$ of X with no finite subcover. Since X is totally bounded for each $n \in \mathbb{N}$ there exists $\Lambda_n \subset\subset X$ such that

$$X = \bigcup_{x \in \Lambda_n} B_x(1/n) \subset \bigcup_{x \in \Lambda_n} C_x(1/n).$$

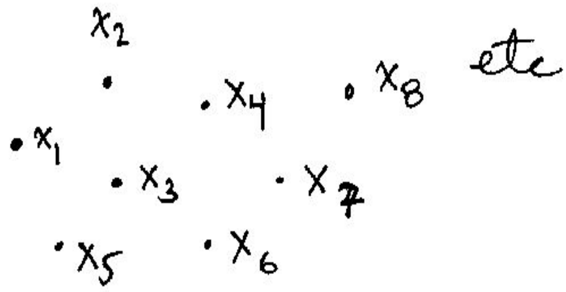


Fig. 11.2. Constructing a set with out an accumulation point.

Choose $x_1 \in \Lambda_1$ such that no finite subset of $\mathcal{V}$ covers $K_1 := C_{x_1}(1)$. Since $K_1 = \cup_{x\in\Lambda_2} K_1 \cap C_x(1/2)$, there exists $x_2 \in \Lambda_2$ such that $K_2 := K_1 \cap C_{x_2}(1/2)$ can not be covered by a finite subset of $\mathcal{V}$, see Figure 11.3. Continuing this way inductively, we construct sets $K_n = K_{n-1} \cap C_{x_n}(1/n)$ with $x_n \in \Lambda_n$ such that no K_n can be covered by a finite subset of $\mathcal{V}$. Now choose $y_n \in K_n$ for each n. Since $\{K_n\}_{n=1}^{\infty}$ is a decreasing sequence of closed sets such that $\mathrm{diam}(K_n) \le 2/n$, it follows that $\{y_n\}$ is a Cauchy and hence convergent with

$$y = \lim_{n\to\infty} y_n \in \cap_{m=1}^{\infty} K_m.$$

Since $\mathcal{V}$ is a cover of X, there exists $V \in \mathcal{V}$ such that $y \in V$. Since $K_n \downarrow \{y\}$ and $\mathrm{diam}(K_n) \to 0$, it now follows that $K_n \subset V$ for some n large. But this violates the assertion that K_n can not be covered by a finite subset of $\mathcal{V}$.

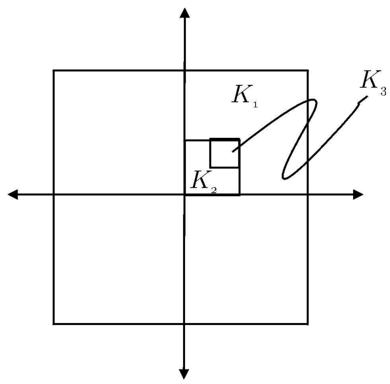


Fig. 11.3. Nested Sequence of cubes.

■

Corollary 11.8. *Any compact metric space (X, d) is second countable and hence also separable by Exercise 10.11. (See Example 12.25 below for an example of a compact topological space which is not separable.)*

Proof. To each integer n, there exists $\Lambda_n \subset\subset X$ such that $X = \cup_{x\in\Lambda_n} B(x, 1/n)$. The collection of open balls,

$$\mathcal{V} := \cup_{n\in\mathbb{N}} \cup_{x\in\Lambda_n} \{B(x, 1/n)\}$$

forms a countable basis for the metric topology on X. To check this, suppose that $x_0 \in X$ and $\varepsilon > 0$ are given and choose $n \in \mathbb{N}$ such that $1/n < \varepsilon/2$ and $x \in \Lambda_n$ such that $d(x_0, x) < 1/n$. Then $B(x, 1/n) \subset B(x_0, \varepsilon)$ because for $y \in B(x, 1/n)$,

$$d(y, x_0) \le d(y, x) + d(x, x_0) < 2/n < \varepsilon.$$

■

Corollary 11.9. *The compact subsets of $\mathbb{R}^n$ are the closed and bounded sets.*

Proof. This is a consequence of Theorem 8.2 and Theorem 11.7. Here is another proof. If K is closed and bounded then K is complete (being the closed subset of a complete space) and K is contained in $[-M, M]^n$ for some positive integer M. For $\delta > 0$, let

$$\Lambda_\delta = \delta\mathbb{Z}^n \cap [-M, M]^n := \{\delta x : x \in \mathbb{Z}^n \text{ and } \delta|x_i| \le M \text{ for } i = 1, 2, \ldots, n\}.$$

We will show, by choosing $\delta > 0$ sufficiently small, that

$$K \subset [-M, M]^n \subset \cup_{x\in\Lambda_\delta} B(x, \varepsilon) \tag{11.1}$$

which shows that K is totally bounded. Hence by Theorem 11.7, K is compact. Suppose that $y \in [-M, M]^n$, then there exists $x \in \Lambda_\delta$ such that $|y_i - x_i| \le \delta$ for $i = 1, 2, \ldots, n$. Hence

$$d^2(x, y) = \sum_{i=1}^n (y_i - x_i)^2 \le n\delta^2$$

which shows that $d(x, y) \le \sqrt{n}\delta$. Hence if choose $\delta < \varepsilon/\sqrt{n}$ we have shows that $d(x, y) < \varepsilon$, i.e. Eq. (11.1) holds. ■

Example 11.10. Let $X = \ell^p(\mathbb{N})$ with $p \in [1, \infty)$ and $\mu \in \ell^p(\mathbb{N})$ such that $\mu(k) \ge 0$ for all $k \in \mathbb{N}$. The set

$$K := \{x \in X : |x(k)| \le \mu(k) \text{ for all } k \in \mathbb{N}\}$$

is compact. To prove this, let $\{x_n\}_{n=1}^\infty \subset K$ be a sequence. By compactness of closed bounded sets in $\mathbb{C}$, for each $k \in \mathbb{N}$ there is a subsequence of $\{x_n(k)\}_{n=1}^\infty \subset \mathbb{C}$ which is convergent. By Cantor's diagonalization trick, we may choose a subsequence $\{y_n\}_{n=1}^\infty$ of $\{x_n\}_{n=1}^\infty$ such that

$y(k) := \lim_{n\to\infty} y_n(k)$ exists for all $k \in \mathbb{N}$.[1] Since $|y_n(k)| \le \mu(k)$ for all n it follows that $|y(k)| \le \mu(k)$, i.e. $y \in K$. Finally

$$\lim_{n\to\infty} \|y - y_n\|_p^p = \lim_{n\to\infty} \sum_{k=1}^{\infty} |y(k) - y_n(k)|^p = \sum_{k=1}^{\infty} \lim_{n\to\infty} |y(k) - y_n(k)|^p = 0$$

wherein we have used the Dominated convergence theorem. (Note $|y(k) - y_n(k)|^p \le 2^p \mu^p(k)$ and μ^p is summable.) Therefore $y_n \to y$ and we are done.

Alternatively, we can prove K is compact by showing that K is closed and totally bounded. It is simple to show K is closed, for if $\{x_n\}_{n=1}^{\infty} \subset K$ is a convergent sequence in X, $x := \lim_{n\to\infty} x_n$, then

$$|x(k)| \le \lim_{n\to\infty} |x_n(k)| \le \mu(k) \ \forall\ k \in \mathbb{N}.$$

This shows that $x \in K$ and hence K is closed. To see that K is totally bounded, let $\varepsilon > 0$ and choose N such that $\left(\sum_{k=N+1}^{\infty} |\mu(k)|^p\right)^{1/p} < \varepsilon$. Since $\prod_{k=1}^{N} C_{\mu(k)}(0) \subset \mathbb{C}^N$ is closed and bounded, it is compact. Therefore there exists a finite subset $\Lambda \subset \prod_{k=1}^{N} C_{\mu(k)}(0)$ such that

$$\prod_{k=1}^{N} C_{\mu(k)}(0) \subset \cup_{z\in\Lambda} B_z^N(\varepsilon)$$

where $B_z^N(\varepsilon)$ is the open ball centered at $z \in \mathbb{C}^N$ relative to the $\ell^p(\{1, 2, 3, \ldots, N\})$ – norm. For each $z \in \Lambda$, let $\tilde{z} \in X$ be defined by $\tilde{z}(k) = z(k)$ if $k \le N$ and $\tilde{z}(k) = 0$ for $k \ge N + 1$. I now claim that

$$K \subset \cup_{z\in\Lambda} B_{\tilde{z}}(2\varepsilon) \tag{11.2}$$

which, when verified, shows K is totally bounded. To verify Eq. (11.2), let $x \in K$ and write $x = u + v$ where $u(k) = x(k)$ for $k \le N$ and $u(k) = 0$ for $k < N$. Then by construction $u \in B_{\tilde{z}}(\varepsilon)$ for some $\tilde{z} \in \Lambda$ and

$$\|v\|_p \le \left(\sum_{k=N+1}^{\infty} |\mu(k)|^p\right)^{1/p} < \varepsilon.$$

So we have

$$\|x - \tilde{z}\|_p = \|u + v - \tilde{z}\|_p \le \|u - \tilde{z}\|_p + \|v\|_p < 2\varepsilon.$$

---

[1] The argument is as follows. Let $\{n_j^1\}_{j=1}^{\infty}$ be a subsequence of $\mathbb{N} = \{n\}_{n=1}^{\infty}$ such that $\lim_{j\to\infty} x_{n_j^1}(1)$ exists. Now choose a subsequence $\{n_j^2\}_{j=1}^{\infty}$ of $\{n_j^1\}_{j=1}^{\infty}$ such that $\lim_{j\to\infty} x_{n_j^2}(2)$ exists and similarly $\{n_j^3\}_{j=1}^{\infty}$ of $\{n_j^2\}_{j=1}^{\infty}$ such that $\lim_{j\to\infty} x_{n_j^3}(3)$ exists. Continue on this way inductively to get

$$\{n\}_{n=1}^{\infty} \supset \{n_j^1\}_{j=1}^{\infty} \supset \{n_j^2\}_{j=1}^{\infty} \supset \{n_j^3\}_{j=1}^{\infty} \supset \ldots$$

such that $\lim_{j\to\infty} x_{n_j^k}(k)$ exists for all $k \in \mathbb{N}$. Let $m_j := n_j^j$ so that eventually $\{m_j\}_{j=1}^{\infty}$ is a subsequence of $\{n_j^k\}_{j=1}^{\infty}$ for all k. Therefore, we may take $y_j := x_{m_j}$.

Exercise 11.5 (Extreme value theorem). Let (X, τ) be a compact topological space and $f : X \to \mathbb{R}$ be a continuous function. Show $-\infty < \inf f \le \sup f < \infty$ and there exists $a, b \in X$ such that $f(a) = \inf f$ and $f(b) = \sup f$[2]. **Hint:** use Exercise 11.2 and Corollary 11.9.

Exercise 11.6 (Uniform Continuity). Let (X, d) be a compact metric space, (Y, ρ) be a metric space and $f : X \to Y$ be a continuous function. Show that f is uniformly continuous, i.e. if $\varepsilon > 0$ there exists $\delta > 0$ such that $\rho(f(y), f(x)) < \varepsilon$ if $x, y \in X$ with $d(x, y) < \delta$. **Hint:** you could follow the argument in the proof of Theorem 8.2.

Definition 11.11. *Let L be a vector space. We say that two norms, $|\cdot|$ and $\|\cdot\|$, on L are equivalent if there exists constants $\alpha, \beta \in (0, \infty)$ such that*

$$\|f\| \le \alpha |f| \text{ and } |f| \le \beta \|f\| \text{ for all } f \in L.$$

Theorem 11.12. *Let L be a finite dimensional vector space. Then any two norms $|\cdot|$ and $\|\cdot\|$ on L are equivalent. (This is typically not true for norms on infinite dimensional spaces, see for example Exercise 7.5.)*

Proof. Let $\{f_i\}_{i=1}^n$ be a basis for L and define a new norm on L by

$$\left\| \sum_{i=1}^n a_i f_i \right\|_2 := \sqrt{\sum_{i=1}^n |a_i|^2} \text{ for } a_i \in \mathbb{F}.$$

By the triangle inequality for the norm $|\cdot|$, we find

$$\left| \sum_{i=1}^n a_i f_i \right| \le \sum_{i=1}^n |a_i| \, |f_i| \le \sqrt{\sum_{i=1}^n |f_i|^2} \sqrt{\sum_{i=1}^n |a_i|^2} \le M \left\| \sum_{i=1}^n a_i f_i \right\|_2$$

where $M = \sqrt{\sum_{i=1}^n |f_i|^2}$. Thus we have

$$|f| \le M \|f\|_2$$

for all $f \in L$ and this inequality shows that $|\cdot|$ is continuous relative to $\|\cdot\|_2$. Since the normed space $(L, \|\cdot\|_2)$ is homeomorphic and isomorphic to $\mathbb{F}^n$ with the standard euclidean norm, the closed bounded set, $S := \{f \in L : \|f\|_2 = 1\} \subset L$, is a compact subset of L relative to $\|\cdot\|_2$. Therefore by Exercise 11.5 there exists $f_0 \in S$ such that

$$m = \inf \{|f| : f \in S\} = |f_0| > 0.$$

[2] Here is a proof if X is a metric space. Let $\{x_n\}_{n=1}^\infty \subset X$ be a sequence such that $f(x_n) \uparrow \sup f$. By compactness of X we may assume, by passing to a subsequence if necessary that $x_n \to b \in X$ as $n \to \infty$. By continuity of f, $f(b) = \sup f$.

Hence given $0 \neq f \in L$, then $\frac{f}{\|f\|_2} \in S$ so that

$$m \leq \left| \frac{f}{\|f\|_2} \right| = |f| \frac{1}{\|f\|_2}$$

or equivalently

$$\|f\|_2 \leq \frac{1}{m} |f| .$$

This shows that $|\cdot|$ and $\|\cdot\|_2$ are equivalent norms. Similarly one shows that $\|\cdot\|$ and $\|\cdot\|_2$ are equivalent and hence so are $|\cdot|$ and $\|\cdot\|$. ■

Corollary 11.13. *If $(L, \|\cdot\|)$ is a finite dimensional normed space, then $A \subset L$ is compact iff A is closed and bounded relative to the given norm, $\|\cdot\|$.*

Corollary 11.14. *Every finite dimensional normed vector space $(L, \|\cdot\|)$ is complete. In particular any finite dimensional subspace of a normed vector space is automatically closed.*

Proof. If $\{f_n\}_{n=1}^{\infty} \subset L$ is a Cauchy sequence, then $\{f_n\}_{n=1}^{\infty}$ is bounded and hence has a convergent subsequence, $g_k = f_{n_k}$, by Corollary 11.13. It is now routine to show $\lim_{n\to\infty} f_n = f := \lim_{k\to\infty} g_k$. ■

Theorem 11.15. *Suppose that $(X, \|\cdot\|)$ is a normed vector in which the unit ball, $V := B_0(1)$, is precompact. Then $\dim X < \infty$.*

Proof. Since $\bar{V}$ is compact, we may choose $\Lambda \subset\subset X$ such that

$$\bar{V} \subset \cup_{x\in\Lambda} \left(x + \frac{1}{2} V \right) \tag{11.3}$$

where, for any $\delta > 0$,

$$\delta V := \{\delta x : x \in V\} = B_0(\delta) .$$

Let $Y := \text{span}(\Lambda)$, then Eq. (11.3) implies,

$$V \subset \bar{V} \subset Y + \frac{1}{2} V.$$

Multiplying this equation by $\frac{1}{2}$ then shows

$$\frac{1}{2} V \subset \frac{1}{2} Y + \frac{1}{4} V = Y + \frac{1}{4} V$$

and hence

$$V \subset Y + \frac{1}{2} V \subset Y + Y + \frac{1}{4} V = Y + \frac{1}{4} V.$$

Continuing this way inductively then shows that

$$V \subset Y + \frac{1}{2^n} V \text{ for all } n \in \mathbb{N}. \tag{11.4}$$

Indeed, if Eq. (11.4) holds, then

$$V \subset Y + \frac{1}{2} V \subset Y + \frac{1}{2}\left(Y + \frac{1}{2^n} V\right) = Y + \frac{1}{2^{n+1}} V.$$

Hence if $x \in V$, there exists $y_n \in Y$ and $z_n \in B_0\left(2^{-n}\right)$ such that $y_n + z_n \to x$. Since $\lim_{n\to\infty} z_n = 0$, it follows that $x = \lim_{n\to\infty} y_n \in \bar{Y}$. Since $\dim Y \leq \#(\Lambda) < \infty$, Corollary 11.14 implies $Y = \bar{Y}$ and so we have shown that $V \subset Y$. Since for any $x \in X$, $\frac{1}{2\|x\|} x \in V \subset Y$, we have $x \in Y$ for all $x \in X$, i.e. $X = Y$. ■

Exercise 11.7. Suppose $(Y, \|\cdot\|_Y)$ is a normed space and $(X, \|\cdot\|_X)$ is a finite dimensional normed space. Show **every** linear transformation $T : X \to Y$ is necessarily bounded.

11.2 Compact Operators

Definition 11.16. *Let $A : X \to Y$ be a bounded operator between two Banach spaces. Then A is* ***compact*** *if $A\left[B_X(0,1)\right]$ is precompact in Y or equivalently for any $\{x_n\}_{n=1}^\infty \subset X$ such that $\|x_n\| \leq 1$ for all n the sequence $y_n := Ax_n \in Y$ has a convergent subsequence.*

Example 11.17. Let $X = \ell^2 = Y$ and $\lambda_n \in \mathbb{C}$ such that $\lim_{n\to\infty} \lambda_n = 0$, then $A : X \to Y$ defined by $(Ax)(n) = \lambda_n x(n)$ is compact.

Proof. Suppose $\{x_j\}_{j=1}^\infty \subset \ell^2$ such that $\|x_j\|^2 = \sum |x_j(n)|^2 \leq 1$ for all j. By Cantor's Diagonalization argument, there exists $\{j_k\} \subset \{j\}$ such that, for each n, $\tilde{x}_k(n) = x_{j_k}(n)$ converges to some $\tilde{x}(n) \in \mathbb{C}$ as $k \to \infty$. By Fatou's Lemma 4.12,

$$\sum_{n=1}^\infty |\tilde{x}(n)|^2 = \sum_{n=1}^\infty \liminf_{k\to\infty} |\tilde{x}_k(n)|^2 \leq \liminf_{k\to\infty} \sum_{n=1}^\infty |\tilde{x}_k(n)|^2 \leq 1,$$

which shows $\tilde{x} \in \ell^2$.

Let $\lambda_M^* = \max_{n\geq M} |\lambda_n|$. Then

$$\begin{aligned}\|A\tilde{x}_k - A\tilde{x}\|^2 &= \sum_{n=1}^{\infty} |\lambda_n|^2 \, |\tilde{x}_k(n) - \tilde{x}(n)|^2 \\ &\le \sum_{n=1}^{M} |\lambda_n|^2 |\tilde{x}_k(n) - \tilde{x}(n)|^2 + |\lambda_M^*|^2 \sum_{M+1}^{\infty} |\tilde{x}_k(n) - \tilde{x}(n)|^2 \\ &\le \sum_{n=1}^{M} |\lambda_n|^2 |\tilde{x}_k(n) - \tilde{x}(n)|^2 + |\lambda_M^*|^2 \, \|\tilde{x}_k - \tilde{x}\|^2 \\ &\le \sum_{n=1}^{M} |\lambda_n|^2 |\tilde{x}_k(n) - \tilde{x}(n)|^2 + 4|\lambda_M^*|^2 .\end{aligned}$$

Passing to the limit in this inequality then implies

$$\limsup_{k\to\infty} \|A\tilde{x}_k - A\tilde{x}\|^2 \le 4|\lambda_M^*|^2 \to 0 \text{ as } M \to \infty$$

and this completes the proof the A is a compact operator. ∎

Lemma 11.18. *If $X \xrightarrow{A} Y \xrightarrow{B} Z$ are bounded operators such the either A or B is compact then the composition $BA : X \to Z$ is also compact.*

Proof. Let $B_X(0,1)$ be the open unit ball in X. If A is compact and B is bounded, then $BA(B_X(0,1)) \subset B(\overline{AB_X(0,1)})$ which is compact since the image of compact sets under continuous maps are compact. Hence we conclude that $\overline{BA(B_X(0,1))}$ is compact, being the closed subset of the compact set $B(\overline{AB_X(0,1)})$. If A is continuous and B is compact, then $A(B_X(0,1))$ is a bounded set and so by the compactness of B, $BA(B_X(0,1))$ is a precompact subset of Z, i.e. BA is compact. ∎

11.3 Local and σ − Compactness

Notation 11.19 *If X is a topological spaces and Y is a normed space, let*

$$BC(X,Y) := \{f \in C(X,Y) : \sup_{x\in X} \|f(x)\|_Y < \infty\}$$

and

$$C_c(X,Y) := \{f \in C(X,Y) : \mathrm{supp}(f) \text{ is compact}\}.$$

If $Y = \mathbb{R}$ or $\mathbb{C}$ we will simply write $C(X)$, $BC(X)$ and $C_c(X)$ for $C(X,Y)$, $BC(X,Y)$ and $C_c(X,Y)$ respectively.

Remark 11.20. Let X be a topological space and Y be a Banach space. By combining Exercise 11.2 and Theorem 11.7 it follows that $C_c(X,Y) \subset BC(X,Y)$.

Definition 11.21 (Local and σ – compactness). *Let (X, τ) be a topological space.*

1. *(X, τ) is **locally compact** if for all $x \in X$ there exists an open neighborhood $V \subset X$ of x such that $\bar{V}$ is compact. (Alternatively, in light of Definition 10.29 (also see Definition 6.5), this is equivalent to requiring that to each $x \in X$ there exists a compact neighborhood N_x of x.)*
2. *(X, τ) is σ – **compact** if there exists compact sets $K_n \subset X$ such that $X = \cup_{n=1}^{\infty} K_n$. (Notice that we may assume, by replacing K_n by $K_1 \cup K_2 \cup \cdots \cup K_n$ if necessary, that $K_n \uparrow X$.)*

Example 11.22. Any open subset of $U \subset \mathbb{R}^n$ is a locally compact and σ – compact metric space. The proof of local compactness is easy and is left to the reader. To see that U is σ – compact, for $k \in \mathbb{N}$, let

$$K_k := \{x \in U : |x| \leq k \text{ and } d_{U^c}(x) \geq 1/k\}.$$

Then K_k is a closed and bounded subset of $\mathbb{R}^n$ and hence compact. Moreover $K_k^o \uparrow U$ as $k \to \infty$ since[3]

$$K_k^o \supset \{x \in U : |x| < k \text{ and } d_{U^c}(x) > 1/k\} \uparrow U \text{ as } k \to \infty.$$

Exercise 11.8. If (X, τ) is locally compact and second countable, then there is a countable basis $\mathcal{B}_0$ for the topology consisting of precompact open sets. Use this to show (X, τ) is σ - compact.

Exercise 11.9. Every separable locally compact metric space is σ – compact.

Exercise 11.10. Every σ – compact metric space is second countable (or equivalently separable), see Corollary 11.8.

Exercise 11.11. Suppose that (X, d) is a metric space and $U \subset X$ is an open subset.

1. If X is locally compact then (U, d) is locally compact.
2. If X is σ – compact then (U, d) is σ – compact. **Hint:** Mimic Example 11.22, replacing $\{x \in \mathbb{R}^n : |x| \leq k\}$ by compact sets $X_k \sqsubset\sqsubset X$ such that $X_k \uparrow X$.

Lemma 11.23. *Let (X, τ) be locally and σ – compact. Then there exists compact sets $K_n \uparrow X$ such that $K_n \subset K_{n+1}^o \subset K_{n+1}$ for all n.*

Proof. Suppose that $C \subset X$ is a compact set. For each $x \in C$ let $V_x \subset_o X$ be an open neighborhood of x such that $\bar{V}_x$ is compact. Then $C \subset \cup_{x \in C} V_x$ so there exists $\Lambda \subset\subset C$ such that

$$C \subset \cup_{x \in \Lambda} V_x \subset \cup_{x \in \Lambda} \bar{V}_x =: K.$$

[3] In fact this is an equality, but we will not need this here.

Then K is a compact set, being a finite union of compact subsets of X, and $C \subset \cup_{x\in\Lambda} V_x \subset K^o$. Now let $C_n \subset X$ be compact sets such that $C_n \uparrow X$ as $n \to \infty$. Let $K_1 = C_1$ and then choose a compact set K_2 such that $C_2 \subset K_2^o$. Similarly, choose a compact set K_3 such that $K_2 \cup C_3 \subset K_3^o$ and continue inductively to find compact sets K_n such that $K_n \cup C_{n+1} \subset K_{n+1}^o$ for all n. Then $\{K_n\}_{n=1}^\infty$ is the desired sequence. ∎

Remark 11.24. Lemma 11.23 may also be stated as saying there exists precompact open sets $\{G_n\}_{n=1}^\infty$ such that $G_n \subset \bar{G}_n \subset G_{n+1}$ for all n and $G_n \uparrow X$ as $n \to \infty$. Indeed if $\{G_n\}_{n=1}^\infty$ are as above, let $K_n := \bar{G}_n$ and if $\{K_n\}_{n=1}^\infty$ are as in Lemma 11.23, let $G_n := K_n^o$.

Proposition 11.25. *Suppose X is a locally compact metric space and $U \subset_o X$ and $K \sqsubset\sqsubset U$. Then there exists $V \subset_o X$ such that $K \subset V \subset \overline{V} \subset U \subset X$ and $\bar{V}$ is compact.*

Proof. (This is done more generally in Proposition 12.7 below.) By local compactness or X, for each $x \in K$ there exists $\varepsilon_x > 0$ such that $\overline{B_x(\varepsilon_x)}$ is compact and by shrinking ε_x if necessary we may assume,

$$\overline{B_x(\varepsilon_x)} \subset C_x(\varepsilon_x) \subset B_x(2\varepsilon_x) \subset U$$

for each $x \in K$. By compactness of K, there exists $\Lambda \subset\subset K$ such that $K \subset \cup_{x\in\Lambda} B_x(\varepsilon_x) =: V$. Notice that $\bar{V} \subset \cup_{x\in\Lambda} \overline{B_x(\varepsilon_x)} \subset U$ and $\bar{V}$ is a closed subset of the compact set $\cup_{x\in\Lambda} \overline{B_x(\varepsilon_x)}$ and hence compact as well. ∎

Definition 11.26. *Let U be an open subset of a topological space (X, τ). We will write $f \prec U$ to mean a function $f \in C_c(X, [0,1])$ such that* $\mathrm{supp}(f) := \overline{\{f \neq 0\}} \subset U$.

Lemma 11.27 (Urysohn's Lemma for Metric Spaces). *Let X be a locally compact metric space and $K \sqsubset\sqsubset U \subset_o X$. Then there exists $f \prec U$ such that $f = 1$ on K. In particular, if K is compact and C is closed in X such that $K \cap C = \emptyset$, there exists $f \in C_c(X, [0,1])$ such that $f = 1$ on K and $f = 0$ on C.*

Proof. Let V be as in Proposition 11.25 and then use Lemma 6.15 to find a function $f \in C(X, [0,1])$ such that $f = 1$ on K and $f = 0$ on V^c. Then $\mathrm{supp}(f) \subset \bar{V} \subset U$ and hence $f \prec U$. ∎

11.4 Function Space Compactness Criteria

In this section, let (X, τ) be a topological space.

Definition 11.28. *Let $\mathcal{F} \subset C(X)$.*

1. $\mathcal{F}$ is **equicontinuous at** $x \in X$ iff for all $\varepsilon > 0$ there exists $U \in \tau_x$ such that $|f(y) - f(x)| < \varepsilon$ for all $y \in U$ and $f \in \mathcal{F}$.
2. $\mathcal{F}$ is **equicontinuous** if $\mathcal{F}$ is equicontinuous at all points $x \in X$.
3. $\mathcal{F}$ is **pointwise bounded** if $\sup\{|f(x)| : |f \in \mathcal{F}\} < \infty$ for all $x \in X$.

Theorem 11.29 (Ascoli-Arzela Theorem). *Let (X, τ) be a compact topological space and $\mathcal{F} \subset C(X)$. Then $\mathcal{F}$ is precompact in $C(X)$ iff $\mathcal{F}$ is equicontinuous and point-wise bounded.*

Proof. ($\Leftarrow$) Since $C(X) \subset \ell^\infty(X)$ is a complete metric space, we must show $\mathcal{F}$ is totally bounded. Let $\varepsilon > 0$ be given. By equicontinuity, for all $x \in X$, there exists $V_x \in \tau_x$ such that $|f(y) - f(x)| < \varepsilon/2$ if $y \in V_x$ and $f \in \mathcal{F}$. Since X is compact we may choose $\Lambda \subset\subset X$ such that $X = \cup_{x\in\Lambda} V_x$. We have now decomposed X into "blocks" $\{V_x\}_{x\in\Lambda}$ such that each $f \in \mathcal{F}$ is constant to within ε on V_x. Since $\sup\{|f(x)| : x \in \Lambda \text{ and } f \in \mathcal{F}\} < \infty$, it is now evident that

$$\begin{aligned} M &= \sup\{|f(x)| : x \in X \text{ and } f \in \mathcal{F}\} \\ &\le \sup\{|f(x)| : x \in \Lambda \text{ and } f \in \mathcal{F}\} + \varepsilon < \infty. \end{aligned}$$

Let $\mathbb{D} := \{k\varepsilon/2 : k \in \mathbb{Z}\} \cap [-M, M]$. If $f \in \mathcal{F}$ and $\phi \in \mathbb{D}^\Lambda$ (i.e. $\phi : \Lambda \to \mathbb{D}$ is a function) is chosen so that $|\phi(x) - f(x)| \le \varepsilon/2$ for all $x \in \Lambda$, then

$$|f(y) - \phi(x)| \le |f(y) - f(x)| + |f(x) - \phi(x)| < \varepsilon \;\forall\; x \in \Lambda \text{ and } y \in V_x.$$

From this it follows that $\mathcal{F} = \bigcup\left\{\mathcal{F}_\phi : \phi \in \mathbb{D}^\Lambda\right\}$ where, for $\phi \in \mathbb{D}^\Lambda$,

$$\mathcal{F}_\phi := \{f \in \mathcal{F} : |f(y) - \phi(x)| < \varepsilon \text{ for } y \in V_x \text{ and } x \in \Lambda\}.$$

Let $\Gamma := \left\{\phi \in \mathbb{D}^\Lambda : \mathcal{F}_\phi \neq \emptyset\right\}$ and for each $\phi \in \Gamma$ choose $f_\phi \in \mathcal{F}_\phi \cap \mathcal{F}$. For $f \in \mathcal{F}_\phi$, $x \in \Lambda$ and $y \in V_x$ we have

$$|f(y) - f_\phi(y)| \le |f(y) - \phi(x))| + |\phi(x) - f_\phi(y)| < 2\varepsilon.$$

So $\|f - f_\phi\|_\infty < 2\varepsilon$ for all $f \in \mathcal{F}_\phi$ showing that $\mathcal{F}_\phi \subset B_{f_\phi}(2\varepsilon)$. Therefore,

$$\mathcal{F} = \cup_{\phi\in\Gamma}\mathcal{F}_\phi \subset \cup_{\phi\in\Gamma} B_{f_\phi}(2\varepsilon)$$

and because $\varepsilon > 0$ was arbitrary we have shown that $\mathcal{F}$ is totally bounded.

($\Rightarrow$) (*The rest of this proof may safely be skipped.) Since $\|\cdot\|_\infty : C(X) \to [0, \infty)$ is a continuous function on $C(X)$ it is bounded on any compact subset $\mathcal{F} \subset C(X)$. This shows that $\sup\{\|f\|_\infty : f \in \mathcal{F}\} < \infty$ which clearly implies that $\mathcal{F}$ is pointwise bounded.[4] Suppose $\mathcal{F}$ were **not** equicontinuous at some

[4] One could also prove that $\mathcal{F}$ is pointwise bounded by considering the continuous evaluation maps $e_x : C(X) \to \mathbb{R}$ given by $e_x(f) = f(x)$ for all $x \in X$.

point $x \in X$ that is to say there exists $\varepsilon > 0$ such that for all $V \in \tau_x$, $\sup_{y \in V} \sup_{f \in \mathcal{F}} |f(y) - f(x)| > \varepsilon$.[5] Equivalently said, to each $V \in \tau_x$ we may choose

$$f_V \in \mathcal{F} \text{ and } x_V \in V \ni |f_V(x) - f_V(x_V)| \geq \varepsilon. \tag{11.5}$$

Set $\mathcal{C}_V = \overline{\{f_W : W \in \tau_x \text{ and } W \subset V\}}^{\|\cdot\|_\infty} \subset \mathcal{F}$ and notice for any $\mathcal{V} \subset\subset \tau_x$ that

$$\cap_{V \in \mathcal{V}} \mathcal{C}_V \supseteq \mathcal{C}_{\cap \mathcal{V}} \neq \emptyset,$$

so that $\{\mathcal{C}_V\}_V \in \tau_x \subset \mathcal{F}$ has the finite intersection property.[6] Since $\mathcal{F}$ is compact, it follows that there exists some

$$f \in \bigcap_{V \in \tau_x} \mathcal{C}_V \neq \emptyset.$$

Since f is continuous, there exists $V \in \tau_x$ such that $|f(x) - f(y)| < \varepsilon/3$ for all $y \in V$. Because $f \in \mathcal{C}_V$, there exists $W \subset V$ such that $\|f - f_W\| < \varepsilon/3$. We now arrive at a contradiction;

$$\begin{aligned}
\varepsilon &\leq |f_W(x) - f_W(x_W)| \\
&\leq |f_W(x) - f(x)| + |f(x) - f(x_W)| + |f(x_W) - f_W(x_W)| \\
&< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.
\end{aligned}$$

■

Exercise 11.12. Give an alternative proof of the implication, $(\Leftarrow)$, in Theorem 11.29 by showing every subsequence $\{f_n : n \in \mathbb{N}\} \subset \mathcal{F}$ has a convergence sub-sequence.

---

[5] If X is first countable we could finish the proof with the following argument. Let $\{V_n\}_{n=1}^\infty$ be a neighborhood base at x such that $V_1 \supset V_2 \supset V_3 \supset \dots$. By the assumption that $\mathcal{F}$ is not equicontinuous at x, there exist $f_n \in \mathcal{F}$ and $x_n \in V_n$ such that $|f_n(x) - f_n(x_n)| \geq \epsilon \ \forall\ n$. Since $\mathcal{F}$ is a compact metric space by passing to a subsequence if necessary we may assume that f_n converges uniformly to some $f \in \mathcal{F}$. Because $x_n \to x$ as $n \to \infty$ we learn that

$$\begin{aligned}
\epsilon \leq |f_n(x) - f_n(x_n)| &\leq |f_n(x) - f(x)| + |f(x) - f(x_n)| + |f(x_n) - f_n(x_n)| \\
&\leq 2\|f_n - f\| + |f(x) - f(x_n)| \to 0 \text{ as } n \to \infty
\end{aligned}$$

which is a contradiction.

[6] If we are willing to use Net's described in Appendix C below we could finish the proof as follows. Since $\mathcal{F}$ is compact, the net $\{f_V\}_{V \in \tau_x} \subset \mathcal{F}$ has a cluster point $f \in \mathcal{F} \subset C(X)$. Choose a subnet $\{g_\alpha\}_{\alpha \in A}$ of $\{f_V\}_{V \in \tau_X}$ such that $g_\alpha \to f$ uniformly. Then, since $x_V \to x$ implies $x_{V_\alpha} \to x$, we may conclude from Eq. (11.5) that

$$\epsilon \leq |g_\alpha(x) - g_\alpha(x_{V_\alpha})| \to |g(x) - g(x)| = 0$$

which is a contradiction.

Exercise 11.13. Suppose $k \in C\left([0,1]^2, \mathbb{R}\right)$ and for $f \in C([0,1], \mathbb{R})$, let

$$Kf(x) := \int_0^1 k(x,y) f(y)\, dy \text{ for all } x \in [0,1].$$

Show K is a compact operator on $(C([0,1], \mathbb{R}), \|\cdot\|_\infty)$.

The following result is a corollary of Lemma 11.23 and Theorem 11.29.

Corollary 11.30 (Locally Compact Ascoli-Arzela Theorem). *Let (X, τ) be a locally compact and σ – compact topological space and $\{f_m\} \subset C(X)$ be a pointwise bounded sequence of functions such that $\{f_m|_K\}$ is equicontinuous for any compact subset $K \subset X$. Then there exists a subsequence $\{m_n\} \subset \{m\}$ such that $\{g_n := f_{m_n}\}_{n=1}^\infty \subset C(X)$ is a sequence which is uniformly convergent on compact subsets of X.*

Proof. Let $\{K_n\}_{n=1}^\infty$ be the compact subsets of X constructed in Lemma 11.23. We may now apply Theorem 11.29 repeatedly to find a nested family of subsequences

$$\{f_m\} \supset \{g_m^1\} \supset \{g_m^2\} \supset \{g_m^3\} \supset \dots$$

such that the sequence $\{g_m^n\}_{m=1}^\infty \subset C(X)$ is uniformly convergent on K_n. Using Cantor's trick, define the subsequence $\{h_n\}$ of $\{f_m\}$ by $h_n := g_n^n$. Then $\{h_n\}$ is uniformly convergent on K_l for each $l \in \mathbb{N}$. Now if $K \subset X$ is an arbitrary compact set, there exists $l < \infty$ such that $K \subset K_l^o \subset K_l$ and therefore $\{h_n\}$ is uniformly convergent on K as well. ∎

Proposition 11.31. *Let $\Omega \subset_o \mathbb{R}^d$ such that $\bar{\Omega}$ is compact and $0 \le \alpha < \beta \le 1$. Then the inclusion map $i : C^\beta(\overline{\Omega}) \hookrightarrow C^\alpha(\overline{\Omega})$ is a compact operator. See Chapter 9 and Lemma 9.9 for the notation being used here.*

Let $\{u_n\}_{n=1}^\infty \subset C^\beta(\overline{\Omega})$ such that $\|u_n\|_{C^\beta} \le 1$, i.e. $\|u_n\|_\infty \le 1$ and

$$|u_n(x) - u_n(y)| \le |x-y|^\beta \text{ for all } x, y \in \overline{\Omega}.$$

By the Arzela-Ascoli Theorem 11.29, there exists a subsequence of $\{\tilde{u}_n\}_{n=1}^\infty$ of $\{u_n\}_{n=1}^\infty$ and $u \in C^o(\bar{\Omega})$ such that $\tilde{u}_n \to u$ in C^0. Since

$$|u(x) - u(y)| = \lim_{n\to\infty} |\tilde{u}_n(x) - \tilde{u}_n(y)| \le |x-y|^\beta,$$

$u \in C^\beta$ as well. Define $g_n := u - \tilde{u}_n \in C^\beta$, then

$$[g_n]_\beta + \|g_n\|_{C^0} = \|g_n\|_{C^\beta} \le 2$$

and $g_n \to 0$ in C^0. To finish the proof we must show that $g_n \to 0$ in C^α. Given $\delta > 0$,

$$[g_n]_\alpha = \sup_{x \ne y} \frac{|g_n(x) - g_n(y)|}{|x-y|^\alpha} \le A_n + B_n$$

where

$$\begin{aligned}A_n &= \sup\left\{\frac{|g_n(x)-g_n(y)|}{|x-y|^\alpha} : x \neq y \text{ and } |x-y| \le \delta\right\}\\ &= \sup\left\{\frac{|g_n(x)-g_n(y)|}{|x-y|^\beta} \cdot |x-y|^{\beta-\alpha} : x \neq y \text{ and } |x-y| \le \delta\right\}\\ &\le \delta^{\beta-\alpha} \cdot [g_n]_\beta \le 2\delta^{\beta-\alpha}\end{aligned}$$

and

$$B_n = \sup\left\{\frac{|g_n(x)-g_n(y)|}{|x-y|^\alpha} : |x-y| > \delta\right\} \le 2\delta^{-\alpha} \|g_n\|_{C^0} \to 0 \text{ as } n \to \infty.$$

Therefore,

$$\limsup_{n\to\infty} [g_n]_\alpha \le \limsup_{n\to\infty} A_n + \limsup_{n\to\infty} B_n \le 2\delta^{\beta-\alpha} + 0 \to 0 \text{ as } \delta \downarrow 0.$$

This proposition generalizes to the following theorem which the reader is asked to prove in Exercise 11.20 below.

Theorem 11.32. *Let Ω be a precompact open subset of $\mathbb{R}^d$, $\alpha, \beta \in [0,1]$ and $k, j \in \mathbb{N}_0$. If $j+\beta > k+\alpha$, then $C^{j,\beta}\left(\bar{\Omega}\right)$ is compactly contained in $C^{k,\alpha}\left(\bar{\Omega}\right)$.*

11.5 Tychonoff's Theorem

The goal of this section is to show that arbitrary products of compact spaces is still compact. Before going to the general case of an arbitrary number of factors let us start with only two factors.

Proposition 11.33. *Suppose that X and Y are non-empty compact topological spaces, then $X \times Y$ is compact in the product topology.*

Proof. Let $\mathcal{U}$ be an open cover of $X \times Y$. Then for each $(x,y) \in X \times Y$ there exist $U \in \mathcal{U}$ such that $(x,y) \in U$. By definition of the product topology, there also exist $V_x \in \tau_x^X$ and $W_y \in \tau_y^Y$ such that $V_x \times W_y \subset U$. Therefore $\mathcal{V} := \{V_x \times W_y : (x,y) \in X \times Y\}$ is also an open cover of $X \times Y$. We will now show that $\mathcal{V}$ has a finite sub-cover, say $\mathcal{V}_0 \subset\subset \mathcal{V}$. Assuming this is proved for the moment, this implies that $\mathcal{U}$ also has a finite subcover because each $V \in \mathcal{V}_0$ is contained in some $U_V \in \mathcal{U}$. So to complete the proof it suffices to show every cover $\mathcal{V}$ of the form $\mathcal{V} = \{V_\alpha \times W_\alpha : \alpha \in A\}$ where $V_\alpha \subset_o X$ and $W_\alpha \subset_o Y$ has a finite subcover. Given $x \in X$, let $f_x : Y \to X \times Y$ be the map $f_x(y) = (x,y)$ and notice that f_x is continuous since $\pi_X \circ f_x(y) = x$ and $\pi_Y \circ f_x(y) = y$ are continuous maps. From this we conclude that $\{x\} \times Y = f_x(Y)$ is compact. Similarly, it follows that $X \times \{y\}$ is compact for all $y \in Y$. Since $\mathcal{V}$ is a cover of $\{x\} \times Y$, there exist $\Gamma_x \subset\subset A$ such that $\{x\} \times Y \subset \bigcup_{\alpha \in \Gamma_x} (V_\alpha \times W_\alpha)$ without loss of generality we may assume that Γ_x is chosen so that $x \in V_\alpha$ for all $\alpha \in \Gamma_x$. Let $U_x := \bigcap_{\alpha \in \Gamma_x} V_\alpha \subset_o X$, and notice that

$$\bigcup_{\alpha \in \Gamma_x} (V_\alpha \times W_\alpha) \supset \bigcup_{\alpha \in \Gamma_x} (U_x \times W_\alpha) = U_x \times Y, \tag{11.6}$$

see Figure 11.4 below. Since $\{U_x\}_{x \in X}$ is now an open cover of X and X is

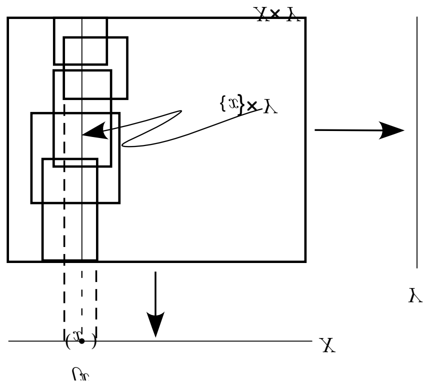


Fig. 11.4. Constructing the open set U_x.

compact, there exists $\Lambda \subset\subset X$ such that $X = \cup_{x \in \Lambda} U_x$. The finite subcollection, $\mathcal{V}_0 := \{V_\alpha \times W_\alpha : \alpha \in \cup_{x \in \Lambda} \Gamma_x\}$, of $\mathcal{V}$ is the desired finite subcover. Indeed using Eq. (11.6),

$$\cup \mathcal{V}_0 = \cup_{x \in \Lambda} \cup_{\alpha \in \Gamma_x} (V_\alpha \times W_\alpha) \supset \cup_{x \in \Lambda} (U_x \times Y) = X \times Y.$$

■

The results of Exercises 11.21 and 10.28 prove Tychonoff's Theorem for a countable product of compact metric spaces. We now state the general version of the theorem.

Theorem 11.34 (Tychonoff's Theorem). *Let $\{X_\alpha\}_{\alpha \in A}$ be a collection of non-empty compact spaces. Then $X := X_A = \prod_{\alpha \in A} X_\alpha$ is compact in the product space topology. (Compare with Exercise 11.21 which covers the special case of a countable product of compact metric spaces.)*

Proof. (The proof is taken from Loomis [1] which followed Bourbaki. Remark 11.35 below should help the reader understand the strategy of the proof to follow.) The proof requires a form of "induction" known as Zorn's lemma which is equivalent to the axiom of choice, see Theorem B.7 of Appendix B below.

For $\alpha \in A$ let π_α denote the projection map from X to X_α. Suppose that $\mathcal{F}$ is a family of closed subsets of X which has the finite intersection property, see Definition 11.3. By Proposition 11.4 the proof will be complete if we can show $\cap\mathcal{F} \neq \emptyset$.

The first step is to apply Zorn's lemma to construct a maximal collection, $\mathcal{F}_0$, of (not necessarily closed) subsets of X with the finite intersection property such that $\mathcal{F} \subset \mathcal{F}_0$. To do this, let $\Gamma := \left\{\mathcal{G} \subset 2^X : \mathcal{F} \subset \mathcal{G}\right\}$ equipped with the partial order, $\mathcal{G}_1 < \mathcal{G}_2$ if $\mathcal{G}_1 \subset \mathcal{G}_2$. If Φ is a linearly ordered subset of Γ, then $\mathcal{G} := \cup\Phi$ is an upper bound for Γ which still has the finite intersection property as the reader should check. So by Zorn's lemma, Γ has a maximal element $\mathcal{F}_0$. The maximal $\mathcal{F}_0$ has the following properties.

1. $\mathcal{F}_0$ is closed under finite intersections. Indeed, if we let $(\mathcal{F}_0)_f$ denote the collection of all finite intersections of elements from $\mathcal{F}_0$, then $(\mathcal{F}_0)_f$ has the finite intersection property and contains $\mathcal{F}_0$. Since $\mathcal{F}_0$ is maximal, this implies $(\mathcal{F}_0)_f = \mathcal{F}_0$.
2. If $B \subset X$ and $B \cap F \neq \emptyset$ for all $F \in \mathcal{F}_0$ then $B \in \mathcal{F}_0$. For if not $\mathcal{F}_0 \cup \{B\}$ would still satisfy the finite intersection property and would properly contain $\mathcal{F}_0$ and this would violate the maximallity of $\mathcal{F}_0$.
3. For each $\alpha \in A$,
$$\pi_a(\mathcal{F}_0) := \{\pi_\alpha(F) \subset X_\alpha : F \in \mathcal{F}_0\}$$
has the finite intersection property. Indeed, if $\{F_i\}_{i=1}^n \subset \mathcal{F}_0$, then $\cap_{i=1}^n \pi_\alpha(F_i) \supset \pi_\alpha\left(\cap_{i=1}^n F_i\right) \neq \emptyset$.

Since X_α is compact, property 3. above along with Proposition 11.4 implies $\cap_{F\in\mathcal{F}_0}\overline{\pi_\alpha(F)} \neq \emptyset$. Since this true for each $\alpha \in A$, using the axiom of choice, there exists $p \in X$ such that $p_\alpha = \pi_\alpha(p) \in \cap_{F\in\mathcal{F}_0}\overline{\pi_\alpha(F)}$ for all $\alpha \in A$. The proof will be completed by showing $\cap\mathcal{F} \neq \emptyset$ by showing $p \in \cap\mathcal{F}$.

Since $C := \cap\left\{\bar{F} : F \in \mathcal{F}_0\right\} \subset \cap\mathcal{F}$, it suffices to show $p \in C$. Let U be an open neighborhood of p in X. By the definition of the product topology (or item 2. of Proposition 10.25), there exists $\Lambda \subset\subset A$ and open sets $U_\alpha \subset X_\alpha$ for all $\alpha \in \Lambda$ such that $p \in \cap_{\alpha\in\Lambda}\pi_\alpha^{-1}(U_\alpha) \subset U$. Since $p_\alpha \in \cap_{F\in\mathcal{F}_0}\overline{\pi_\alpha(F)}$ and $p_\alpha \in U_\alpha$ for all $\alpha \in \Lambda$, it follows that $U_\alpha \cap \pi_\alpha(F) \neq \emptyset$ for all $F \in \mathcal{F}_0$ and all $\alpha \in \Lambda$. This then implies $\pi_\alpha^{-1}(U_\alpha) \cap F \neq \emptyset$ for all $F \in \mathcal{F}_0$ and all $\alpha \in \Lambda$. By property 2.[7] above we concluded that $\pi_\alpha^{-1}(U_\alpha) \in \mathcal{F}_0$ for all $\alpha \in \Lambda$ and then by property 1. that $\cap_{\alpha\in\Lambda}\pi_\alpha^{-1}(U_\alpha) \in \mathcal{F}_0$. In particular

$$\emptyset \neq F \cap \left(\cap_{\alpha\in\Lambda}\pi_\alpha^{-1}(U_\alpha)\right) \subset F \cap U \text{ for all } F \in \mathcal{F}_0$$

which shows $p \in \bar{F}$ for each $F \in \mathcal{F}_0$, i.e. $p \in C$. ■

Remark 11.35. Consider the following simple example where $X = [-1,1] \times [-1,1]$ and $\mathcal{F} = \{F_1, F_2\}$ as in Figure 11.5. Notice that $\pi_i(F_1) \cap \pi_i(F_2) =$

[7] Here is where we use that $\mathcal{F}_0$ is maximal among the collection of all, not just closed, sets having the finite intersection property and containing $\mathcal{F}$.

$[-1, 1]$ for each i and so gives no help in trying to find the i^{th} – coordinate of one of the two points in $F_1 \cap F_2$. This is why it is necessary to introduce the collection $\mathcal{F}_0$ in the proof of Theorem 11.34. In this case one might take $\mathcal{F}_0$ to be the collection of all subsets $F \subset X$ such that $p \in F$. We then have $\cap_{F \in \mathcal{F}_0} \pi_i(F) = \{p_i\}$, so the i^{th} – coordinate of p may now be determined by observing the sets, $\{\pi_i(F) : F \in \mathcal{F}_0\}$.

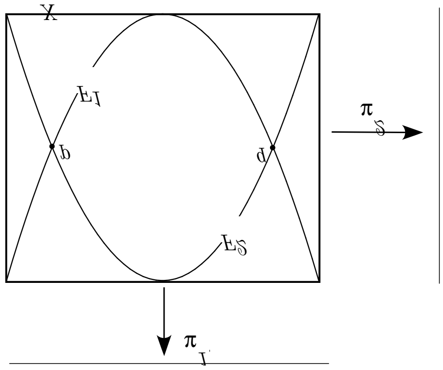


Fig. 11.5. Here $\mathcal{F} = \{F_1, F_2\}$ where F_1 and F_2 are the two parabolic arcs and $F_1 \cap F_2 = \{p, q\}$.

11.6 Exercises

Exercise 11.14. Prove Lemma 11.5.

Exercise 11.15. Let C be a closed proper subset of $\mathbb{R}^n$ and $x \in \mathbb{R}^n \setminus C$. Show there exists a $y \in C$ such that $d(x, y) = d_C(x)$.

Exercise 11.16. Let $\mathbb{F} = \mathbb{R}$ in this problem and $A \subset \ell^2(\mathbb{N})$ be defined by

$$\begin{aligned} A &= \{x \in \ell^2(\mathbb{N}) : x(n) \geq 1 + 1/n \text{ for some } n \in \mathbb{N}\} \\ &= \cup_{n=1}^{\infty} \{x \in \ell^2(\mathbb{N}) : x(n) \geq 1 + 1/n\}. \end{aligned}$$

Show A is a closed subset of $\ell^2(\mathbb{N})$ with the property that $d_A(0) = 1$ while there is no $y \in A$ such that $d(0, y) = 1$. (Remember that in general an infinite union of closed sets need not be closed.)

Exercise 11.17. Let $p \in [1, \infty]$ and X be an infinite set. Show directly, without using Theorem 11.15, the closed unit ball in $\ell^p(X)$ is not compact.

11.6.1 Ascoli-Arzela Theorem Problems

Exercise 11.18. Let $T \in (0,\infty)$ and $\mathcal{F} \subset C([0,T])$ be a family of functions such that:

1. $\dot{f}(t)$ exists for all $t \in (0,T)$ and $f \in \mathcal{F}$.
2. $\sup_{f\in\mathcal{F}} |f(0)| < \infty$ and
3. $M := \sup_{f\in\mathcal{F}} \sup_{t\in(0,T)} \left|\dot{f}(t)\right| < \infty$.

Show $\mathcal{F}$ is precompact in the Banach space $C([0,T])$ equipped with the norm $\|f\|_\infty = \sup_{t\in[0,T]} |f(t)|$.

Exercise 11.19 (Peano's Existence Theorem). Suppose $Z : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ is a bounded continuous function. Then for each $T < \infty$[8] there exists a solution to the differential equation

$$\dot{x}(t) = Z(t, x(t)) \text{ for } -T < t < T \text{ with } x(0) = x_0. \tag{11.7}$$

Do this by filling in the following outline for the proof.

1. Given $\varepsilon > 0$, show there exists a unique function $x_\varepsilon \in C([-\varepsilon,\infty) \to \mathbb{R}^d)$ such that $x_\varepsilon(t) := x_0$ for $-\varepsilon \le t \le 0$ and
$$x_\varepsilon(t) = x_0 + \int_0^t Z(\tau, x_\varepsilon(\tau - \varepsilon))d\tau \text{ for all } t \ge 0. \tag{11.8}$$
Here
$$\int_0^t Z(\tau, x_\varepsilon(\tau-\varepsilon))d\tau = \left(\int_0^t Z_1(\tau, x_\varepsilon(\tau-\varepsilon))d\tau, \dots, \int_0^t Z_d(\tau, x_\varepsilon(\tau-\varepsilon))d\tau\right)$$
where $Z = (Z_1, \dots, Z_d)$ and the integrals are either the Lebesgue or the Riemann integral since they are equal on continuous functions. **Hint:** For $t \in [0,\varepsilon]$, it follows from Eq. (11.8) that
$$x_\varepsilon(t) = x_0 + \int_0^t Z(\tau, x_0)d\tau.$$
Now that $x_\varepsilon(t)$ is known for $t \in [-\varepsilon,\varepsilon]$ it can be found by integration for $t \in [-\varepsilon, 2\varepsilon]$. The process can be repeated.
2. Then use Exercise 11.18 to show there exists $\{\varepsilon_k\}_{k=1}^\infty \subset (0,\infty)$ such that $\lim_{k\to\infty} \varepsilon_k = 0$ and x_{ε_k} converges to some $x \in C([0,T])$ with respect to the sup-norm: $\|x\|_\infty = \sup_{t\in[0,T]} |x(t)|$). Also show for this sequence that
$$\lim_{k\to\infty} \sup_{\varepsilon_k \le \tau \le T} |x_{\varepsilon_k}(\tau - \varepsilon_k) - x(\tau)| = 0.$$

[8] Using Corollary 11.30, we may in fact allow $T = \infty$.

3. Pass to the limit (**with justification**) in Eq. (11.8) with ε replaced by ε_k to show x satisfies
$$x(t) = x_0 + \int_0^t Z(\tau, x(\tau))d\tau \;\forall\; t \in [0, T].$$
4. Conclude from this that $\dot{x}(t)$ exists for $t \in (0, T)$ and that x solves Eq. (11.7).
5. Apply what you have just proved to the ODE,
$$\dot{y}(t) = -Z(-t, y(t)) \text{ for } 0 \le t < T \text{ with } y(0) = x_0.$$
Then extend $x(t)$ above to $(-T, T)$ by setting $x(t) = y(-t)$ if $t \in (-T, 0]$. Show x so defined solves Eq. (11.7) for $t \in (-T, T)$.

Exercise 11.20. Prove Theorem 11.32. **Hint:** First prove $C^{j,\beta}\left(\bar{\Omega}\right) \sqsubset\sqsubset C^{j,\alpha}\left(\bar{\Omega}\right)$ is compact if $0 \le \alpha < \beta \le 1$. Then use Lemma 11.18 repeatedly to handle all of the other cases.

11.6.2 Tychonoff's Theorem Problem

Exercise 11.21 (Tychonoff's Theorem for Compact Metric Spaces). Let us continue the Notation used in Exercise 6.12. Further assume that the spaces X_n are compact for all n. Show, without using Theorem 11.34, (X, d) is compact. **Hint:** Either use Cantor's method to show every sequence $\{x_m\}_{m=1}^\infty \subset X$ has a convergent subsequence or alternatively show (X, d) is complete and totally bounded. (Compare with Example 11.10.)

*** Beginning of WORK material. ***

Exercise 11.22. Let $X := \{0, 1\}^{\mathbb{R}}$ and τ be the product topology on X where $\{0, 1\}$ is equipped with the discrete topology. By Tychonoff's Theorem 11.34, (X, τ) is a compact space. Show (X, τ) is not separable and hence not metrizable.

Solution to Exercise (11.22). I don't know how to do this but here is a start. Suppose $D := \{f_n\}_{n=1}^\infty \subset X$ were a dense set and $g \in X$. We begin by observing that a basic open neighborhood of g is of the form
$$V_\Lambda := \{f \in X : f = g \text{ on } \Lambda\}$$
where $\Lambda \subset\subset \mathbb{R}$. Therefore to see that X is not separable, we must find a finite set $\Lambda \subset \mathbb{R}$ and a function $g : \mathbb{R} \to \{0, 1\}$ such that $f_n \ne g$ on Λ for any $n \in \mathbb{N}$. It is clear, to be able to do this we must assume that $\#(\Lambda) > 1$ for if $\#(\Lambda) = 1$, and $f_1 \equiv 0$ and $f_2 \equiv 1$, then $g = f_1$ or $g = f_2$ on Λ.

So the idea now is to show, there must exist some $s < t$ in $\mathbb{R}$ such that
$$(0, 1) \notin \{(f_n(s), f_n(t)) : n \in \mathbb{N}\}.$$
If this can be done, then we may choose $g : \mathbb{R} \to \{0, 1\}$ such that $g(s) = 0$ and $g(t) = 1$ and we will have shown that $g \notin \bar{D}$.

*** End of WORK material. ***

12

Locally Compact Hausdorff Spaces

In this section X will always be a topological space with topology τ. We are now interested in restrictions on τ in order to insure there are "plenty" of continuous functions. One such restriction is to assume $\tau = \tau_d$ – is the topology induced from a metric on X. For example the results in Lemma 6.15 and Theorem 7.4 above shows that metric spaces have lots of continuous functions.

The main thrust of this section is to study locally compact (and σ – compact) "Hausdorff" spaces as defined in Definitions 12.2 and 11.21. We will see again that this class of topological spaces have an ample supply of continuous functions. We will start out with the notion of a Hausdorff topology. The following example shows a pathology which occurs when there are not enough open sets in a topology.

Example 12.1. As in Example 10.36, let

$$X := \{1, 2, 3\} \text{ with } \tau := \{X, \emptyset, \{1, 2\}, \{2, 3\}, \{2\}\}.$$

Example 10.36 shows limits need not be unique in this space and moreover it is easy to verify that the only continuous functions, $f : Y \to \mathbb{R}$, are necessarily constant.

Definition 12.2 (Hausdorff Topology). *A topological space, (X, τ), is **Hausdorff** if for each pair of distinct points, $x, y \in X$, there exists disjoint open neighborhoods, U and V of x and y respectively. (Metric spaces are typical examples of Hausdorff spaces.)*

Remark 12.3. When τ is Hausdorff the "pathologies" appearing in Example 12.1 do not occur. Indeed if $x_n \to x \in X$ and $y \in X \setminus \{x\}$ we may choose $V \in \tau_x$ and $W \in \tau_y$ such that $V \cap W = \emptyset$. Then $x_n \in V$ a.a. implies $x_n \notin W$ for all but a finite number of n and hence $x_n \nrightarrow y$, so limits are unique.

Proposition 12.4. *Let (X_α, τ_α) be Hausdorff topological spaces. Then the product space $X_A = \prod_{\alpha \in A} X_\alpha$ equipped with the product topology is Hausdorff.*

Proof. Let $x, y \in X_A$ be distinct points. Then there exists $\alpha \in A$ such that $\pi_\alpha(x) = x_\alpha \neq y_\alpha = \pi_\alpha(y)$. Since X_α is Hausdorff, there exists disjoint open sets $U, V \subset X_\alpha$ such $\pi_\alpha(x) \in U$ and $\pi_\alpha(y) \in V$. Then $\pi_\alpha^{-1}(U)$ and $\pi_\alpha^{-1}(V)$ are disjoint open sets in X_A containing x and y respectively. ∎

Proposition 12.5. *Suppose that (X, τ) is a Hausdorff space, $K \sqsubset\sqsubset X$ and $x \in K^c$. Then there exists $U, V \in \tau$ such that $U \cap V = \emptyset$, $x \in U$ and $K \subset V$. In particular K is closed. (So compact subsets of Hausdorff topological spaces are closed.) More generally if K and F are two disjoint compact subsets of X, there exist disjoint open sets $U, V \in \tau$ such that $K \subset V$ and $F \subset U$.*

Proof. Because X is Hausdorff, for all $y \in K$ there exists $V_y \in \tau_y$ and $U_y \in \tau_x$ such that $V_y \cap U_y = \emptyset$. The cover $\{V_y\}_{y \in K}$ of K has a finite subcover, $\{V_y\}_{y \in \Lambda}$ for some $\Lambda \subset\subset K$. Let $V = \cup_{y \in \Lambda} V_y$ and $U = \cap_{y \in \Lambda} U_y$, then $U, V \in \tau$ satisfy $x \in U$, $K \subset V$ and $U \cap V = \emptyset$. This shows that K^c is open and hence that K is closed. Suppose that K and F are two disjoint compact subsets of X. For each $x \in F$ there exists disjoint open sets U_x and V_x such that $K \subset V_x$ and $x \in U_x$. Since $\{U_x\}_{x \in F}$ is an open cover of F, there exists a finite subset Λ of F such that $F \subset U := \cup_{x \in \Lambda} U_x$. The proof is completed by defining $V := \cap_{x \in \Lambda} V_x$. ∎

Exercise 12.1. Show any finite set X admits exactly one Hausdorff topology τ.

Exercise 12.2. Let (X, τ) and (Y, τ_Y) be topological spaces.

1. Show τ is Hausdorff iff $\Delta := \{(x, x) : x \in X\}$ is a closed in $X \times X$ equipped with the product topology $\tau \otimes \tau$.
2. Suppose τ is Hausdorff and $f, g : Y \to X$ are continuous maps. If $\overline{\{f = g\}}^Y = Y$ then $f = g$. **Hint:** make use of the map $f \times g : Y \to X \times X$ defined by $(f \times g)(y) = (f(y), g(y))$.

Exercise 12.3. Given an example of a topological space which has a non-closed compact subset.

Proposition 12.6. *Suppose that X is a compact topological space, Y is a Hausdorff topological space, and $f : X \to Y$ is a continuous bijection then f is a homeomorphism, i.e. $f^{-1} : Y \to X$ is continuous as well.*

Proof. Since closed subsets of compact sets are compact, continuous images of compact subsets are compact and compact subsets of Hausdorff spaces are closed, it follows that $\left(f^{-1}\right)^{-1}(C) = f(C)$ is closed in X for all closed subsets C of X. Thus f^{-1} is continuous. ∎

The next two results shows that locally compact Hausdorff spaces have plenty of open sets and plenty of continuous functions.

Proposition 12.7. *Suppose X is a locally compact Hausdorff space and $U \subset_o X$ and $K \sqsubset\sqsubset U$. Then there exists $V \subset_o X$ such that $K \subset V \subset \overline{V} \subset U \subset X$ and $\bar{V}$ is compact. (Compare with Proposition 11.25 above.)*

Proof. By local compactness, for all $x \in K$, there exists $U_x \in \tau_x$ such that $\bar{U}_x$ is compact. Since K is compact, there exists $\Lambda \subset\subset K$ such that $\{U_x\}_{x\in\Lambda}$ is a cover of K. The set $O = U \cap (\cup_{x\in\Lambda} U_x)$ is an open set such that $K \subset O \subset U$ and O is precompact since $\bar{O}$ is a closed subset of the compact set $\cup_{x\in\Lambda}\bar{U}_x$. ($\cup_{x\in\Lambda}\bar{U}_x$. is compact because it is a finite union of compact sets.) So by replacing U by O if necessary, we may assume that $\bar{U}$ is compact. Since $\bar{U}$ is compact and $\mathrm{bd}(U) = \bar{U} \cap U^c$ is a closed subset of $\bar{U}$, $\mathrm{bd}(U)$ is compact. Because $\mathrm{bd}(U) \subset U^c$, it follows that $\mathrm{bd}(U) \cap K = \emptyset$, so by Proposition 12.5, there exists disjoint open sets V and W such that $K \subset V$ and $\mathrm{bd}(U) \subset W$. By replacing V by $V \cap U$ if necessary we may further assume that $K \subset V \subset U$, see Figure 12.1. Because $\bar{U} \cap W^c$ is a closed set containing V and $\mathrm{bd}(U) \cap W^c = \emptyset$,

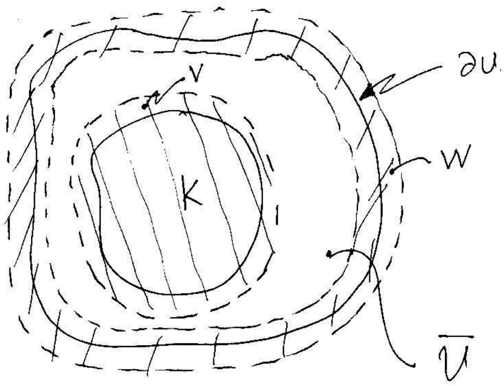


Fig. 12.1. The construction of V.

$$\bar{V} \subset \bar{U} \cap W^c = (U \cup \mathrm{bd}(U)) \cap W^c = U \cap W^c \subset U \subset \bar{U}.$$

Since $\bar{U}$ is compact it follows that $\bar{V}$ is compact and the proof is complete. ■

The following Lemma is analogous to Lemma 11.27.

Lemma 12.8 (Urysohn's Lemma for LCH Spaces). *Let X be a locally compact Hausdorff space and $K \sqsubset\sqsubset U \subset_o X$. Then there exists $f \prec U$ (see Definition 11.26) such that $f = 1$ on K. In particular, if K is compact and C is closed in X such that $K \cap C = \emptyset$, there exists $f \in C_c(X, [0,1])$ such that $f = 1$ on K and $f = 0$ on C.*

Proof. For notational ease later it is more convenient to construct $g := 1 - f$ rather than f. To motivate the proof, suppose $g \in C(X, [0,1])$ such that $g = 0$ on K and $g = 1$ on U^c. For $r > 0$, let $U_r = \{g < r\}$. Then for $0 < r < s \le 1$, $U_r \subset \{g \le r\} \subset U_s$ and since $\{g \le r\}$ is closed this implies

$$K \subset U_r \subset \bar{U}_r \subset \{g \le r\} \subset U_s \subset U.$$

Therefore associated to the function g is the collection open sets $\{U_r\}_{r>0} \subset \tau$ with the property that $K \subset U_r \subset \bar{U}_r \subset U_s \subset U$ for all $0 < r < s \le 1$ and $U_r = X$ if $r > 1$. Finally let us notice that we may recover the function g from the sequence $\{U_r\}_{r>0}$ by the formula

$$g(x) = \inf\{r > 0 : x \in U_r\}. \tag{12.1}$$

The idea of the proof to follow is to turn these remarks around and define g by Eq. (12.1).

Step 1. (Construction of the U_r.) Let

$$\mathbb{D} := \left\{k2^{-n} : k = 1, 2, \dots, 2^{-n}, n = 1, 2, \dots\right\}$$

be the dyadic rationals in $(0, 1]$. Use Proposition 12.7 to find a precompact open set U_1 such that $K \subset U_1 \subset \bar{U}_1 \subset U$. Apply Proposition 12.7 again to construct an open set $U_{1/2}$ such that

$$K \subset U_{1/2} \subset \bar{U}_{1/2} \subset U_1$$

and similarly use Proposition 12.7 to find open sets $U_{1/2}, U_{3/4} \subset_o X$ such that

$$K \subset U_{1/4} \subset \bar{U}_{1/4} \subset U_{1/2} \subset \bar{U}_{1/2} \subset U_{3/4} \subset \bar{U}_{3/4} \subset U_1.$$

Likewise there exists open set $U_{1/8}, U_{3/8}, U_{5/8}, U_{7/8}$ such that

$$\begin{aligned} K &\subset U_{1/8} \subset \bar{U}_{1/8} \subset U_{1/4} \subset \bar{U}_{1/4} \subset U_{3/8} \subset \bar{U}_{3/8} \subset U_{1/2} \\ &\subset \bar{U}_{1/2} \subset U_{5/8} \subset \bar{U}_{5/8} \subset U_{3/4} \subset \bar{U}_{3/4} \subset U_{7/8} \subset \bar{U}_{7/8} \subset U_1. \end{aligned}$$

Continuing this way inductively, one shows there exists precompact open sets $\{U_r\}_{r\in\mathbb{D}} \subset \tau$ such that

$$K \subset U_r \subset \overline{U}_r \subset U_s \subset U_1 \subset \bar{U}_1 \subset U$$

for all $r, s \in \mathbb{D}$ with $0 < r < s \le 1$.

Step 2. Let $U_r := X$ if $r > 1$ and define

$$g(x) = \inf\{r \in \mathbb{D} \cup (1, 2) : x \in U_r\},$$

see Figure 12.2. Then $g(x) \in [0, 1]$ for all $x \in X$, $g(x) = 0$ for $x \in K$ since $x \in K \subset U_r$ for all $r \in \mathbb{D}$. If $x \in U_1^c$, then $x \notin U_r$ for all $r \in \mathbb{D}$ and hence $g(x) = 1$. Therefore $f := 1 - g$ is a function such that $f = 1$ on K and $\{f \neq 0\} = \{g \neq 1\} \subset U_1 \subset \bar{U}_1 \subset U$ so that $\operatorname{supp}(f) = \overline{\{f \neq 0\}} \subset \bar{U}_1 \subset U$ is

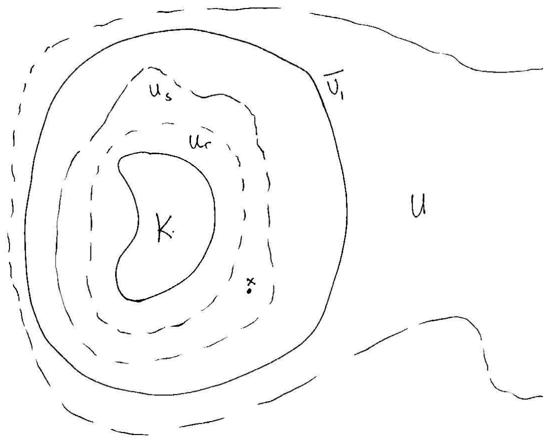


Fig. 12.2. Determining g from $\{U_r\}$.

a compact subset of U. Thus it only remains to show f, or equivalently g, is continuous.

Since $\mathcal{E} = \{(\alpha, \infty), (-\infty, \alpha) : \alpha \in \mathbb{R}\}$ generates the standard topology on $\mathbb{R}$, to prove g is continuous it suffices to show $\{g < \alpha\}$ and $\{g > \alpha\}$ are open sets for all $\alpha \in \mathbb{R}$. But $g(x) < \alpha$ iff there exists $r \in \mathbb{D} \cup (1, \infty)$ with $r < \alpha$ such that $x \in U_r$. Therefore

$$\{g < \alpha\} = \bigcup \{U_r : r \in \mathbb{D} \cup (1, \infty) \ni r < \alpha\}$$

which is open in X. If $\alpha \geq 1$, $\{g > \alpha\} = \emptyset$ and if $\alpha < 0$, $\{g > \alpha\} = X$. If $\alpha \in (0, 1)$, then $g(x) > \alpha$ iff there exists $r \in \mathbb{D}$ such that $r > \alpha$ and $x \notin U_r$. Now if $r > \alpha$ and $x \notin U_r$ then for $s \in \mathbb{D} \cap (\alpha, r)$, $x \notin \bar{U}_s \subset U_r$. Thus we have shown that

$$\{g > \alpha\} = \bigcup \left\{ \left(\overline{U}_s\right)^c : s \in \mathbb{D} \ni s > \alpha \right\}$$

which is again an open subset of X. ∎

Theorem 12.9 (Locally Compact Tietz Extension Theorem). *Let (X, τ) be a locally compact Hausdorff space, $K \sqsubset\sqsubset U \subset_o X$, $f \in C(K, \mathbb{R})$, $a = \min f(K)$ and $b = \max f(K)$. Then there exists $F \in C(X, [a, b])$ such that $F|_K = f$. Moreover given $c \in [a, b]$, F can be chosen so that $\operatorname{supp}(F - c) = \overline{\{F \neq c\}} \subset U$.*

The proof of this theorem is similar to Theorem 7.4 and will be left to the reader, see Exercise 12.5.

12.1 Locally compact form of Urysohn's Metrization Theorem

Notation 12.10 *Let* $Q := [0,1]^{\mathbb{N}}$ *denote the (infinite dimensional)* ***unit cube*** *in* $\mathbb{R}^{\mathbb{N}}$. *For* $a, b \in Q$ *let*

$$d(a,b) := \sum_{n=1}^{\infty} \frac{1}{2^n} |a_n - b_n| . \tag{12.2}$$

The metric introduced in Exercise 11.21 would be defined, in this context, as $\tilde{d}(a,b) := \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{|a_n - b_n|}{1+|a_n-b_n|}$. Since $1 \le 1+|a_n - b_n| \le 2$, it follows that $\tilde{d} \le d \le 2d$. So the metrics d and $\tilde{d}$ are equivalent and in particular the topologies induced by d and $\tilde{d}$ are the same. By Exercises 10.28, the $d-$topology on Q is the same as the product topology and by Tychonoff's Theorem 11.34 or by Exercise 11.21, (Q,d) is a compact metric space.

Theorem 12.11. *To every separable metric space* (X,ρ), *there exists a continuous injective map* $G : X \to Q$ *such that* $G : X \to G(X) \subset Q$ *is a homeomorphism. In short, any separable metrizable space* X *is homeomorphic to a subset of* (Q,d).

Remark 12.12. Notice that if we let $\rho'(x,y) := d(G(x), G(y))$, then ρ' induces the same topology on X as ρ and $G : (X,\rho') \to (Q,d)$ is isometric.

Proof. Let $D = \{x_n\}_{n=1}^{\infty}$ be a countable dense subset of X,

$$\phi(t) := \begin{cases} 1 & \text{if} \quad t \le 0 \\ 1-t & \text{if}\; 0 \le t \le 1 \\ 0 & \text{if} \quad t \ge 1, \end{cases}$$

(see Figure 12.3) and for $m, n \in \mathbb{N}$ let

$$f_{m,n}(x) := 1 - \phi\left(m\rho(x_n, x)\right).$$

Then $f_{m,n} = 0$ if $\rho(x,x_n) < 1/m$ and $f_{m,n} = 1$ if $\rho(x,x_n) > 2/m$. Let $\{g_k\}_{k=1}^{\infty}$ be an enumeration of $\{f_{m,n} : m,n \in \mathbb{N}\}$ and define $G : X \to Q$ by

$$G(x) = (g_1(x), g_2(x), \dots) \in Q.$$

We will now show $G : X \to G(X) \subset Q$ is a homeomorphism. To show G is injective suppose $x, y \in X$ and $\rho(x,y) = \delta \ge 1/m$. In this case we may find $x_n \in X$ such that $\rho(x,x_n) \le \frac{1}{2m}$, $\rho(y,x_n) \ge \delta - \frac{1}{2m} \ge \frac{1}{2m}$ and hence $f_{4m,n}(y) = 1$ while $f_{4m,n}(y) = 0$. From this it follows that $G(x) \ne G(y)$ if $x \ne y$ and hence G is injective. The continuity of G is a consequence of the continuity of each of the components g_i of G. So it only remains to show $G^{-1} : G(X) \to X$ is continuous. Given $a = G(x) \in G(X) \subset Q$ and $\varepsilon > 0$, choose $m \in \mathbb{N}$ and $x_n \in X$ such that $\rho(x_n, x) < \frac{1}{2m} < \frac{\varepsilon}{2}$. Then $f_{m,n}(x) = 0$

Fig. 12.3. The graph of the function ϕ.

and for $y \notin B(x_n, \frac{2}{m})$, $f_{m,n}(y) = 1$. So if k is chosen so that $g_k = f_{m,n}$, we have shown that for

$$d(G(y), G(x)) \geq 2^{-k} \text{ for } y \notin B(x_n, 2/m)$$

or equivalently put, if

$$d(G(y), G(x)) < 2^{-k} \text{ then } y \in B(x_n, 2/m) \subset B(x, 1/m) \subset B(x, \varepsilon).$$

This shows that if $G(y)$ is sufficiently close to $G(x)$ then $\rho(y, x) < \varepsilon$, i.e. G^{-1} is continuous at $a = G(x)$. ■

Theorem 12.13 (Urysohn Metrization Theorem for LCH's). *Every second countable locally compact Hausdorff space, (X, τ), is metrizable, i.e. there is a metric ρ on X such that $\tau = \tau_\rho$. Moreover, ρ may be chosen so that X is isometric to a subset $Q_0 \subset Q$ equipped with the metric d in Eq. (12.2). In this metric X is totally bounded and hence the completion of X (which is isometric to $\bar{Q}_0 \subset Q$) is compact. (Also see Theorem 12.43.)*

Proof. Let $\mathcal{B}$ be a countable base for τ and set

$$\Gamma := \{(U, V) \in \mathcal{B} \times \mathcal{B} \mid \bar{U} \subset V \text{ and } \bar{U} \text{ is compact}\}.$$

To each $O \in \tau$ and $x \in O$ there exist $(U, V) \in \Gamma$ such that $x \in U \subset V \subset O$. Indeed, since $\mathcal{B}$ is a base for τ, there exists $V \in \mathcal{B}$ such that $x \in V \subset O$. Now apply Proposition 12.7 to find $U' \subset_o X$ such that $x \in U' \subset \bar{U}' \subset V$ with $\bar{U}'$ being compact. Since $\mathcal{B}$ is a base for τ, there exists $U \in \mathcal{B}$ such that $x \in U \subset U'$ and since $\bar{U} \subset \bar{U}'$, $\bar{U}$ is compact so $(U, V) \in \Gamma$. In particular this shows that $\mathcal{B}' := \{U \in \mathcal{B} : (U, V) \in \Gamma \text{ for some } V \in \mathcal{B}\}$ is still a base for τ. If Γ is a finite, then $\mathcal{B}'$ is finite and τ only has a finite number of elements as well. Since (X, τ) is Hausdorff, it follows that X is a finite set. Letting $\{x_n\}_{n=1}^N$ be an enumeration of X, define $T : X \to Q$ by $T(x_n) = e_n$ for $n = 1, 2, \dots, N$ where $e_n = (0, 0, \dots, 0, 1, 0, \dots)$, with the 1 occurring in the n^{th} spot. Then $\rho(x, y) := d(T(x), T(y))$ for $x, y \in X$ is the desired metric.

So we may now assume that Γ is an infinite set and let $\{(U_n, V_n)\}_{n=1}^{\infty}$ be an enumeration of Γ. By Urysohn's Lemma 12.8 there exists $f_{U,V} \in C(X,[0,1])$ such that $f_{U,V} = 0$ on $\bar{U}$ and $f_{U,V} = 1$ on V^c. Let $\mathcal{F} := \{f_{U,V} \mid (U,V) \in \Gamma\}$ and set $f_n := f_{U_n,V_n}$ – an enumeration of $\mathcal{F}$. We will now show that

$$\rho(x,y) := \sum_{n=1}^{\infty} \frac{1}{2^n} |f_n(x) - f_n(y)|$$

is the desired metric on X. The proof will involve a number of steps.

1. (ρ is a metric on X.) It is routine to show ρ satisfies the triangle inequality and ρ is symmetric. If $x, y \in X$ are distinct points then there exists $(U_{n_0}, V_{n_0}) \in \Gamma$ such that $x \in U_{n_0}$ and $V_{n_0} \subset O := \{y\}^c$. Since $f_{n_0}(x) = 0$ and $f_{n_0}(y) = 1$, it follows that $\rho(x,y) \geq 2^{-n_0} > 0$.
2. (Let $\tau_0 = \tau(f_n : n \in \mathbb{N})$, then $\tau = \tau_0 = \tau_\rho$.) As usual we have $\tau_0 \subset \tau$. Since, for each $x \in X$, $y \to \rho(x,y)$ is τ_0 – continuous (being the uniformly convergent sum of continuous functions), it follows that $B_x(\varepsilon) := \{y \in X : \rho(x,y) < \varepsilon\} \in \tau_0$ for all $x \in X$ and $\varepsilon > 0$. Thus $\tau_\rho \subset \tau_0 \subset \tau$. Suppose that $O \in \tau$ and $x \in O$. Let $(U_{n_0}, V_{n_0}) \in \Gamma$ be such that $x \in U_{n_0}$ and $V_{n_0} \subset O$. Then $f_{n_0}(x) = 0$ and $f_{n_0} = 1$ on O^c. Therefore if $y \in X$ and $f_{n_0}(y) < 1$, then $y \in O$ so $x \in \{f_{n_0} < 1\} \subset O$. This shows that O may be written as a union of elements from τ_0 and therefore $O \in \tau_0$. So $\tau \subset \tau_0$ and hence $\tau = \tau_0$. Moreover, if $y \in B_x(2^{-n_0})$ then $2^{-n_0} > \rho(x,y) \geq 2^{-n_0} f_{n_0}(y)$ and therefore $x \in B_x(2^{-n_0}) \subset \{f_{n_0} < 1\} \subset O$. This shows O is ρ – open and hence $\tau_\rho \subset \tau_0 \subset \tau \subset \tau_\rho$.
3. (X is isometric to some $Q_0 \subset Q$.) Let $T : X \to Q$ be defined by $T(x) = (f_1(x), f_2(x), \dots, f_n(x), \dots)$. Then T is an isometry by the very definitions of d and ρ and therefore X is isometric to $Q_0 := T(X)$. Since Q_0 is a subset of the compact metric space (Q,d), Q_0 is totally bounded and therefore X is totally bounded.

■

BRUCE: Add Stone Chech Compactification results.

12.2 Partitions of Unity

Definition 12.14. *Let (X,τ) be a topological space and $X_0 \subset X$ be a set. A collection of sets $\{B_\alpha\}_{\alpha \in A} \subset 2^X$ is* ***locally finite*** *on X_0 if for all $x \in X_0$, there is an open neighborhood $N_x \in \tau$ of x such that $\#\{\alpha \in A : B_\alpha \cap N_x \neq \emptyset\} < \infty$.*

Definition 12.15. *Suppose that $\mathcal{U}$ is an open cover of $X_0 \subset X$. A collection $\{\phi_\alpha\}_{\alpha \in A} \subset C(X,[0,1])$ ($N = \infty$ is allowed here) is a* ***partition of unity*** *on X_0 subordinate to the cover $\mathcal{U}$ if:*

1. *for all α there is a $U \in \mathcal{U}$ such that* $\operatorname{supp}(\phi_\alpha) \subset U$,

2. *the collection of sets,* $\{\mathrm{supp}(\phi_\alpha)\}_{\alpha\in A}$*, is locally finite on* X*, and*
3. $\sum_{\alpha\in A}\phi_\alpha = 1$ *on* X_0.

Notice by item 2. that, for each $x \in X$*, there is a neighborhood* N_x *such that*

$$\Lambda := \{\alpha \in A : \mathrm{supp}(\phi_\alpha) \cap N_x \neq \emptyset\}$$

is a finite set. Therefore, $\sum_{\alpha\in A}\phi_\alpha|_{N_x} = \sum_{\alpha\in\Lambda}\phi_\alpha|_{N_x}$ *which shows the sum* $\sum_{\alpha\in A}\phi_\alpha$ *is well defined and defines a continuous function on* N_x *and therefore on* X *since continuity is a local property. We will summarize these last comments by saying the sum,* $\sum_{\alpha\in A}\phi_\alpha$*, is* ***locally finite****.*

Proposition 12.16 (Partitions of Unity: The Compact Case). *Suppose that* X *is a locally compact Hausdorff space,* $K \subset X$ *is a compact set and* $\mathcal{U} = \{U_j\}_{j=1}^n$ *is an open cover of* K*. Then there exists a partition of unity* $\{h_j\}_{j=1}^n$ *of* K *such that* $h_j \prec U_j$ *for all* $j = 1, 2, \ldots, n$*.*

Proof. For all $x \in K$ choose a precompact open neighborhood, V_x, of x such that $\overline{V}_x \subset U_j$. Since K is compact, there exists a finite subset, Λ, of K such that $K \subset \bigcup_{x\in\Lambda} V_x$. Let

$$F_j = \cup\left\{\bar{V}_x : x \in \Lambda \text{ and } \overline{V}_x \subset U_j\right\}.$$

Then F_j is compact, $F_j \subset U_j$ for all j, and $K \subset \cup_{j=1}^n F_j$. By Urysohn's Lemma 12.8 there exists $f_j \prec U_j$ such that $f_j = 1$ on F_j for $j = 1, 2, \ldots, n$ and by convention let $f_{n+1} \equiv 1$. We will now give two methods to finish the proof.

Method 1. Let $h_1 = f_1$, $h_2 = f_2(1 - h_1) = f_2(1 - f_1)$,

$$h_3 = f_3(1 - h_1 - h_2) = f_3(1 - f_1 - (1 - f_1)f_2) = f_3(1 - f_1)(1 - f_2)$$

and continue on inductively to define

$$h_k = (1 - h_1 - \cdots - h_{k-1})f_k = f_k \cdot \prod_{j=1}^{k-1}(1 - f_j)\,\forall\, k = 2, 3, \ldots, n \tag{12.3}$$

and to show

$$h_{n+1} = (1 - h_1 - \cdots - h_n) \cdot 1 = 1 \cdot \prod_{j=1}^{n}(1 - f_j). \tag{12.4}$$

From these equations it clearly follows that $h_j \in C_c(X, [0, 1])$ and that $\mathrm{supp}(h_j) \subset \mathrm{supp}(f_j) \subset U_j$, i.e. $h_j \prec U_j$. Since $\prod_{j=1}^n(1 - f_j) = 0$ on K, $\sum_{j=1}^n h_j = 1$ on K and $\{h_j\}_{j=1}^n$ is the desired partition of unity.

Method 2. Let $g := \sum_{j=1}^n f_j \in C_c(X)$. Then $g \geq 1$ on K and hence $K \subset \{g > \frac{1}{2}\}$. Choose $\phi \in C_c(X, [0, 1])$ such that $\phi = 1$ on K and $\mathrm{supp}(\phi) \subset$

$\{g > \frac{1}{2}\}$ and define $f_0 := 1 - \phi$. Then $f_0 = 0$ on K, $f_0 = 1$ if $g \leq \frac{1}{2}$ and therefore,

$$f_0 + f_1 + \cdots + f_n = f_0 + g > 0$$

on X. The desired partition of unity may be constructed as

$$h_j(x) = \frac{f_j(x)}{f_0(x) + \cdots + f_n(x)}.$$

Indeed $\mathrm{supp}\,(h_j) = \mathrm{supp}\,(f_j) \subset U_j$, $h_j \in C_c(X, [0,1])$ and on K,

$$h_1 + \cdots + h_n = \frac{f_1 + \cdots + f_n}{f_0 + f_1 + \cdots + f_n} = \frac{f_1 + \cdots + f_n}{f_1 + \cdots + f_n} = 1.$$

■

Proposition 12.17. *Let (X, τ) be a locally compact and σ – compact Hausdorff space. Suppose that $\mathcal{U} \subset \tau$ is an open cover of X. Then we may construct two locally finite open covers $\mathcal{V} = \{V_i\}_{i=1}^N$ and $\mathcal{W} = \{W_i\}_{i=1}^N$ of X ($N = \infty$ is allowed here) such that:*

1. *$W_i \subset \bar{W}_i \subset V_i \subset \bar{V}_i$ and $\bar{V}_i$ is compact for all i.*
2. *For each i there exist $U \in \mathcal{U}$ such that $\bar{V}_i \subset U$.*

Proof. By Remark 11.24, there exists an open cover of $\mathcal{G} = \{G_n\}_{n=1}^\infty$ of X such that $G_n \subset \bar{G}_n \subset G_{n+1}$. Then $X = \cup_{k=1}^\infty (\bar{G}_k \setminus \bar{G}_{k-1})$, where by convention $G_{-1} = G_0 = \emptyset$. For the moment fix $k \geq 1$. For each $x \in \bar{G}_k \setminus G_{k-1}$, let $U_x \in \mathcal{U}$ be chosen so that $x \in U_x$ and by Proposition 12.7 choose an open neighborhood N_x of x such that $\bar{N}_x \subset U_x \cap (G_{k+1} \setminus \bar{G}_{k-2})$, see Figure 12.4 below. Since $\{N_x\}_{x \in \bar{G}_k \setminus G_{k-1}}$ is an open cover of the compact set $\bar{G}_k \setminus G_{k-1}$, there exist a finite subset $\Gamma_k \subset \{N_x\}_{x \in \bar{G}_k \setminus G_{k-1}}$ which also covers $\bar{G}_k \setminus G_{k-1}$.

By construction, for each $W \in \Gamma_k$, there is a $U \in \mathcal{U}$ such that $\bar{W} \subset U \cap (G_{k+1} \setminus \bar{G}_{k-2})$ and by another application of Proposition 12.7, there exists an open set V_W such that $\bar{W} \subset V_W \subset \bar{V}_W \subset U \cap (G_{k+1} \setminus \bar{G}_{k-2})$. We now choose and enumeration $\{W_i\}_{i=1}^N$ of the countable open cover, $\cup_{k=1}^\infty \Gamma_k$, of X and define $V_i = V_{W_i}$. Then the collection $\{W_i\}_{i=1}^N$ and $\{V_i\}_{i=1}^N$ are easily checked to satisfy all the conclusions of the proposition. In particular notice that for each k; $V_i \cap G_k \neq \emptyset$ for only a finite number of i's. ■

Theorem 12.18 (Partitions of Unity for σ – Compact LCH Spaces). *Let (X, τ) be locally compact, σ – compact and Hausdorff and let $\mathcal{U} \subset \tau$ be an open cover of X. Then there exists a partition of unity of $\{h_i\}_{i=1}^N$ ($N = \infty$ is allowed here) subordinate to the cover $\mathcal{U}$ such that* $\mathrm{supp}(h_i)$ *is compact for all i.*

Proof. Let $\mathcal{V} = \{V_i\}_{i=1}^N$ and $\mathcal{W} = \{W_i\}_{i=1}^N$ be open covers of X with the properties described in Proposition 12.17. By Urysohn's Lemma 12.8, there

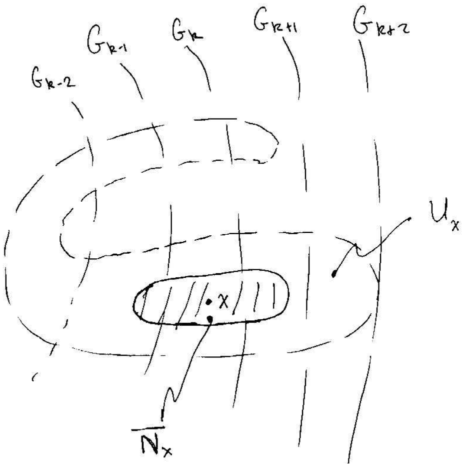


Fig. 12.4. Constructing the $\{W_i\}_{i=1}^N$.

exists $f_i \prec V_i$ such that $f_i = 1$ on $\bar{W}_i$ for each i. As in the proof of Proposition 12.16 there are two methods to finish the proof.

Method 1. Define $h_1 = f_1$, h_j by Eq. (12.3) for all other j. Then as in Eq. (12.4), for all $n < N + 1$,

$$1 - \sum_{j=1}^{\infty} h_j = \lim_{n\to\infty} \left(f_n \prod_{j=1}^{n} (1 - f_j) \right) = 0$$

since for $x \in X$, $f_j(x) = 1$ for some j. As in the proof of Proposition 12.16, it is easily checked that $\{h_i\}_{i=1}^N$ is the desired partition of unity.

Method 2. Let $f := \sum_{i=1}^N f_i$, a locally finite sum, so that $f \in C(X)$. Since $\{W_i\}_{i=1}^\infty$ is a cover of X, $f \geq 1$ on X so that $1/f \in C(X))$ as well. The functions $h_i := f_i/f$ for $i = 1, 2, \ldots, N$ give the desired partition of unity. ■

Lemma 12.19. *Let (X, τ) be a locally compact Hausdorff space.*

1. *A subset $E \subset X$ is closed iff $E \cap K$ is closed for all $K \sqsubset\sqsubset X$.*
2. *Let $\{C_\alpha\}_{\alpha\in A}$ be a locally finite collection of closed subsets of X, then $C = \cup_{\alpha\in A} C_\alpha$ is closed in X. (Recall that in general closed sets are only closed under finite unions.)*

Proof. 1. Since compact subsets of Hausdorff spaces are closed, $E \cap K$ is closed if E is closed and K is compact. Now suppose that $E \cap K$ is closed for all compact subsets $K \subset X$ and let $x \in E^c$. Since X is locally compact, there exists a precompact open neighborhood, V, of x.[1] By assumption $E \cap \bar{V}$

[1] If X were a metric space we could finish the proof as follows. If there does not exist an open neighborhood of x which is disjoint from E, then there would exists

is closed so $x \in \left(E \cap \bar{V}\right)^c$ – an open subset of X. By Proposition 12.7 there exists an open set U such that $x \in U \subset \bar{U} \subset \left(E \cap \bar{V}\right)^c$, see Figure 12.5. Let

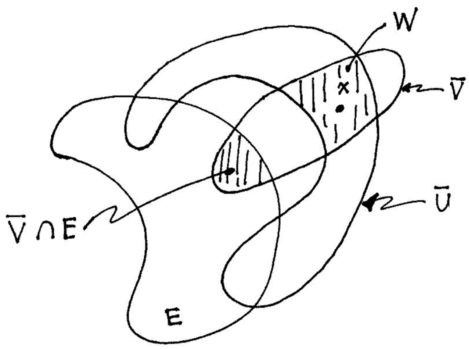


Fig. 12.5. Showing E^c is open.

$W := U \cap V$. Since

$$W \cap E = U \cap V \cap E \subset U \cap \bar{V} \cap E = \emptyset,$$

and W is an open neighborhood of x and $x \in E^c$ was arbitrary, we have shown E^c is open hence E is closed.

2. Let K be a compact subset of X and for each $x \in K$ let N_x be an open neighborhood of x such that $\#\{\alpha \in A : C_\alpha \cap N_x \neq \emptyset\} < \infty$. Since K is compact, there exists a finite subset $\Lambda \subset K$ such that $K \subset \cup_{x \in \Lambda} N_x$. Letting $\Lambda_0 := \{\alpha \in A : C_\alpha \cap K \neq \emptyset\}$, then

$$\#\left(\Lambda_0\right) \leq \sum_{x \in \Lambda} \#\{\alpha \in A : C_\alpha \cap N_x \neq \emptyset\} < \infty$$

and hence $K \cap \left(\cup_{\alpha \in A} C_\alpha\right) = K \cap \left(\cup_{\alpha \in \Lambda_0} C_\alpha\right)$. The set $\left(\cup_{\alpha \in \Lambda_0} C_\alpha\right)$ is a finite union of closed sets and hence closed. Therefore, $K \cap \left(\cup_{\alpha \in A} C_\alpha\right)$ is closed and by item 1. it follows that $\cup_{\alpha \in A} C_\alpha$ is closed as well. ■

Corollary 12.20. *Let (X, τ) be a locally compact and σ – compact Hausdorff space and $\mathcal{U} = \{U_\alpha\}_{\alpha \in A} \subset \tau$ be an open cover of X. Then there exists a partition of unity of $\{h_\alpha\}_{\alpha \in A}$ subordinate to the cover $\mathcal{U}$ such that* $\operatorname{supp}(h_\alpha) \subset U_\alpha$ *for all $\alpha \in A$. (Notice that we do not assert that h_α has compact support. However if $\bar{U}_\alpha$ is compact then* $\operatorname{supp}(h_\alpha)$ *will be compact.)*

$x_n \in E$ such that $x_n \to x$. Since $E \cap \bar{V}$ is closed and $x_n \in E \cap \bar{V}$ for all large n, it follows (see Exercise 6.4) that $x \in E \cap \bar{V}$ and in particular that $x \in E$. But we chose $x \in E^c$.

Proof. By the σ – compactness of X, we may choose a countable subset, $\{\alpha_i\}_{i=1}^N$ ($N = \infty$ allowed here), of A such that $\{U_i := U_{\alpha_i}\}_{i=1}^N$ is still an open cover of X. Let $\{g_j\}_{j=1}^\infty$ be a partition of unity[2] subordinate to the cover $\{U_i\}_{i=1}^N$ as in Theorem 12.18. Define $\tilde{\Gamma}_k := \{j : \operatorname{supp}(g_j) \subset U_k\}$ and $\Gamma_k := \tilde{\Gamma}_k \setminus \left(\cup_{j=1}^{k-1} \tilde{\Gamma}_k\right)$, where by convention $\tilde{\Gamma}_0 = \emptyset$. Then

$$\mathbb{N} = \bigcup_{k=1}^{\infty} \tilde{\Gamma}_k = \coprod_{k=1}^{\infty} \Gamma_k.$$

If $\Gamma_k = \emptyset$ let $h_k := 0$ otherwise let $h_k := \sum_{j \in \Gamma_k} g_j$, a locally finite sum. Then

$$\sum_{k=1}^{N} h_k = \sum_{j=1}^{\infty} g_j = 1.$$

By Item 2. of Lemma 12.19, $\cup_{j \in \Gamma_k} \operatorname{supp}(g_j)$ is closed and therefore,

$$\operatorname{supp}(h_k) = \overline{\{h_k \neq 0\}} = \overline{\cup_{j \in \Gamma_k} \{g_j \neq 0\}} \subset \cup_{j \in \Gamma_k} \operatorname{supp}(g_j) \subset U_k$$

and hence $h_k \prec U_k$ and the sum $\sum_{k=1}^N h_k$ is still locally finite. (Why?) The desired partition of unity is now formed by letting $h_{\alpha_k} := h_k$ for $k < N + 1$ and $h_\alpha \equiv 0$ if $\alpha \notin \{\alpha_i\}_{i=1}^N$. ∎

Corollary 12.21. *Let (X, τ) be a locally compact and σ – compact Hausdorff space and A, B be disjoint closed subsets of X. Then there exists $f \in C(X, [0, 1])$ such that $f = 1$ on A and $f = 0$ on B. In fact f can be chosen so that* $\operatorname{supp}(f) \subset B^c$.

Proof. Let $U_1 = A^c$ and $U_2 = B^c$, then $\{U_1, U_2\}$ is an open cover of X. By Corollary 12.20 there exists $h_1, h_2 \in C(X, [0, 1])$ such that $\operatorname{supp}(h_i) \subset U_i$ for $i = 1, 2$ and $h_1 + h_2 = 1$ on X. The function $f = h_2$ satisfies the desired properties. ∎

12.3 $C_0(X)$ and the Alexanderov Compactification

Definition 12.22. *Let (X, τ) be a topological space. A continuous function $f : X \to \mathbb{C}$ is said to **vanish at infinity** if $\{|f| \geq \varepsilon\}$ is compact in X for all $\varepsilon > 0$. The functions, $f \in C(X)$, vanishing at infinity will be denoted by $C_0(X)$. (Notice that $C_0(X) = C(X)$ whenever X is compact.)*

Proposition 12.23. *Let X be a topological space, $BC(X)$ be the space of bounded continuous functions on X with the supremum norm topology. Then*

[2] So as to simplify the indexing we assume there countable number of g_j's. This can always be arranged by taking $g_k \equiv 0$ for large k if necessary.

1. *$C_0(X)$ is a closed subspace of $BC(X)$.*
2. *If we further assume that X is a locally compact Hausdorff space, then $C_0(X) = \overline{C_c(X)}$.*

Proof.

1. If $f \in C_0(X)$, $K_1 := \{|f| \geq 1\}$ is a compact subset of X and therefore $f(K_1)$ is a compact and hence bounded subset of $\mathbb{C}$ and so $M := \sup_{x \in K_1} |f(x)| < \infty$. Therefore $\|f\|_\infty \leq M \vee 1 < \infty$ showing $f \in BC(X)$. Now suppose $f_n \in C_0(X)$ and $f_n \to f$ in $BC(X)$. Let $\varepsilon > 0$ be given and choose n sufficiently large so that $\|f - f_n\|_\infty \leq \varepsilon/2$. Since
$$|f| \leq |f_n| + |f - f_n| \leq |f_n| + \|f - f_n\|_\infty \leq |f_n| + \varepsilon/2,$$
$$\{|f| \geq \varepsilon\} \subset \{|f_n| + \varepsilon/2 \geq \varepsilon\} = \{|f_n| \geq \varepsilon/2\}.$$
Because $\{|f| \geq \varepsilon\}$ is a closed subset of the compact set $\{|f_n| \geq \varepsilon/2\}$, $\{|f| \geq \varepsilon\}$ is compact and we have shown $f \in C_0(X)$.
2. Since $C_0(X)$ is a closed subspace of $BC(X)$ and $C_c(X) \subset C_0(X)$, we always have $\overline{C_c(X)} \subset C_0(X)$. Now suppose that $f \in C_0(X)$ and let $K_n := \{|f| \geq \frac{1}{n}\} \sqsubset\sqsubset X$. By Lemma 12.8 we may choose $\phi_n \in C_c(X, [0,1])$ such that $\phi_n \equiv 1$ on K_n. Define $f_n := \phi_n f \in C_c(X)$. Then
$$\|f - f_n\|_u = \|(1 - \phi_n) f\|_\infty \leq \frac{1}{n} \to 0 \text{ as } n \to \infty.$$
This shows that $f \in \overline{C_c(X)}$.

■

Proposition 12.24 (Alexanderov Compactification). *Suppose that (X, τ) is a non-compact locally compact Hausdorff space. Let $X^* = X \cup \{\infty\}$, where $\{\infty\}$ is a new symbol not in X. The collection of sets,*
$$\tau^* = \tau \cup \{X^* \setminus K : K \sqsubset\sqsubset X\} \subset 2^{X^*},$$
is a topology on X^* and (X^*, τ^*) is a compact Hausdorff space. Moreover $f \in C(X)$ extends continuously to X^* iff $f = g + c$ with $g \in C_0(X)$ and $c \in \mathbb{C}$ in which case the extension is given by $f(\infty) = c$.

Proof. 1. (τ^* is a topology.) Let $\mathcal{F} := \{F \subset X^* : X^* \setminus F \in \tau^*\}$, i.e. $F \in \mathcal{F}$ iff F is a compact subset of X or $F = F_0 \cup \{\infty\}$ with F_0 being a closed subset of X. Since the finite union of compact (closed) subsets is compact (closed), it is easily seen that $\mathcal{F}$ is closed under finite unions. Because arbitrary intersections of closed subsets of X are closed and closed subsets of compact subsets of X are compact, it is also easily checked that $\mathcal{F}$ is closed under arbitrary intersections. Therefore $\mathcal{F}$ satisfies the axioms of the closed subsets associated to a topology and hence τ^* is a topology.

2. ((X^*,τ^*) is a Hausdorff space.) It suffices to show any point $x \in X$ can be separated from ∞. To do this use Proposition 12.7 to find an open precompact neighborhood, U, of x. Then U and $V := X^* \setminus \bar{U}$ are disjoint open subsets of X^* such that $x \in U$ and $\infty \in V$.

3. ((X^*,τ^*) is compact.) Suppose that $\mathcal{U} \subset \tau^*$ is an open cover of X^*. Since $\mathcal{U}$ covers ∞, there exists a compact set $K \subset X$ such that $X^* \setminus K \in \mathcal{U}$. Clearly X is covered by $\mathcal{U}_0 := \{V \setminus \{\infty\} : V \in \mathcal{U}\}$ and by the definition of τ^* (or using (X^*,τ^*) is Hausdorff), $\mathcal{U}_0$ is an open cover of X. In particular $\mathcal{U}_0$ is an open cover of K and since K is compact there exists $\Lambda \subset\subset \mathcal{U}$ such that $K \subset \cup \{V \setminus \{\infty\} : V \in \Lambda\}$. It is now easily checked that $\Lambda \cup \{X^* \setminus K\} \subset \mathcal{U}$ is a finite subcover of X^*.

4. (Continuous functions on $C(X^*)$ statements.) Let $i : X \to X^*$ be the inclusion map. Then i is continuous and open, i.e. $i(V)$ is open in X^* for all V open in X. If $f \in C(X^*)$, then $g = f|_X - f(\infty) = f \circ i - f(\infty)$ is continuous on X. Moreover, for all $\varepsilon > 0$ there exists an open neighborhood $V \in \tau^*$ of ∞ such that

$$|g(x)| = |f(x) - f(\infty)| < \varepsilon \text{ for all } x \in V.$$

Since V is an open neighborhood of ∞, there exists a compact subset, $K \subset X$, such that $V = X^* \setminus K$. By the previous equation we see that $\{x \in X : |g(x)| \geq \varepsilon\} \subset K$, so $\{|g| \geq \varepsilon\}$ is compact and we have shown g vanishes at ∞.

Conversely if $g \in C_0(X)$, extend g to X^* by setting $g(\infty) = 0$. Given $\varepsilon > 0$, the set $K = \{|g| \geq \varepsilon\}$ is compact, hence $X^* \setminus K$ is open in X^*. Since $g(X^* \setminus K) \subset (-\varepsilon, \varepsilon)$ we have shown that g is continuous at ∞. Since g is also continuous at all points in X it follows that g is continuous on X^*. Now it $f = g + c$ with $c \in \mathbb{C}$ and $g \in C_0(X)$, it follows by what we just proved that defining $f(\infty) = c$ extends f to a continuous function on X^*. ■

Example 12.25. Let X be an uncountable set and τ be the discrete topology on X. Let $(X^* = X \cup \{\infty\}, \tau^*)$ be the one point compactification of X. The smallest dense subset of X^* is the uncountable set X. Hence X^* is a compact but non-separable and hence non-metrizable space.

The next proposition gathers a number of results involving countability assumptions which have appeared in the exercises.

Proposition 12.26 (Summary). *Let (X,τ) be a topological space.*

1. *If (X,τ) is second countable, then (X,τ) is separable; see Exercise 10.11.*
2. *If (X,τ) is separable and metrizable then (X,τ) is second countable; see Exercise 10.12.*
3. *If (X,τ) is locally compact and metrizable then (X,τ) is σ – compact iff (X,τ) is separable; see Exercises 11.10 and 11.11.*
4. *If (X,τ) is locally compact and second countable, then (X,τ) is σ - compact, see Exercise 11.8.*
5. *If (X,τ) is locally compact and metrizable, then (X,τ) is σ – compact iff (X,τ) is separable, see Exercises 11.9 and 11.10.*

12.4 Stone-Weierstrass Theorem

We now wish to generalize Theorem 8.34 to more general topological spaces. We will first need some definitions.

Definition 12.27. *Let X be a topological space and $\mathcal{A} \subset C(X) = C(X,\mathbb{R})$ or $C(X,\mathbb{C})$ be a collection of functions. Then*

1. *$\mathcal{A}$ is said to **separate points** if for all distinct points $x, y \in X$ there exists $f \in \mathcal{A}$ such that $f(x) \neq f(y)$.*
2. *$\mathcal{A}$ is an **algebra** if $\mathcal{A}$ is a vector subspace of $C(X)$ which is closed under pointwise multiplication. (**Note well:** we do not assume $1 \in \mathcal{A}$.)*
3. *$\mathcal{A} \subset C(X,\mathbb{R})$ is called a **lattice** if $f \vee g := \max(f,g)$ and $f \wedge g = \min(f,g) \in \mathcal{A}$ for all $f, g \in \mathcal{A}$.*
4. *$\mathcal{A} \subset C(X,\mathbb{C})$ is closed under conjugation if $\bar{f} \in \mathcal{A}$ whenever $f \in \mathcal{A}$.*

Remark 12.28. If X is a topological space such that $C(X,\mathbb{R})$ separates points then X is Hausdorff. Indeed if $x, y \in X$ and $f \in C(X,\mathbb{R})$ such that $f(x) \neq f(y)$, then $f^{-1}(J)$ and $f^{-1}(I)$ are disjoint open sets containing x and y respectively when I and J are disjoint intervals containing $f(x)$ and $f(y)$ respectively.

Lemma 12.29. *If $\mathcal{A}$ is a closed sub-algebra of $BC(X,\mathbb{R})$ then $|f| \in \mathcal{A}$ for all $f \in \mathcal{A}$ and $\mathcal{A}$ is a lattice.*

Proof. Let $f \in \mathcal{A}$ and let $M = \sup_{x \in X} |f(x)|$. Using Theorem 8.34 or Exercise 12.12, there are polynomials $p_n(t)$ such that

$$\lim_{n\to\infty} \sup_{|t| \le M} ||t| - p_n(t)| = 0.$$

By replacing p_n by $p_n - p_n(0)$ if necessary we may assume that $p_n(0) = 0$. Since $\mathcal{A}$ is an algebra, it follows that $f_n = p_n(f) \in \mathcal{A}$ and $|f| \in \mathcal{A}$, because $|f|$ is the uniform limit of the f_n's. Since

$$\begin{aligned} f \vee g &= \frac{1}{2}\,(f + g + |f - g|) \text{ and} \\ f \wedge g &= \frac{1}{2}\,(f + g - |f - g|), \end{aligned}$$

we have shown $\mathcal{A}$ is a lattice. ∎

Lemma 12.30. *Let $\mathcal{A} \subset C(X,\mathbb{R})$ be an algebra which separates points and suppose x and y are distinct points of X. If there exits such that $f, g \in \mathcal{A}$ such that*

$$f(x) \neq 0 \text{ and } g(y) \neq 0, \tag{12.5}$$

then

$$V := \{(f(x), f(y)) : f \in \mathcal{A}\} = \mathbb{R}^2. \tag{12.6}$$

Proof. It is clear that V is a non-zero subspace of $\mathbb{R}^2$. If $\dim(V) = 1$, then $V = \operatorname{span}(a, b)$ for some $(a, b) \in \mathbb{R}^2$ which, necessarily by Eq. (12.5), satisfy $a \neq 0 \neq b$. Since $(a, b) = (f(x), f(y))$ for some $f \in \mathcal{A}$ and $f^2 \in \mathcal{A}$, it follows that $(a^2, b^2) = (f^2(x), f^2(y)) \in V$ as well. Since $\dim V = 1$, (a, b) and (a^2, b^2) are linearly dependent and therefore

$$0 = \det \begin{pmatrix} a & b \\ a^2 & b^2 \end{pmatrix} = ab^2 - a^2 b = ab(b - a)$$

which implies that $a = b$. But this the implies that $f(x) = f(y)$ for all $f \in \mathcal{A}$, violating the assumption that $\mathcal{A}$ separates points. Therefore we conclude that $\dim(V) = 2$, i.e. $V = \mathbb{R}^2$. ∎

Theorem 12.31 (Stone-Weierstrass Theorem). *Suppose X is a locally compact Hausdorff space and $\mathcal{A} \subset C_0(X, \mathbb{R})$ is a **closed** subalgebra which separates points. For $x \in X$ let*

$$\mathcal{A}_x := \{f(x) : f \in \mathcal{A}\} \text{ and}$$
$$\mathcal{I}_x = \{f \in C_0(X, \mathbb{R}) : f(x) = 0\}.$$

Then either one of the following two cases hold.

1. *$\mathcal{A} = C_0(X, \mathbb{R})$ or*
2. *there exists a unique point $x_0 \in X$ such that $\mathcal{A} = \mathcal{I}_{x_0}$.*

Moreover, case 1. holds iff $\mathcal{A}_x = \mathbb{R}$ for all $x \in X$ and case 2. holds iff there exists a point $x_0 \in X$ such that $\mathcal{A}_{x_0} = \{0\}$.

Proof. If there exists x_0 such that $\mathcal{A}_{x_0} = \{0\}$ (x_0 is unique since $\mathcal{A}$ separates points) then $\mathcal{A} \subset \mathcal{I}_{x_0}$. If such an x_0 exists let $\mathcal{C} = \mathcal{I}_{x_0}$ and if $\mathcal{A}_x = \mathbb{R}$ for all x, set $\mathcal{C} = C_0(X, \mathbb{R})$. Let $f \in \mathcal{C}$ be given. By Lemma 12.30, for all $x, y \in X$ such that $x \neq y$, there exists $g_{xy} \in \mathcal{A}$ such that $f = g_{xy}$ on $\{x, y\}$.[3] When X is compact the basic idea of the proof is contained in the following identity,

$$f(z) = \inf_{x \in X} \sup_{y \in X} g_{xy}(z) \text{ for all } z \in X. \tag{12.7}$$

To prove this identity, let $g_x := \sup_{y \in X} g_{xy}$ and notice that $g_x \geq f$ since $g_{xy}(y) = f(y)$ for all $y \in X$. Moreover, $g_x(x) = f(x)$ for all $x \in X$ since $g_{xy}(x) = f(x)$ for all x. Therefore,

$$\inf_{x \in X} \sup_{y \in X} g_{xy} = \inf_{x \in X} g_x = f.$$

The rest of the proof is devoted to replacing the inf and the sup above by min and max over finite sets at the expense of Eq. (12.7) becoming only an approximate identity. We also have to modify Eq. (12.7) slightly to take care of the non-compact case.

[3] If $\mathcal{A}_{x_0} = \{0\}$ and $x = x_0$ or $y = x_0$, then g_{xy} exists merely by the fact that $\mathcal{A}$ separates points.

Claim. Given $\varepsilon > 0$ and $x \in X$ there exists $g_x \in \mathcal{A}$ such that $g_x(x) = f(x)$ and $f < g_x + \varepsilon$ on X.

To prove this, let V_y be an open neighborhood of y such that $|f - g_{xy}| < \varepsilon$ on V_y; in particular $f < \varepsilon + g_{xy}$ on V_y. Also let $g_{x,\infty}$ be any fixed element in $\mathcal{A}$ such that $g_{x,\infty}(x) = f(x)$ and let

$$K = \left\{|f| \geq \frac{\varepsilon}{2}\right\} \cup \left\{|g_{x,\infty}| \geq \frac{\varepsilon}{2}\right\}. \tag{12.8}$$

Since K is compact, there exists $\Lambda \subset\subset K$ such that $K \subset \bigcup_{y \in \Lambda} V_y$. Define

$$g_x(z) = \max\{g_{xy} : y \in \Lambda \cup \{\infty\}\}.$$

Since

$$f < \varepsilon + g_{xy} < \varepsilon + g_x \text{ on } V_y,$$

for any $y \in \Lambda$, and

$$f < \frac{\varepsilon}{2} < \varepsilon + g_{x,\infty} \leq g_x + \varepsilon \text{ on } K^c,$$

$f < \varepsilon + g_x$ on X and by construction $f(x) = g_x(x)$, see Figure **??**. This completes the proof of the claim.

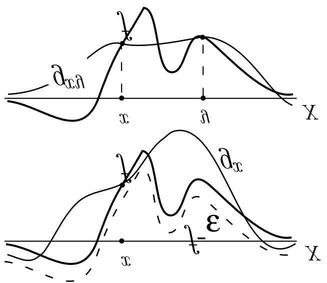


Fig. 12.6. Constructing the "dominating approximates," g_x for each $x \in X$.

To complete the proof of the theorem, let g_∞ be a fixed element of $\mathcal{A}$ such that $f < g_\infty + \varepsilon$ on X; for example let $g_\infty = g_{x_0} \in \mathcal{A}$ for some fixed $x_0 \in X$. For each $x \in X$, let U_x be a neighborhood of x such that $|f - g_x| < \varepsilon$ on U_x. Choose

$$\Gamma \subset\subset F := \left\{|f| \geq \frac{\varepsilon}{2}\right\} \cup \left\{|g_\infty| \geq \frac{\varepsilon}{2}\right\}$$

such that $F \subset \bigcup_{x \in \Gamma} U_x$ (Γ exists since F is compact) and define

$$g = \min\{g_x : x \in \Gamma \cup \{\infty\}\} \in \mathcal{A}.$$

Then, for $x \in F$, $g_x < f + \varepsilon$ on U_x and hence $g < f + \varepsilon$ on $\bigcup_{x \in \Gamma} U_x \supset F$. Likewise,

$$g \leq g_\infty < \varepsilon/2 < f + \varepsilon \text{ on } F^c.$$

Therefore we have now shown,

$$f < g + \varepsilon \text{ and } g < f + \varepsilon \text{ on } X,$$

i.e. $|f - g| < \varepsilon$ on X. Since $\varepsilon > 0$ is arbitrary it follows that $f \in \bar{\mathcal{A}} = \mathcal{A}$ and so $\mathcal{A} = \mathcal{C}$. ∎

Corollary 12.32 (Complex Stone-Weierstrass Theorem). *Let X be a locally compact Hausdorff space. Suppose $\mathcal{A} \subset C_0(X, \mathbb{C})$ is closed in the uniform topology, separates points, and is closed under complex conjugation. Then either $\mathcal{A} = C_0(X, \mathbb{C})$ or*

$$\mathcal{A} = \mathcal{I}_{x_0}^{\mathbb{C}} := \{f \in C_0(X, \mathbb{C}) : f(x_0) = 0\}$$

for some $x_0 \in X$.

Proof. Since

$$\text{Re } f = \frac{f + \bar{f}}{2} \text{ and } \operatorname{Im} f = \frac{f - \bar{f}}{2i},$$

Re f and $\operatorname{Im} f$ are both in $\mathcal{A}$. Therefore

$$\mathcal{A}_{\mathbb{R}} = \{\text{Re } f, \operatorname{Im} f : f \in \mathcal{A}\}$$

is a real sub-algebra of $C_0(X, \mathbb{R})$ which separates points. Therefore either $\mathcal{A}_{\mathbb{R}} = C_0(X, \mathbb{R})$ or $\mathcal{A}_{\mathbb{R}} = \mathcal{I}_{x_0} \cap C_0(X, \mathbb{R})$ for some x_0 and hence $\mathcal{A} = C_0(X, \mathbb{C})$ or $\mathcal{I}_{x_0}^{\mathbb{C}}$ respectively. ∎

As an easy application, Theorem 12.31 and Corollary 12.32 imply Theorem 8.34 and Corollary 8.36 respectively. Here are a few more applications.

Example 12.33. Let $f \in C([a, b])$ be a positive function which is injective. Then functions of the form $\sum_{k=1}^{N} a_k f^k$ with $a_k \in \mathbb{C}$ and $N \in \mathbb{N}$ are dense in $C([a, b])$. For example if $a = 1$ and $b = 2$, then one may take $f(x) = x^\alpha$ for any $\alpha \neq 0$, or $f(x) = e^x$, etc.

Exercise 12.4. Let (X, d) be a separable compact metric space. Show that $C(X)$ is also separable. **Hint:** Let $E \subset X$ be a countable dense set and then consider the algebra, $\mathcal{A} \subset C(X)$, generated by $\{d(x, \cdot)\}_{x \in E}$.

Example 12.34. Let $X = [0, \infty)$, $\lambda > 0$ be fixed, $\mathcal{A}$ be the real algebra generated by $t \to e^{-\lambda t}$. So the general element $f \in \mathcal{A}$ is of the form $f(t) = p(e^{-\lambda t})$, where $p(x)$ is a polynomial function in x with real coefficients. Since $\mathcal{A} \subset C_0(X, \mathbb{R})$ separates points and $e^{-\lambda t} \in \mathcal{A}$ is pointwise positive, $\bar{\mathcal{A}} = C_0(X, \mathbb{R})$.

As an application of Example 12.34, suppose that $g \in C_c(X, \mathbb{R})$ satisfies,

$$\int_0^\infty g(t) e^{-\lambda t} dt = 0 \text{ for all } \lambda > 0. \tag{12.9}$$

(Note well that the integral in Eq. (12.9) is really over a finite interval since g is compactly supported.) Equation (12.9) along with linearity of the Riemann integral implies

$$\int_0^\infty g(t) f(t) dt = 0 \text{ for all } f \in \mathcal{A}.$$

We may now choose $f_n \in \mathcal{A}$ such that $f_n \to g$ uniformly and therefore, using the continuity of the Riemann integral under uniform convergence (see Proposition 8.5),

$$0 = \lim_{n \to \infty} \int_0^\infty g(t) f_n(t) dt = \int_0^\infty g^2(t) dt.$$

From this last equation it is easily deduced, using the continuity of g, that $g \equiv 0$. See Theorem 22.12 below, where this is done in greater generality.

12.5 *More on Separation Axioms: Normal Spaces

(This section may safely be omitted on the first reading.)

Definition 12.35 ($T_0 - T_2$ Separation Axioms). *Let (X, τ) be a topological space. The topology τ is said to be:*

1. *T_0 if for $x \neq y$ in X there exists $V \in \tau$ such that $x \in V$ and $y \notin V$ or V such that $y \in V$ but $x \notin V$.*
2. *T_1 if for every $x, y \in X$ with $x \neq y$ there exists $V \in \tau$ such that $x \in V$ and $y \notin V$. Equivalently, τ is T_1 iff all one point subsets of X are closed.*[4]
3. *T_2 if it is Hausdorff.*

Note T_2 implies T_1 which implies T_0. The topology in Example 12.1 is T_0 but not T_1. If X is a finite set and τ is a T_1 – topology on X then $\tau = 2^X$. To prove this let $x \in X$ be fixed. Then for every $y \neq x$ in X there exists $V_y \in \tau$ such that $x \in V_y$ while $y \notin V_y$. Thus $\{x\} = \cap_{y \neq x} V_y \in \tau$ showing τ contains all one point subsets of X and therefore all subsets of X. So we have to look to infinite sets for an example of T_1 topology which is not T_2.

[4] If one point subsets are closed and $x \neq y$ in X then $V := \{x\}^c$ is an open set containing y but not x. Conversely if τ is T_1 and $x \in X$ there exists $V_y \in \tau$ such that $y \in V_y$ and $x \notin V_y$ for all $y \neq x$. Therefore, $\{x\}^c = \cup_{y \neq x} V_y \in \tau$.

Example 12.36. Let X be any infinite set and let $\tau = \{A \subset X : \#(A^c) < \infty\} \cup \{\emptyset\}$ – the so called **cofinite** topology. This topology is T_1 because if $x \neq y$ in X, then $V = \{x\}^c \in \tau$ with $x \notin V$ while $y \in V$. This topology however is not T_2. Indeed if $U, V \in \tau$ are open sets such that $x \in U$, $y \in V$ and $U \cap V = \emptyset$ then $U \subset V^c$. But this implies $\#(U) < \infty$ which is impossible unless $U = \emptyset$ which is impossible since $x \in U$.

The uniqueness of limits of sequences which occurs for Hausdorff topologies (see Remark 12.3) need not occur for T_1 – spaces. For example, let $X = \mathbb{N}$ and τ be the cofinite topology on X as in Example 12.36. Then $x_n = n$ is a sequence in X such that $x_n \to x$ as $n \to \infty$ **for all** $x \in \mathbb{N}$. For the most part we will avoid these pathologies in the future by only considering Hausdorff topologies.

Definition 12.37 (Normal Spaces: T_4 – Separation Axiom). *A topological space (X, τ) is said to be **normal** or T_4 if:*

1. *X is Hausdorff and*
2. *if for any two closed disjoint subsets $A, B \subset X$ there exists disjoint open sets $V, W \subset X$ such that $A \subset V$ and $B \subset W$.*

Example 12.38. By Lemma 6.15 and Corollary 12.21 it follows that metric spaces and topological spaces which are locally compact, σ – compact and Hausdorff (in particular compact Hausdorff spaces) are normal. Indeed, in each case if A, B are disjoint closed subsets of X, there exists $f \in C(X, [0,1])$ such that $f = 1$ on A and $f = 0$ on B. Now let $U = \left\{f > \frac{1}{2}\right\}$ and $V = \{f < \frac{1}{2}\}$.

Remark 12.39. A topological space, (X, τ), is normal iff for any $C \subset W \subset X$ with C being closed and W being open there exists an open set $U \subset_o X$ such that

$$C \subset U \subset \bar{U} \subset W.$$

To prove this first suppose X is normal. Since W^c is closed and $C \cap W^c = \emptyset$, there exists disjoint open sets U and V such that $C \subset U$ and $W^c \subset V$. Therefore $C \subset U \subset V^c \subset W$ and since V^c is closed, $C \subset U \subset \bar{U} \subset V^c \subset W$.

For the converse direction suppose A and B are disjoint closed subsets of X. Then $A \subset B^c$ and B^c is open, and so by assumption there exists $U \subset_o X$ such that $A \subset U \subset \bar{U} \subset B^c$ and by the same token there exists $W \subset_o X$ such that $\bar{U} \subset W \subset \bar{W} \subset B^c$. Taking complements of the last expression implies

$$B \subset \bar{W}^c \subset W^c \subset \bar{U}^c.$$

Let $V = \bar{W}^c$. Then $A \subset U \subset_o X$, $B \subset V \subset_o X$ and $U \cap V \subset U \cap W^c = \emptyset$.

Theorem 12.40 (Urysohn's Lemma for Normal Spaces). *Let X be a normal space. Assume A, B are disjoint closed subsets of X. Then there exists $f \in C(X, [0,1])$ such that $f = 0$ on A and $f = 1$ on B.*

Proof. To make the notation match Lemma 12.8, let $U = A^c$ and $K = B$. Then $K \subset U$ and it suffices to produce a function $f \in C(X, [0,1])$ such that $f = 1$ on K and $\mathrm{supp}(f) \subset U$. The proof is now identical to that for Lemma 12.8 except we now use Remark 12.39 in place of Proposition 12.7. ∎

Theorem 12.41 (Tietze Extension Theorem). *Let (X, τ) be a normal space, D be a closed subset of X, $-\infty < a < b < \infty$ and $f \in C(D, [a,b])$. Then there exists $F \in C(X, [a,b])$ such that $F|_D = f$.*

Proof. The proof is identical to that of Theorem 7.4 except we now use Theorem 12.40 in place of Lemma 6.15. ∎

Corollary 12.42. *Suppose that X is a normal topological space, $D \subset X$ is closed, $F \in C(D, \mathbb{R})$. Then there exists $F \in C(X)$ such that $F|_D = f$.*

Proof. Let $g = \arctan(f) \in C(D, (-\frac{\pi}{2}, \frac{\pi}{2}))$. Then by the Tietze extension theorem, there exists $G \in C(X, [-\frac{\pi}{2}, \frac{\pi}{2}])$ such that $G|_D = g$. Let $B := G^{-1}(\{-\frac{\pi}{2}, \frac{\pi}{2}\}) \sqsubset X$, then $B \cap D = \emptyset$. By Urysohn's lemma (Theorem 12.40) there exists $h \in C(X, [0,1])$ such that $h \equiv 1$ on D and $h = 0$ on B and in particular $hG \in C(D, (-\frac{\pi}{2}, \frac{\pi}{2}))$ and $(hG)|_D = g$. The function $F := \tan(hG) \in C(X)$ is an extension of f. ∎

Theorem 12.43 (Urysohn Metrization Theorem for Normal Spaces). *Every second countable normal space, (X, τ), is metrizable, i.e. there is a metric ρ on X such that $\tau = \tau_\rho$. Moreover, ρ may be chosen so that X is isometric to a subset $Q_0 \subset Q$ (Q is as in Notation 12.10) equipped with the metric d in Eq. (12.2). In this metric X is totally bounded and hence the completion of X (which is isometric to $\bar{Q}_0 \subset Q$) is compact.*

Proof. (The proof here will be very similar to the proof of Theorem 12.13.) Let $\mathcal{B}$ be a countable base for τ and set

$$\Gamma := \{(U, V) \in \mathcal{B} \times \mathcal{B} \mid \bar{U} \subset V\}.$$

To each $O \in \tau$ and $x \in O$ there exist $(U, V) \in \Gamma$ such that $x \in U \subset V \subset O$. Indeed, since $\mathcal{B}$ is a base for τ, there exists $V \in \mathcal{B}$ such that $x \in V \subset O$. Because $\{x\} \cap V^c = \emptyset$, there exists disjoint open sets $\widetilde{U}$ and W such that $x \in \widetilde{U}$, $V^c \subset W$ and $\widetilde{U} \cap W = \emptyset$. Choose $U \in \mathcal{B}$ such that $x \in U \subset \widetilde{U}$. Since $U \subset \widetilde{U} \subset W^c$, $\overline{U} \subset W^c \subset V$ and hence $(U, V) \in \Gamma$. See Figure 12.7 below. In particular this shows that

$$\mathcal{B}_0 := \{U \in \mathcal{B} : (U, V) \in \Gamma \text{ for some } V \in \mathcal{B}\}$$

is still a base for τ.

If Γ is a finite set, the previous comment shows that τ only has a finite number of elements as well. Since (X, τ) is Hausdorff, it follows that X is a finite set. Letting $\{x_n\}_{n=1}^N$ be an enumeration of X, define $T : X \to Q$ by

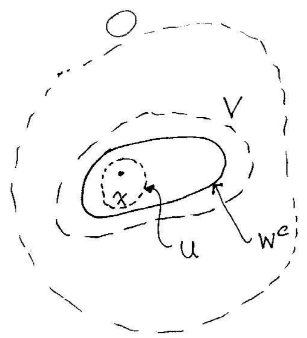


Fig. 12.7. Constructing $(U, V) \in \Gamma$.

$T(x_n) = e_n$ for $n = 1, 2, \dots, N$ where $e_n = (0, 0, \dots, 0, 1, 0, \dots)$, with the 1 occurring in the n^{th} spot. Then $\rho(x, y) := d(T(x), T(y))$ for $x, y \in X$ is the desired metric.

So we may now assume that Γ is an infinite set and let $\{(U_n, V_n)\}_{n=1}^\infty$ be an enumeration of Γ. By Urysohn's Lemma for normal spaces (Theorem 12.40) there exists $f_{U,V} \in C(X, [0,1])$ such that $f_{U,V} = 0$ on $\bar{U}$ and $f_{U,V} = 1$ on V^c. Let $\mathcal{F} := \{f_{U,V} \mid (U, V) \in \Gamma\}$ and set $f_n := f_{U_n, V_n}$ – an enumeration of $\mathcal{F}$. The proof that

$$\rho(x, y) := \sum_{n=1}^\infty \frac{1}{2^n} |f_n(x) - f_n(y)|$$

is the desired metric on X now follows exactly as the corresponding argument in the proof of Theorem 12.13. ■

12.6 Exercises

Exercise 12.5. Prove Theorem 12.9. **Hints:**

1. By Proposition 12.7, there exists a precompact open set V such that $K \subset V \subset \bar{V} \subset U$. Now suppose that $f : K \to [0, \alpha]$ is continuous with $\alpha \in (0, 1]$ and let $A := f^{-1}([0, \frac{1}{3}\alpha])$ and $B := f^{-1}([\frac{2}{3}\alpha, 1])$. Appeal to Lemma 12.8 to find a function $g \in C(X, [0, \alpha/3])$ such that $g = \alpha/3$ on B and $\text{supp}(g) \subset V \setminus A$.
2. Now follow the argument in the proof of Theorem 7.4 to construct $F \in C(X, [a, b])$ such that $F|_K = f$.
3. For $c \in [a, b]$, choose $\phi \prec U$ such that $\phi = 1$ on K and replace F by $F_c := \phi F + (1 - \phi)c$.

Exercise 12.6 (Sterographic Projection). Let $X = \mathbb{R}^n$, $X^* := X \cup \{\infty\}$ be the one point compactification of X, $S^n := \{y \in \mathbb{R}^{n+1} : |y| = 1\}$ be the unit sphere in $\mathbb{R}^{n+1}$ and $N = (0, \ldots, 0, 1) \in \mathbb{R}^{n+1}$. Define $f : S^n \to X^*$ by $f(N) = \infty$, and for $y \in S^n \setminus \{N\}$ let $f(y) = b \in \mathbb{R}^n$ be the unique point such that $(b, 0)$ is on the line containing N and y, see Figure 12.8 below. Find a formula for f and show $f : S^n \to X^*$ is a homeomorphism. (So the one point compactification of $\mathbb{R}^n$ is homeomorphic to the n sphere.)

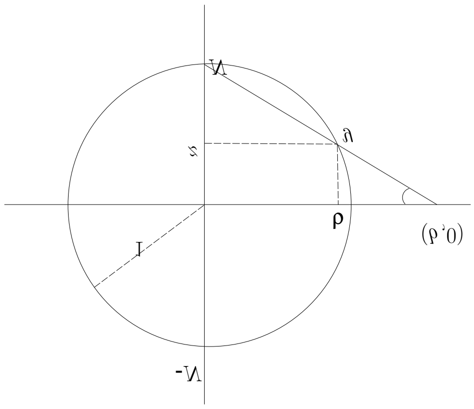


Fig. 12.8. Sterographic projection and the one point compactification of $\mathbb{R}^n$.

Exercise 12.7. Let (X, τ) be a locally compact Hausdorff space. Show (X, τ) is separable iff (X^*, τ^*) is separable.

Exercise 12.8. Show by example that there exists a locally compact metric space (X, d) such that the one point compactification, $(X^* := X \cup \{\infty\}, \tau^*)$, is **not** metrizable. **Hint**: use exercise 12.7.

Exercise 12.9. Suppose (X, d) is a locally compact and σ − compact metric space. Show the one point compactification, $(X^* := X \cup \{\infty\}, \tau^*)$, is metrizable.

Exercise 12.10. In this problem, suppose Theorem 12.31 has only been proved when X is compact. Show that it is possible to prove Theorem 12.31 by using Proposition 12.24 to reduce the non-compact case to the compact case.

Hints.

Exercise 12.11. 1. If $\mathcal{A}_x = \mathbb{R}$ for all $x \in X$ let $X^* = X \cup \{\infty\}$ be the one point compactification of X.
2. If $\mathcal{A}_{x_0} = \{0\}$ for some $x_0 \in X$, let $Y := X \setminus \{x_0\}$ and $Y^* = Y \cup \{\infty\}$ be the one point compactification of Y.

For $f \in \mathcal{A}$ define $f(\infty) = 0$. In this way $\mathcal{A}$ may be considered to be a sub-algebra of $C(X^*, \mathbb{R})$ in case 1. or a sub-algebra of $C(Y^*, \mathbb{R})$ in case 2.

Exercise 12.12. Let $M < \infty$, show there are polynomials $p_n(t)$ such that

$$\lim_{n\to\infty} \sup_{|t|\le M} ||t| - p_n(t)| = 0$$

using the following outline.

1. Let $f(x) = \sqrt{1-x}$ for $|x| \le 1$ and use Taylor's theorem with integral remainder (see Eq. A.15 of Appendix A), or analytic function theory if you know it, to show there are constants[5] $c_n > 0$ for $n \in \mathbb{N}$ such that

$$\sqrt{1-x} = 1 - \sum_{n=1}^{\infty} c_n x^n \text{ for all } |x| < 1. \tag{12.10}$$

2. Let $q_m(x) := 1 - \sum_{n=1}^{m} c_n x^n$. Use (12.10) to show $\sum_{n=1}^{\infty} c_n = 1$ and conclude from this that

$$\lim_{m\to\infty} \sup_{|x|\le 1} |\sqrt{1-x} - q_m(x)| = 0. \tag{12.11}$$

3. Let $1 - x = t^2/M^2$, i.e. $x = 1 - t^2/M^2$, then

$$\lim_{m\to\infty} \sup_{|t|\le M} \left| \frac{|t|}{M} - q_m(1 - t^2/M^2) \right| = 0$$

   so that $p_m(t) := M q_m(1 - t^2/M^2)$ are the desired polynomials.

Exercise 12.13. Given a continuous function $f : \mathbb{R} \to \mathbb{C}$ which is 2π - periodic and $\varepsilon > 0$. Show there exists a trigonometric polynomial, $p(\theta) = \sum_{n=-N}^{n} \alpha_n e^{in\theta}$, such that $|f(\theta) - P(\theta)| < \varepsilon$ for all $\theta \in \mathbb{R}$. **Hint:** show that there exists a unique function $F \in C(S^1)$ such that $f(\theta) = F(e^{i\theta})$ for all $\theta \in \mathbb{R}$.

Remark 12.44. Exercise 12.13 generalizes to 2π – periodic functions on $\mathbb{R}^d$, i.e. functions such that $f(\theta + 2\pi e_i) = f(\theta)$ for all $i = 1, 2, \ldots, d$ where $\{e_i\}_{i=1}^d$ is the standard basis for $\mathbb{R}^d$. A trigonometric polynomial $p(\theta)$ is a function of $\theta \in \mathbb{R}^d$ of the form

$$p(\theta) = \sum_{n\in\Gamma} \alpha_n e^{in\cdot\theta}$$

where Γ is a finite subset of $\mathbb{Z}^d$. The assertion is again that these trigonometric polynomials are dense in the 2π – periodic functions relative to the supremum norm.

[5] In fact $c_n := \frac{(2n-3)!!}{2^n n!}$, but this is not needed.

13

Baire Category Theorem

Definition 13.1. *Let (X, τ) be a topological space. A set $E \subset X$ is said to be* ***nowhere dense*** *if $\left(\bar{E}\right)^o = \emptyset$ i.e. $\bar{E}$ has empty interior.*

Notice that E is nowhere dense is equivalent to

$$X = \left(\left(\bar{E}\right)^o\right)^c = \overline{\left(\bar{E}\right)^c} = \overline{(E^c)^o}.$$

That is to say E is nowhere dense iff E^c has dense interior.

13.1 Metric Space Baire Category Theorem

Theorem 13.2 (Baire Category Theorem). *Let (X, ρ) be a complete metric space.*

1. *If $\{V_n\}_{n=1}^\infty$ is a sequence of dense open sets, then $G := \bigcap_{n=1}^{\infty} V_n$ is dense in X.*
2. *If $\{E_n\}_{n=1}^\infty$ is a sequence of nowhere dense sets, then $\bigcup_{n=1}^\infty E_n \subset \bigcup_{n=1}^\infty \bar{E}_n \subsetneq X$ and in particular $X \neq \bigcup_{n=1}^\infty E_n$.*

Proof. 1) We must shows that $\bar{G} = X$ which is equivalent to showing that $W \cap G \neq \emptyset$ for all non-empty open sets $W \subset X$. Since V_1 is dense, $W \cap V_1 \neq \emptyset$ and hence there exists $x_1 \in X$ and $\varepsilon_1 > 0$ such that

$$\overline{B(x_1, \varepsilon_1)} \subset W \cap V_1.$$

Since V_2 is dense, $B(x_1, \varepsilon_1) \cap V_2 \neq \emptyset$ and hence there exists $x_2 \in X$ and $\varepsilon_2 > 0$ such that

$$\overline{B(x_2, \varepsilon_2)} \subset B(x_1, \varepsilon_1) \cap V_2.$$

Continuing this way inductively, we may choose $\{x_n \in X \text{ and } \varepsilon_n > 0\}_{n=1}^\infty$ such that

$$\overline{B(x_n,\varepsilon_n)} \subset B(x_{n-1},\varepsilon_{n-1}) \cap V_n \ \forall n.$$

Furthermore we can clearly do this construction in such a way that $\varepsilon_n \downarrow 0$ as $n \uparrow \infty$. Hence $\{x_n\}_{n=1}^\infty$ is Cauchy sequence and $x = \lim_{n\to\infty} x_n$ exists in X since X is complete. Since $\overline{B(x_n,\varepsilon_n)}$ is closed, $x \in \overline{B(x_n,\varepsilon_n)} \subset V_n$ so that $x \in V_n$ for all n and hence $x \in G$. Moreover, $x \in \overline{B(x_1,\varepsilon_1)} \subset W \cap V_1$ implies $x \in W$ and hence $x \in W \cap G$ showing $W \cap G \neq \emptyset$. 2) The second assertion is equivalently to showing

$$\emptyset \neq \left(\bigcup_{n=1}^\infty \bar{E}_n\right)^c = \bigcap_{n=1}^\infty \left(\bar{E}_n\right)^c = \bigcap_{n=1}^\infty \left(E_n^c\right)^o.$$

As we have observed, E_n is nowhere dense is equivalent to $(E_n^c)^o$ being a dense open set, hence by part 1), $\bigcap_{n=1}^\infty (E_n^c)^o$ is dense in X and hence not empty. ∎

13.2 Locally Compact Hausdorff Space Baire Category Theorem

Here is another version of the Baire Category theorem when X is a locally compact Hausdorff space.

Proposition 13.3. *Let X be a locally compact Hausdorff space.*

1. *If $\{V_n\}_{n=1}^\infty$ is a sequence of dense open sets, then $G := \bigcap_{n=1}^\infty V_n$ is dense in X.*
2. *If $\{E_n\}_{n=1}^\infty$ is a sequence of nowhere dense sets, then $X \neq \bigcup_{n=1}^\infty E_n$.*

Proof. As in the previous proof, the second assertion is a consequence of the first. To finish the proof, if suffices to show $G \cap W \neq \emptyset$ for all open sets $W \subset X$. Since V_1 is dense, there exists $x_1 \in V_1 \cap W$ and by Proposition 12.7 there exists $U_1 \subset_o X$ such that $x_1 \in U_1 \subset \bar{U}_1 \subset V_1 \cap W$ with $\bar{U}_1$ being compact. Similarly, there exists a non-empty open set U_2 such that $U_2 \subset \bar{U}_2 \subset U_1 \cap V_2$. Working inductively, we may find non-empty open sets $\{U_k\}_{k=1}^\infty$ such that $U_k \subset \bar{U}_k \subset U_{k-1} \cap V_k$. Since $\cap_{k=1}^n \bar{U}_k = \bar{U}_n \neq \emptyset$ for all n, the finite intersection characterization of $\bar{U}_1$ being compact implies that

$$\emptyset \neq \cap_{k=1}^\infty \bar{U}_k \subset G \cap W.$$

∎

Definition 13.4. *A subset $E \subset X$ is **meager** or of the **first category** if $E = \bigcup_{n=1}^\infty E_n$ where each E_n is nowhere dense. And a set $R \subset X$ is called **residual** if R^c is meager.*

Remarks 13.5 *For those readers that already know some measure theory may want to think of meager as being the topological analogue of sets of measure* 0 *and residual as being the topological analogue of sets of full measure. (This analogy should not be taken too seriously, see Exercise 19.19.)*

1. *R is residual iff R contains a countable intersection of dense open sets. Indeed if R is a residual set, then there exists nowhere dense sets $\{E_n\}$ such that*
$$R^c = \cup_{n=1}^\infty E_n \subset \cup_{n=1}^\infty \bar{E}_n.$$
Taking complements of this equation shows that
$$\cap_{n=1}^\infty \bar{E}_n^c \subset R,$$
i.e. R contains a set of the form $\cap_{n=1}^\infty V_n$ with each V_n $(= \bar{E}_n^c)$ being an open dense subset of X.
Conversely, if $\cap_{n=1}^\infty V_n \subset R$ with each V_n being an open dense subset of X, then $R^c \subset \cup_{n=1}^\infty V_n^c$ and hence $R^c = \cup_{n=1}^\infty E_n$ where each $E_n = R^c \cap V_n^c$, is a nowhere dense subset of X.
2. *A countable union of meager sets is meager and any subset of a meager set is meager.*
3. *A countable intersection of residual sets is residual.*

Remarks 13.6 *The Baire Category Theorems may now be stated as follows. If X is a complete metric space or X is a locally compact Hausdorff space, then*

1. *all residual sets are dense in X and*
2. *X is not meager.*

It should also be remarked that incomplete metric spaces may be meager. For example, let $X \subset C([0,1])$ be the subspace of polynomial functions on $[0,1]$ equipped with the supremum norm. Then $X = \cup_{n=1}^\infty E_n$ where $E_n \subset X$ denotes the subspace of polynomials of degree less than or equal to n. You are asked to show in Exercise 13.1 below that E_n is nowhere dense for all n. Hence X is meager and the empty set is residual in X.

Here is an application of Theorem 13.2.

Theorem 13.7. *Let $\mathcal{N} \subset C([0,1],\mathbb{R})$ be the set of nowhere differentiable functions. (Here a function f is said to be differentiable at* 0 *if $f'(0) := \lim_{t\downarrow 0} \frac{f(t)-f(0)}{t}$ exists and at* 1 *if $f'(1) := \lim_{t\uparrow 0} \frac{f(1)-f(t)}{1-t}$ exists.) Then $\mathcal{N}$ is a residual set so the "generic" continuous functions is nowhere differentiable.*

Proof. If $f \notin \mathcal{N}$, then $f'(x_0)$ exists for some $x_0 \in [0,1]$ and by the definition of the derivative and compactness of $[0,1]$, there exists $n \in \mathbb{N}$ such that $|f(x) - f(x_0)| \le n|x - x_0|\ \forall\ x \in [0,1]$. Thus if we define

$$E_n := \{f \in C([0,1]) : \exists\ x_0 \in [0,1]\ \ni\ |f(x) - f(x_0)| \le n|x - x_0|\ \forall\ x \in [0,1]\},$$

then we have just shown $\mathcal{N}^c \subset E := \cup_{n=1}^{\infty} E_n$. So to finish the proof it suffices to show (for each n) E_n is a closed subset of $C([0,1],\mathbb{R})$ with empty interior.
1) To prove E_n is closed, let $\{f_m\}_{m=1}^{\infty} \subset E_n$ be a sequence of functions such that there exists $f \in C([0,1],\mathbb{R})$ such that $\|f - f_m\|_\infty \to 0$ as $m \to \infty$. Since $f_m \in E_n$, there exists $x_m \in [0,1]$ such that

$$|f_m(x) - f_m(x_m)| \le n|x - x_m| \ \forall\ x \in [0,1]. \tag{13.1}$$

Since $[0,1]$ is a compact metric space, by passing to a subsequence if necessary, we may assume $x_0 = \lim_{m\to\infty} x_m \in [0,1]$ exists. Passing to the limit in Eq. (13.1), making use of the uniform convergence of $f_n \to f$ to show $\lim_{m\to\infty} f_m(x_m) = f(x_0)$, implies

$$|f(x) - f(x_0)| \le n|x - x_0| \ \forall\ x \in [0,1]$$

and therefore that $f \in E_n$. This shows E_n is a closed subset of $C([0,1],\mathbb{R})$.
2) To finish the proof, we will show $E_n^0 = \emptyset$ by showing for each $f \in E_n$ and $\varepsilon > 0$ given, there exists $g \in C([0,1],\mathbb{R}) \setminus E_n$ such that $\|f - g\|_\infty < \varepsilon$. We now construct g. Since $[0,1]$ is compact and f is continuous there exists $N \in \mathbb{N}$ such that $|f(x) - f(y)| < \varepsilon/2$ whenever $|y - x| < 1/N$. Let k denote the piecewise linear function on $[0,1]$ such that $k(\frac{m}{N}) = f(\frac{m}{N})$ for $m = 0, 1, \dots, N$ and $k''(x) = 0$ for $x \notin \pi_N := \{m/N : m = 0, 1, \dots, N\}$. Then it is easily seen that $\|f - k\|_u < \varepsilon/2$ and for $x \in (\frac{m}{N}, \frac{m+1}{N})$ that

$$|k'(x)| = \frac{|f(\frac{m+1}{N}) - f(\frac{m}{N})|}{\frac{1}{N}} < N\varepsilon/2.$$

We now make k "rougher" by adding a small wiggly function h which we define as follows. Let $M \in \mathbb{N}$ be chosen so that $4\varepsilon M > 2n$ and define h uniquely by $h(\frac{m}{M}) = (-1)^m \varepsilon/2$ for $m = 0, 1, \dots, M$ and $h''(x) = 0$ for $x \notin \pi_M$. Then $\|h\|_\infty < \varepsilon$ and $|h'(x)| = 4\varepsilon M > 2n$ for $x \notin \pi_M$. See Figure 13.1 below. Finally define $g := k + h$. Then

$$\|f - g\|_\infty \le \|f - k\|_\infty + \|h\|_\infty < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

and

$$|g'(x)| \ge |h'(x)| - |k'(x)| > 2n - n = n \ \forall x \notin \pi_M \cup \pi_N.$$

It now follows from this last equation and the mean value theorem that for any $x_0 \in [0,1]$,

$$\left| \frac{g(x) - g(x_0)}{x - x_0} \right| > n$$

for all $x \in [0,1]$ sufficiently close to x_0. This shows $g \notin E_n$ and so the proof is complete. ∎

Here is an application of the Baire Category Theorem in Proposition 13.3.

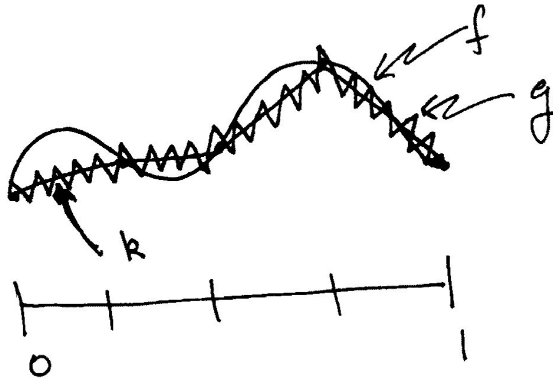


Fig. 13.1. Constgructing a rough approximation, g, to a continuous function f.

Proposition 13.8. *Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a function such that $f'(x)$ exists for all $x \in \mathbb{R}$. Let*

$$U := \cup_{\varepsilon>0} \left\{ x \in \mathbb{R} : \sup_{|y|<\varepsilon} |f'(x+y)| < \infty \right\}.$$

*Then U is a dense open set. (It is not true that $U = \mathbb{R}$ in general, see Example **??** below.)*

Proof. It is easily seen from the definition of U that U is open. Let $W \subset_o \mathbb{R}$ be an open subset of $\mathbb{R}$. For $k \in \mathbb{N}$, let

$$\begin{aligned} E_k &:= \left\{ x \in W : |f(y) - f(x)| \le k\,|y-x| \text{ when } |y-x| \le \frac{1}{k} \right\} \\ &= \bigcap_{z:|z|\le k^{-1}} \left\{ x \in W : |f(x+z) - f(x)| \le k\,|z| \right\}, \end{aligned}$$

which is a closed subset of $\mathbb{R}$ since f is continuous. Moreover, if $x \in W$ and $M = |f'(x)|$, then

$$\begin{aligned} |f(y) - f(x)| &= |f'(x)\,(y-x) + o\,(y-x)| \\ &\le (M+1)\,|y-x| \end{aligned}$$

for y close to x. (Here $o(y-x)$ denotes a function such that $\lim_{y\to x} o(y-x)/(y-x) = 0$.) In particular, this shows that $x \in E_k$ for all k sufficiently large. Therefore $W = \cup_{k=1}^{\infty} E_k$ and since W is not meager by the Baire category Theorem in Proposition 13.3, some E_k has non-empty interior. That is there exists $x_0 \in E_k \subset W$ and $\varepsilon > 0$ such that

$$J := (x_0 - \varepsilon, x_0 + \varepsilon) \subset E_k \subset W.$$

For $x \in J$, we have $|f(x+z) - f(x)| \le k\,|z|$ provided that $|z| \le k^{-1}$ and therefore that $|f'(x)| \le k$ for $x \in J$. Therefore $x_0 \in U \cap W$ showing U is dense. ■

Remark 13.9. This proposition generalizes to functions $f : \mathbb{R}^n \to \mathbb{R}^m$ in an obvious way.

For our next application of Theorem 13.2, let $X := BC^\infty((-1,1))$ denote the set of smooth functions f on $(-1,1)$ such that f and all of its derivatives are bounded. In the metric

$$\rho(f,g) := \sum_{k=0}^{\infty} 2^{-k} \frac{\left\|f^{(k)} - g^{(k)}\right\|_\infty}{1 + \left\|f^{(k)} - g^{(k)}\right\|_\infty} \text{ for } f, g \in X,$$

X becomes a complete metric space.

Theorem 13.10. *Given an increasing sequence of positive numbers $\{M_n\}_{n=1}^\infty$, the set*

$$\mathcal{F} := \left\{ f \in X : \limsup_{n\to\infty} \left| \frac{f^{(n)}(0)}{M_n} \right| \ge 1 \right\}$$

is dense in X. In particular, there is a dense set of $f \in X$ such that the power series expansion of f at 0 has zero radius of convergence.

Proof. Step 1. Let $n \in \mathbb{N}$. Choose $g \in C_c^\infty((-1,1))$ such that $\|g\|_\infty < 2^{-n}$ while $g'(0) = 2M_n$ and define

$$f_n(x) := \int_0^x dt_{n-1} \int_0^{t_{n-1}} dt_{n-2} \dots \int_0^{t_2} dt_1 g(t_1).$$

Then for $k < n$,

$$f_n^{(k)}(x) = \int_0^x dt_{n-k-1} \int_0^{t_{n-k-1}} dt_{n-k-2} \dots \int_0^{t_2} dt_1 g(t_1),$$

$f^{(n)}(x) = g'(x)$, $f_n^{(n)}(0) = 2M_n$ and $f_n^{(k)}$ satisfies

$$\left\|f_n^{(k)}\right\|_\infty \le \frac{2^{-n}}{(n-1-k)!} \le 2^{-n} \text{ for } k < n.$$

Consequently,

$$\begin{aligned} \rho(f_n, 0) &= \sum_{k=0}^{\infty} 2^{-k} \frac{\left\|f_n^{(k)}\right\|_\infty}{1 + \left\|f_n^{(k)}\right\|_\infty} \\ &\le \sum_{k=0}^{n-1} 2^{-k} 2^{-n} + \sum_{k=n}^{\infty} 2^{-k} \cdot 1 \le 2\left(2^{-n} + 2^{-n}\right) = 4 \cdot 2^{-n}. \end{aligned}$$

Thus we have constructed $f_n \in X$ such that $\lim_{n\to\infty} \rho(f_n, 0) = 0$ while $f_n^{(n)}(0) = 2M_n$ for all n. Step 2. The set

$$G_n := \cup_{m\geq n} \left\{ f \in X : \left| f^{(m)}(0) \right| > M_m \right\}$$

is a dense open subset of X. The fact that G_n is open is clear. To see that G_n is dense, let $g \in X$ be given and define $g_m := g + \varepsilon_m f_m$ where $\varepsilon_m := sgn(g^{(m)}(0))$. Then

$$\left| g_m^{(m)}(0) \right| = \left| g^{(m)}(0) \right| + \left| f_m^{(m)}(0) \right| \geq 2M_m > M_m \text{ for all } m.$$

Therefore, $g_m \in G_n$ for all $m \geq n$ and since

$$\rho(g_m, g) = \rho(f_m, 0) \to 0 \text{ as } m \to \infty$$

it follows that $g \in \bar{G}_n$. Step 3. By the Baire Category theorem, $\cap G_n$ is a dense subset of X. This completes the proof of the first assertion since

$$\begin{aligned} \mathcal{F} &= \left\{ f \in X : \limsup_{n\to\infty} \left| \frac{f^{(n)}(0)}{M_n} \right| \geq 1 \right\} \\ &= \cap_{n=1}^{\infty} \left\{ f \in X : \left| \frac{f^{(n)}(0)}{M_n} \right| \geq 1 \text{ for some } n \geq m \right\} \supset \cap_{n=1}^{\infty} G_n. \end{aligned}$$

Step 4. Take $M_n = (n!)^2$ and recall that the power series expansion for f near 0 is given by $\sum_{n=0}^{\infty} \frac{f_n(0)}{n!} x^n$. This series can not converge for any $f \in \mathcal{F}$ and any $x \neq 0$ because

$$\begin{aligned} \limsup_{n\to\infty} \left| \frac{f_n(0)}{n!} x^n \right| &= \limsup_{n\to\infty} \left| \frac{f_n(0)}{(n!)^2} n! x^n \right| \\ &= \limsup_{n\to\infty} \left| \frac{f_n(0)}{(n!)^2} \right| \cdot \lim_{n\to\infty} n! \left| x^n \right| = \infty \end{aligned}$$

where we have used $\lim_{n\to\infty} n! \left| x^n \right| = \infty$ and $\limsup_{n\to\infty} \left| \frac{f_n(0)}{(n!)^2} \right| \geq 1$. ∎

Remark 13.11. Given a sequence of real number $\{a_n\}_{n=0}^{\infty}$ there always exists $f \in X$ such that $f^{(n)}(0) = a_n$. To construct such a function f, let $\phi \in C_c^{\infty}(-1, 1)$ be a function such that $\phi = 1$ in a neighborhood of 0 and $\varepsilon_n \in (0, 1)$ be chosen so that $\varepsilon_n \downarrow 0$ as $n \to \infty$ and $\sum_{n=0}^{\infty} |a_n| \varepsilon_n^n < \infty$. The desired function f can then be defined by

$$f(x) = \sum_{n=0}^{\infty} \frac{a_n}{n!} x^n \phi(x/\varepsilon_n) =: \sum_{n=0}^{\infty} g_n(x). \tag{13.2}$$

The fact that f is well defined and continuous follows from the estimate:

$$|g_n(x)| = \left|\frac{a_n}{n!} x^n \phi(x/\varepsilon_n)\right| \le \frac{\|\phi\|_\infty}{n!} |a_n| \, \varepsilon_n^n$$

and the assumption that $\sum_{n=0}^\infty |a_n| \, \varepsilon_n^n < \infty$. The estimate

$$\begin{aligned}
|g_n'(x)| &= \left|\frac{a_n}{(n-1)!} x^{n-1} \phi(x/\varepsilon_n) + \frac{a_n}{n!\varepsilon_n} x^n \phi'(x/\varepsilon_n)\right| \\
&\le \frac{\|\phi\|_\infty}{(n-1)!} |a_n| \, \varepsilon_n^{n-1} + \frac{\|\phi'\|_\infty}{n!} |a_n| \, \varepsilon_n^n \\
&\le \left(\|\phi\|_\infty + \|\phi'\|_\infty\right) |a_n| \, \varepsilon_n^n
\end{aligned}$$

and the assumption that $\sum_{n=0}^\infty |a_n| \, \varepsilon_n^n < \infty$ shows $f \in C^1(-1,1)$ and $f'(x) = \sum_{n=0}^\infty g_n'(x)$. Similar arguments show $f \in C_c^k(-1,1)$ and $f^{(k)}(x) = \sum_{n=0}^\infty g_n^{(k)}(x)$ for all x and $k \in \mathbb{N}$. This completes the proof since, using $\phi(x/\varepsilon_n) = 1$ for x in a neighborhood of 0, $g_n^{(k)}(0) = \delta_{k,n} a_k$ and hence

$$f^{(k)}(0) = \sum_{n=0}^\infty g_n^{(k)}(0) = a_k.$$

13.3 Exercises

Exercise 13.1. Let $(X, \|\cdot\|)$ be an infinite dimensional normed space and $E \subset X$ be a finite dimensional subspace. Show that $E \subset X$ is nowhere dense.

Exercise 13.2. Now suppose that $(X, \|\cdot\|)$ is an infinite dimensional Banach space. Show that X can not have a countable **algebraic** basis. More explicitly, there is no countable subset $S \subset X$ such that every element $x \in X$ may be written as a **finite** linear combination of elements from S. **Hint:** make use of Exercise 13.1 and the Baire category theorem.

14

Hilbert Space Basics

(BRUCE: Perhaps this should be move to between Chapters 7 & 8?)

Definition 14.1. *Let H be a complex vector space. An inner product on H is a function, $\langle\cdot|\cdot\rangle : H \times H \to \mathbb{C}$, such that*

1. *$\langle ax + by|z\rangle = a\langle x|z\rangle + b\langle y|z\rangle$ i.e. $x \to \langle x|z\rangle$ is linear.*
2. *$\overline{\langle x|y\rangle} = \langle y|x\rangle$.*
3. *$\|x\|^2 := \langle x|x\rangle \geq 0$ with equality $\|x\|^2 = 0$ iff $x = 0$.*

Notice that combining properties (1) and (2) that $x \to \langle z|x\rangle$ is anti-linear for fixed $z \in H$, i.e.

$$\langle z|ax + by\rangle = \bar{a}\langle z|x\rangle + \bar{b}\langle z|y\rangle.$$

The following identity will be used frequently in the sequel without further mention,

$$\begin{aligned} \|x + y\|^2 &= \langle x + y|x + y\rangle = \|x\|^2 + \|y\|^2 + \langle x|y\rangle + \langle y|x\rangle \\ &= \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\langle x|y\rangle. \end{aligned} \tag{14.1}$$

Theorem 14.2 (Schwarz Inequality). *Let $(H, \langle\cdot|\cdot\rangle)$ be an inner product space, then for all $x, y \in H$*

$$|\langle x|y\rangle| \leq \|x\|\|y\|$$

and equality holds iff x and y are linearly dependent.

Proof. If $y = 0$, the result holds trivially. So assume that $y \neq 0$ and observe; if $x = \alpha y$ for some $\alpha \in \mathbb{C}$, then $\langle x|y\rangle = \alpha \|y\|^2$ and hence

$$|\langle x|y\rangle| = |\alpha| \|y\|^2 = \|x\|\|y\|.$$

Now suppose that $x \in H$ is arbitrary, let $z := x - \|y\|^{-2}\langle x|y\rangle y$. (So z is the "orthogonal projection" of x onto y, see Figure 14.1.) Then

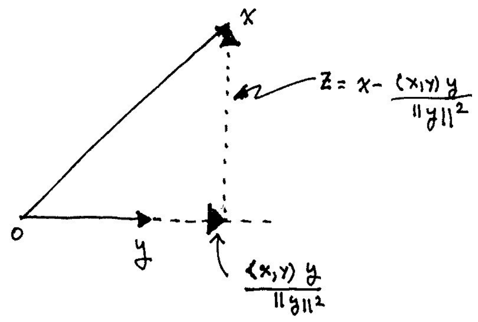


Fig. 14.1. The picture behind the proof of the Schwarz inequality.

$$\begin{aligned}0 \le \|z\|^2 = \left\| x - \frac{\langle x|y\rangle}{\|y\|^2} y \right\|^2 &= \|x\|^2 + \frac{|\langle x|y\rangle|^2}{\|y\|^4}\|y\|^2 - 2\mathrm{Re}\langle x|\frac{\langle x|y\rangle}{\|y\|^2} y\rangle \\ &= \|x\|^2 - \frac{|\langle x|y\rangle|^2}{\|y\|^2}\end{aligned}$$

from which it follows that $0 \le \|y\|^2\|x\|^2 - |\langle x|y\rangle|^2$ with equality iff $z = 0$ or equivalently iff $x = \|y\|^{-2}\langle x|y\rangle y$. ∎

Corollary 14.3. *Let $(H, \langle\cdot|\cdot\rangle)$ be an inner product space and $\|x\| := \sqrt{\langle x|x\rangle}$. Then the **Hilbertian norm,** $\|\cdot\|$, is a norm on H. Moreover $\langle\cdot|\cdot\rangle$ is continuous on $H \times H$, where H is viewed as the normed space $(H, \|\cdot\|)$.*

Proof. If $x, y \in H$, then, using the Schwarz's inequality,

$$\begin{aligned}\|x+y\|^2 &= \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\langle x|y\rangle \\ &\le \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| = (\|x\| + \|y\|)^2.\end{aligned}$$

Taking the square root of this inequality shows $\|\cdot\|$ satisfies the triangle inequality.

Checking that $\|\cdot\|$ satisfies the remaining axioms of a norm is now routine and will be left to the reader. If $x, x'y, y' \in H$, then

$$\begin{aligned}|\langle x|y\rangle - \langle x'|y'\rangle| &= |\langle x - x'|y\rangle + \langle x'|y - y'\rangle| \\ &\le \|y\|\|x - x'\| + \|x'\|\|y - y'\| \\ &\le \|y\|\|x - x'\| + (\|x\| + \|x - x'\|)\,\|y - y'\| \\ &= \|y\|\|x - x'\| + \|x\|\|y - y'\| + \|x - x'\|\|y - y'\|\end{aligned}$$

from which it follows that $\langle\cdot|\cdot\rangle$ is continuous. ∎

Definition 14.4. *Let $(H, \langle\cdot|\cdot\rangle)$ be an inner product space, we say $x, y \in H$ are **orthogonal** and write $x \perp y$ iff $\langle x|y\rangle = 0$. More generally if $A \subset H$ is a set, $x \in H$ is **orthogonal to** A (write $x \perp A$) iff $\langle x|y\rangle = 0$ for all $y \in A$. Let*

$A^\perp = \{x \in H : x \perp A\}$ be the set of vectors orthogonal to A. A subset $S \subset H$ is an ***orthogonal set*** *if $x \perp y$ for all distinct elements $x, y \in S$. If S further satisfies, $\|x\| = 1$ for all $x \in S$, then S is said to be* ***orthonormal set.***

Proposition 14.5. *Let $(H, \langle\cdot|\cdot\rangle)$ be an inner product space then*

1. *(**Parallelogram Law**)*
$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2 \tag{14.2}$$
for all $x, y \in H$.
2. *(**Pythagorean Theorem**) If $S \subset H$ is a finite orthogonal set, then*
$$\left\|\sum_{x\in S} x\right\|^2 = \sum_{x\in S}\|x\|^2. \tag{14.3}$$
3. *If $A \subset H$ is a set, then $A^\perp$ is a* ***closed*** *linear subspace of H.*

Remark 14.6. See Proposition 14.54 for the "converse" of the parallelogram law.

Proof. I will assume that H is a complex Hilbert space, the real case being easier. Items 1. and 2. are proved by the following elementary computations;

$$\begin{aligned}
&\|x+y\|^2 + \|x-y\|^2 \\
&\qquad = \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\langle x|y\rangle + \|x\|^2 + \|y\|^2 - 2\mathrm{Re}\langle x|y\rangle \\
&\qquad = 2\|x\|^2 + 2\|y\|^2,
\end{aligned}$$

and

$$\begin{aligned}
\left\|\sum_{x\in S} x\right\|^2 &= \langle\sum_{x\in S} x|\sum_{y\in S} y\rangle = \sum_{x,y\in S}\langle x|y\rangle \\
&= \sum_{x\in S}\langle x|x\rangle = \sum_{x\in S}\|x\|^2.
\end{aligned}$$

Item 3. is a consequence of the continuity of $\langle\cdot|\cdot\rangle$ and the fact that

$$A^\perp = \cap_{x\in A}\,\mathrm{Nul}(\langle\cdot|x\rangle)$$

where $\mathrm{Nul}(\langle\cdot|x\rangle) = \{y \in H : \langle y|x\rangle = 0\}$ – a closed subspace of H. ∎

Definition 14.7. *A* ***Hilbert space*** *is an inner product space $(H, \langle\cdot|\cdot\rangle)$ such that the induced Hilbertian norm is complete.*

Example 14.8. Suppose X is a set and $\mu : X \to (0,\infty)$, then $H := \ell^2(\mu)$ is a Hilbert space when equipped with the inner product,

$$\langle f|g\rangle := \sum_{x\in X} f(x)\,\bar{g}(x)\,\mu(x).$$

In Exercise 14.7 you will show every Hilbert space H is "equivalent" to a Hilbert space of this form with $\mu \equiv 1$.

More example of Hilbert spaces will be given later after we develop the Lebesgue integral, see Example 23.1 below.

Definition 14.9. *A subset C of a vector space X is said to be convex if for all $x, y \in C$ the line segment $[x,y] := \{tx + (1-t)y : 0 \le t \le 1\}$ joining x to y is contained in C as well. (Notice that any vector subspace of X is convex.)*

Theorem 14.10 (Best Approximation Theorem). *Suppose that H is a Hilbert space and $M \subset H$ is a closed convex subset of H. Then for any $x \in H$ there exists a unique $y \in M$ such that*

$$\|x-y\| = d(x,M) = \inf_{z\in M} \|x-z\|.$$

Moreover, if M is a vector subspace of H, then the point y may also be characterized as the unique point in M such that $(x-y) \perp M$.

Proof. Uniqueness. By replacing M by $M - x := \{m - x : m \in M\}$ we may assume $x = 0$. Let $\delta := d(0,M) = \inf_{m\in M} \|m\|$ and $y, z \in M$, see Figure 14.2.

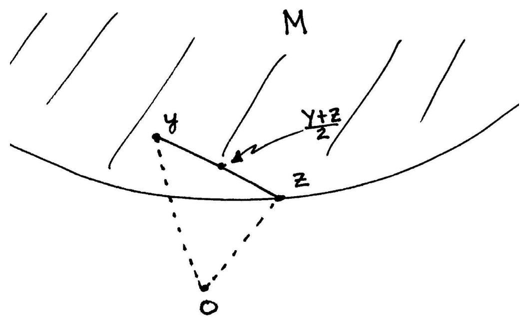


Fig. 14.2. The geometry of convex sets.

By the parallelogram law and the convexity of M,

$$\begin{aligned} 2\|y\|^2 + 2\|z\|^2 &= \|y+z\|^2 + \|y-z\|^2 \\ &= 4\left\|\frac{y+z}{2}\right\|^2 + \|y-z\|^2 \ge 4\delta^2 + \|y-z\|^2. \end{aligned} \tag{14.4}$$

Hence if $\|y\| = \|z\| = \delta$, then $2\delta^2 + 2\delta^2 \geq 4\delta^2 + \|y - z\|^2$, so that $\|y - z\|^2 = 0$. Therefore, if a minimizer for $d(0, \cdot)|_M$ exists, it is unique.

Existence. Let $y_n \in M$ be chosen such that $\|y_n\| = \delta_n \to \delta \equiv d(0, M)$. Taking $y = y_m$ and $z = y_n$ in Eq. (14.4) shows

$$2\delta_m^2 + 2\delta_n^2 \geq 4\delta^2 + \|y_n - y_m\|^2.$$

Passing to the limit $m, n \to \infty$ in this equation implies,

$$2\delta^2 + 2\delta^2 \geq 4\delta^2 + \limsup_{m,n\to\infty} \|y_n - y_m\|^2,$$

i.e. $\limsup_{m,n\to\infty} \|y_n - y_m\|^2 = 0$. Therefore, by completeness of H, $\{y_n\}_{n=1}^\infty$ is convergent. Because M is closed, $y := \lim\limits_{n\to\infty} y_n \in M$ and because the norm is continuous,

$$\|y\| = \lim_{n\to\infty} \|y_n\| = \delta = d(0, M).$$

So y is the desired point in M which is closest to 0.

Now suppose M is a closed subspace of H and $x \in H$. Let $y \in M$ be the closest point in M to x. Then for $w \in M$, the function

$$g(t) := \|x - (y + tw)\|^2 = \|x - y\|^2 - 2t\mathrm{Re}\langle x - y|w\rangle + t^2\|w\|^2$$

has a minimum at $t = 0$ and therefore $0 = g'(0) = -2\mathrm{Re}\langle x - y|w\rangle$. Since $w \in M$ is arbitrary, this implies that $(x - y) \perp M$.

Finally suppose $y \in M$ is any point such that $(x - y) \perp M$. Then for $z \in M$, by Pythagorean's theorem,

$$\|x - z\|^2 = \|x - y + y - z\|^2 = \|x - y\|^2 + \|y - z\|^2 \geq \|x - y\|^2$$

which shows $d(x, M)^2 \geq \|x - y\|^2$. That is to say y is the point in M closest to x. ■

Definition 14.11. *Suppose that $A : H \to H$ is a bounded operator. The **adjoint** of A, denote A^*, is the unique operator $A^* : H \to H$ such that $\langle Ax|y\rangle = \langle x|A^*y\rangle$. (The proof that A^* exists and is unique will be given in Proposition 14.16 below.) A bounded operator $A : H \to H$ is **self - adjoint** or **Hermitian** if $A = A^*$.*

Definition 14.12. *Let H be a Hilbert space and $M \subset H$ be a closed subspace. The orthogonal projection of H onto M is the function $P_M : H \to H$ such that for $x \in H$, $P_M(x)$ is the unique element in M such that $(x - P_M(x)) \perp M$.*

Theorem 14.13 (Projection Theorem). *Let H be a Hilbert space and $M \subset H$ be a closed subspace. The orthogonal projection P_M satisfies:*

1. *P_M is linear and hence we will write $P_M x$ rather than $P_M(x)$.*
2. *$P_M^2 = P_M$ (P_M is a projection).*

3. *$P_M^* = P_M$, (P_M is self-adjoint).*
4. *$\mathrm{Ran}(P_M) = M$ and $\mathrm{Nul}(P_M) = M^\perp$.*

Proof.

1. Let $x_1, x_2 \in H$ and $\alpha \in \mathbb{F}$, then $P_M x_1 + \alpha P_M x_2 \in M$ and
$$P_M x_1 + \alpha P_M x_2 - (x_1 + \alpha x_2) = [P_M x_1 - x_1 + \alpha(P_M x_2 - x_2)] \in M^\perp$$
showing $P_M x_1 + \alpha P_M x_2 = P_M(x_1 + \alpha x_2)$, i.e. P_M is linear.
2. Obviously $\mathrm{Ran}(P_M) = M$ and $P_M x = x$ for all $x \in M$. Therefore $P_M^2 = P_M$.
3. Let $x, y \in H$, then since $(x - P_M x)$ and $(y - P_M y)$ are in $M^\perp$,
$$\begin{aligned}\langle P_M x|y\rangle &= \langle P_M x|P_M y + y - P_M y\rangle = \langle P_M x|P_M y\rangle \\ &= \langle P_M x + (x - P_M x)|P_M y\rangle = \langle x|P_M y\rangle.\end{aligned}$$
4. We have already seen, $\mathrm{Ran}(P_M) = M$ and $P_M x = 0$ iff $x = x - 0 \in M^\perp$, i.e. $\mathrm{Nul}(P_M) = M^\perp$.

■

Corollary 14.14. *If $M \subset H$ is a proper closed subspace of a Hilbert space H, then $H = M \oplus M^\perp$.*

Proof. Given $x \in H$, let $y = P_M x$ so that $x - y \in M^\perp$. Then $x = y + (x - y) \in M + M^\perp$. If $x \in M \cap M^\perp$, then $x \perp x$, i.e. $\|x\|^2 = \langle x|x\rangle = 0$. So $M \cap M^\perp = \{0\}$. ■

Exercise 14.1. Suppose M is a subset of H, then $M^{\perp\perp} = \overline{\mathrm{span}(M)}$.

Theorem 14.15 (Riesz Theorem). *Let H^* be the dual space of H (Notation 7.9). The map*
$$z \in H \xrightarrow{j} \langle \cdot|z\rangle \in H^* \tag{14.5}$$
is a conjugate linear[1] isometric isomorphism.

Proof. The map j is conjugate linear by the axioms of the inner products. Moreover, for $x, z \in H$,
$$|\langle x|z\rangle| \le \|x\|\,\|z\| \text{ for all } x \in H$$
with equality when $x = z$. This implies that $\|jz\|_{H^*} = \|\langle \cdot|z\rangle\|_{H^*} = \|z\|$. Therefore j is isometric and this implies j is injective. To finish the proof we

[1] Recall that j is conjugate linear if
$$j(z_1 + \alpha z_2) = jz_1 + \bar{\alpha} j z_2$$
for all $z_1, z_2 \in H$ and $\alpha \in \mathbb{C}$.

must show that j is surjective. So let $f \in H^*$ which we assume, with out loss of generality, is non-zero. Then $M = \text{Nul}(f)$ – a closed proper subspace of H. Since, by Corollary 14.14, $H = M \oplus M^\perp$, $f : H/M \cong M^\perp \to \mathbb{F}$ is a linear isomorphism. This shows that $\dim(M^\perp) = 1$ and hence $H = M \oplus \mathbb{F}x_0$ where $x_0 \in M^\perp \setminus \{0\}$.[2] Choose $z = \lambda x_0 \in M^\perp$ such that $f(x_0) = \langle x_0|z\rangle$, i.e. $\lambda = \bar{f}(x_0)/\|x_0\|^2$. Then for $x = m + \lambda x_0$ with $m \in M$ and $\lambda \in \mathbb{F}$,

$$f(x) = \lambda f(x_0) = \lambda\langle x_0|z\rangle = \langle \lambda x_0|z\rangle = \langle m + \lambda x_0|z\rangle = \langle x|z\rangle$$

which shows that $f = jz$. ∎

Proposition 14.16 (Adjoints). *Let H and K be Hilbert spaces and $A : H \to K$ be a bounded operator. Then there exists a unique bounded operator $A^* : K \to H$ such that*

$$\langle Ax|y\rangle_K = \langle x|A^*y\rangle_H \text{ for all } x \in H \text{ and } y \in K. \tag{14.6}$$

Moreover, for all $A, B \in L(H,K)$ and $\lambda \in \mathbb{C}$,

1. $(A + \lambda B)^* = A^* + \bar{\lambda}B^*$,
2. $A^{**} := (A^*)^* = A$,
3. $\|A^*\| = \|A\|$ *and*
4. $\|A^*A\| = \|A\|^2$.
5. *If $K = H$, then $(AB)^* = B^*A^*$. In particular $A \in L(H)$ has a bounded inverse iff A^* has a bounded inverse and $(A^*)^{-1} = (A^{-1})^*$.*

Proof. For each $y \in K$, the map $x \to \langle Ax|y\rangle_K$ is in H^* and therefore there exists, by Theorem 14.15, a unique vector $z \in H$ such that

$$\langle Ax|y\rangle_K = \langle x|z\rangle_H \text{ for all } x \in H.$$

This shows there is a unique map $A^* : K \to H$ such that $\langle Ax|y\rangle_K = \langle x|A^*(y)\rangle_H$ for all $x \in H$ and $y \in K$.

To see A^* is linear, let $y_1, y_2 \in K$ and $\lambda \in \mathbb{C}$, then for any $x \in H$,

$$\begin{aligned}\langle Ax|y_1 + \lambda y_2\rangle_K &= \langle Ax|y_1\rangle_K + \bar{\lambda}\langle Ax|y_2\rangle_K \\ &= \langle x|A^*(y_1)\rangle_K + \bar{\lambda}\langle x|A^*(y_2)\rangle_K \\ &= \langle x|A^*(y_1) + \lambda A^*(y_2)\rangle_K\end{aligned}$$

and by the uniqueness of $A^*(y_1 + \lambda y_2)$ we find

$$A^*(y_1 + \lambda y_2) = A^*(y_1) + \lambda A^*(y_2).$$

This shows A^* is linear and so we will now write A^*y instead of $A^*(y)$.

[2] Alternatively, choose $x_0 \in M^\perp \setminus \{0\}$ such that $f(x_0) = 1$. For $x \in M^\perp$ we have $f(x - \lambda x_0) = 0$ provided that $\lambda := f(x)$. Therefore $x - \lambda x_0 \in M \cap M^\perp = \{0\}$, i.e. $x = \lambda x_0$. This again shows that $M^\perp$ is spanned by x_0.

Since

$$\langle A^*y|x\rangle_H = \overline{\langle x|A^*y\rangle_H} = \overline{\langle Ax|y\rangle_K} = \langle y|Ax\rangle_K$$

it follows that $A^{**} = A$. The assertion that $(A+\lambda B)^* = A^* + \bar{\lambda} B^*$ is Exercise 14.2.

Items 3. and 4. Making use of the Schwarz inequality (Theorem 14.2), we have

$$\begin{aligned}\|A^*\| &= \sup_{k\in K:\|k\|=1} \|A^*k\| \\ &= \sup_{k\in K:\|k\|=1} \sup_{h\in H:\|h\|=1} |\langle A^*k|h\rangle| \\ &= \sup_{h\in H:\|h\|=1} \sup_{k\in K:\|k\|=1} |\langle k|Ah\rangle| = \sup_{h\in H:\|h\|=1} \|Ah\| = \|A\|\end{aligned}$$

so that $\|A^*\| = \|A\|$. Since

$$\|A^*A\| \le \|A^*\| \, \|A\| = \|A\|^2$$

and

$$\begin{aligned}\|A\|^2 &= \sup_{h\in H:\|h\|=1} \|Ah\|^2 = \sup_{h\in H:\|h\|=1} |\langle Ah|Ah\rangle| \\ &= \sup_{h\in H:\|h\|=1} |\langle h|A^*Ah\rangle| \le \sup_{h\in H:\|h\|=1} \|A^*Ah\| = \|A^*A\| \end{aligned} \tag{14.7}$$

we also have $\|A^*A\| \le \|A\|^2 \le \|A^*A\|$ which shows $\|A\|^2 = \|A^*A\|$.

Alternatively, from Eq. (14.7),

$$\|A\|^2 \le \|A^*A\| \le \|A\| \, \|A^*\| \tag{14.8}$$

which then implies $\|A\| \le \|A^*\|$. Replacing A by A^* in this last inequality shows $\|A^*\| \le \|A\|$ and hence that $\|A^*\| = \|A\|$. Using this identity back in Eq. (14.8) proves $\|A\|^2 = \|A^*A\|$.

Now suppose that $K = H$. Then

$$\langle ABh|k\rangle = \langle Bh|A^*k\rangle = \langle h|B^*A^*k\rangle$$

which shows $(AB)^* = B^*A^*$. If A^{-1} exists then

$$\begin{aligned}&\left(A^{-1}\right)^* A^* = \left(AA^{-1}\right)^* = I^* = I \text{ and} \\ &A^* \left(A^{-1}\right)^* = \left(A^{-1}A\right)^* = I^* = I.\end{aligned}$$

This shows that A^* is invertible and $(A^*)^{-1} = \left(A^{-1}\right)^*$. Similarly if A^* is invertible then so is $A = A^{**}$. ■

Exercise 14.2. Let H, K, M be Hilbert spaces, $A, B \in L(H,K)$, $C \in L(K,M)$ and $\lambda \in \mathbb{C}$. Show $(A+\lambda B)^* = A^* + \bar{\lambda} B^*$ and $(CA)^* = A^*C^* \in L(M,H)$.

Exercise 14.3. Let $H = \mathbb{C}^n$ and $K = \mathbb{C}^m$ equipped with the usual inner products, i.e. $\langle z|w\rangle_H = z \cdot \bar{w}$ for $z, w \in H$. Let A be an $m \times n$ matrix thought of as a linear operator from H to K. Show the matrix associated to $A^* : K \to H$ is the conjugate transpose of A.

Lemma 14.17. *Suppose $A : H \to K$ is a bounded operator, then:*

1. $\mathrm{Nul}(A^*) = \mathrm{Ran}(A)^\perp$.
2. $\overline{\mathrm{Ran}(A)} = \mathrm{Nul}(A^*)^\perp$.
3. *if $K = H$ and $V \subset H$ is an A – invariant subspace (i.e. $A(V) \subset V$), then $V^\perp$ is A^* – invariant.*

Proof. An element $y \in K$ is in $\mathrm{Nul}(A^*)$ iff $0 = \langle A^*y|x\rangle = \langle y|Ax\rangle$ for all $x \in H$ which happens iff $y \in \mathrm{Ran}(A)^\perp$. Because, by Exercise 14.1, $\overline{\mathrm{Ran}(A)} = \mathrm{Ran}(A)^{\perp\perp}$, and so by the first item, $\overline{\mathrm{Ran}(A)} = \mathrm{Nul}(A^*)^\perp$. Now suppose $A(V) \subset V$ and $y \in V^\perp$, then

$$\langle A^*y|x\rangle = \langle y|Ax\rangle = 0 \text{ for all } x \in V$$

which shows $A^*y \in V^\perp$. ∎

14.1 Hilbert Space Basis

Proposition 14.18 (Bessel's Inequality). *Let T be an orthonormal set, then for any $x \in H$,*

$$\sum_{v \in T} |\langle x|v\rangle|^2 \le \|x\|^2 \text{ for all } x \in H. \tag{14.9}$$

In particular the set $T_x := \{v \in T : \langle x|v\rangle \ne 0\}$ is at most countable for all $x \in H$.

Proof. Let $\Gamma \subset T$ be any finite set. Then

$$\begin{aligned} 0 \le \|x - \sum_{v \in \Gamma} \langle x|v\rangle v\|^2 &= \|x\|^2 - 2\mathrm{Re} \sum_{v \in \Gamma} \langle x|v\rangle \; \langle v|x\rangle + \sum_{v \in \Gamma} |\langle x|v\rangle|^2 \\ &= \|x\|^2 - \sum_{v \in \Gamma} |\langle x|v\rangle|^2 \end{aligned}$$

showing that $\sum_{v \in \Gamma} |\langle x|v\rangle|^2 \le \|x\|^2$. Taking the supremum of this inequality over $\Gamma \subset\subset T$ then proves Eq. (14.9). ∎

Proposition 14.19. *Suppose $T \subset H$ is an orthogonal set. Then $s = \sum_{v \in T} v$ exists in H (see Definition 7.15) iff $\sum_{v \in T} \|v\|^2 < \infty$. (In particular T must be at most a countable set.) Moreover, if $\sum_{v \in T} \|v\|^2 < \infty$, then*

1. $\|s\|^2 = \sum_{v\in T} \|v\|^2$ *and*
2. $\langle s|x\rangle = \sum_{v\in T}\langle v|x\rangle$ *for all* $x \in H$.

Similarly if $\{v_n\}_{n=1}^{\infty}$ *is an orthogonal set, then* $s = \sum_{n=1}^{\infty} v_n$ *exists in* H *iff* $\sum_{n=1}^{\infty} \|v_n\|^2 < \infty$. *In particular if* $\sum_{n=1}^{\infty} v_n$ *exists, then it is independent of rearrangements of* $\{v_n\}_{n=1}^{\infty}$.

Proof. Suppose $s = \sum_{v\in T} v$ exists. Then there exists $\Gamma \subset\subset T$ such that

$$\sum_{v\in\Lambda} \|v\|^2 = \left\|\sum_{v\in\Lambda} v\right\|^2 \le 1$$

for all $\Lambda \subset\subset T \setminus \Gamma$, wherein the first inequality we have used Pythagorean's theorem. Taking the supremum over such Λ shows that $\sum_{v\in T\setminus\Gamma} \|v\|^2 \le 1$ and therefore

$$\sum_{v\in T} \|v\|^2 \le 1 + \sum_{v\in\Gamma} \|v\|^2 < \infty.$$

Conversely, suppose that $\sum_{v\in T} \|v\|^2 < \infty$. Then for all $\varepsilon > 0$ there exists $\Gamma_\varepsilon \subset\subset T$ such that if $\Lambda \subset\subset T \setminus \Gamma_\varepsilon$,

$$\left\|\sum_{v\in\Lambda} v\right\|^2 = \sum_{v\in\Lambda} \|v\|^2 < \varepsilon^2. \tag{14.10}$$

Hence by Lemma 7.16, $\sum_{v\in T} v$ exists.

For item 1, let Γ_ε be as above and set $s_\varepsilon := \sum_{v\in\Gamma_\varepsilon} v$. Then

$$|\|s\| - \|s_\varepsilon\|| \le \|s - s_\varepsilon\| < \varepsilon$$

and by Eq. (14.10),

$$0 \le \sum_{v\in T} \|v\|^2 - \|s_\varepsilon\|^2 = \sum_{v\notin\Gamma_\varepsilon} \|v\|^2 \le \varepsilon^2.$$

Letting $\varepsilon \downarrow 0$ we deduce from the previous two equations that $\|s_\varepsilon\| \to \|s\|$ and $\|s_\varepsilon\|^2 \to \sum_{v\in T} \|v\|^2$ as $\varepsilon \downarrow 0$ and therefore $\|s\|^2 = \sum_{v\in T} \|v\|^2$.

Item 2. is a special case of Lemma 7.16. For the final assertion, let $s_N := \sum_{n=1}^{N} v_n$ and suppose that $\lim_{N\to\infty} s_N = s$ exists in H and in particular $\{s_N\}_{N=1}^{\infty}$ is Cauchy. So for $N > M$.

$$\sum_{n=M+1}^{N} \|v_n\|^2 = \|s_N - s_M\|^2 \to 0 \text{ as } M, N \to \infty$$

which shows that $\sum_{n=1}^{\infty} \|v_n\|^2$ is convergent, i.e. $\sum_{n=1}^{\infty} \|v_n\|^2 < \infty$.

Alternative proof of item 1. We could use the last result to prove Item 1. Indeed, if $\sum_{v \in T} \|v\|^2 < \infty$, then T is countable and so we may write $T = \{v_n\}_{n=1}^{\infty}$. Then $s = \lim_{N \to \infty} s_N$ with s_N as above. Since the norm, $\|\cdot\|$, is continuous on H,

$$\begin{aligned} \|s\|^2 &= \lim_{N \to \infty} \|s_N\|^2 = \lim_{N \to \infty} \left\| \sum_{n=1}^{N} v_n \right\|^2 = \lim_{N \to \infty} \sum_{n=1}^{N} \|v_n\|^2 \\ &= \sum_{n=1}^{\infty} \|v_n\|^2 = \sum_{v \in T} \|v\|^2. \end{aligned}$$

■

Corollary 14.20. *Suppose H is a Hilbert space, $\beta \subset H$ is an orthonormal set and $M = \overline{\operatorname{span} \beta}$. Then*

$$P_M x = \sum_{u \in \beta} \langle x|u \rangle u, \tag{14.11}$$

$$\sum_{u \in \beta} |\langle x|u \rangle|^2 = \|P_M x\|^2 \quad \text{and} \tag{14.12}$$

$$\sum_{u \in \beta} \langle x|u \rangle \langle u|y \rangle = \langle P_M x|y \rangle \tag{14.13}$$

for all $x, y \in H$.

Proof. By Bessel's inequality, $\sum_{u \in \beta} |\langle x|u \rangle|^2 \leq \|x\|^2$ for all $x \in H$ and hence by Proposition 14.18, $Px := \sum_{u \in \beta} \langle x|u \rangle u$ exists in H and for all $x, y \in H$,

$$\langle Px|y \rangle = \sum_{u \in \beta} \langle \langle x|u \rangle u|y \rangle = \sum_{u \in \beta} \langle x|u \rangle \langle u|y \rangle. \tag{14.14}$$

Taking $y \in \beta$ in Eq. (14.14) gives $\langle Px|y \rangle = \langle x|y \rangle$, i.e. that $\langle x - Px|y \rangle = 0$ for all $y \in \beta$. So $(x - Px) \perp \operatorname{span} \beta$ and by continuity we also have $(x - Px) \perp M = \overline{\operatorname{span} \beta}$. Since Px is also in M, it follows from the definition of P_M that $Px = P_M x$ proving Eq. (14.11). Equations (14.12) and (14.13) now follow from (14.14), Proposition 14.19 and the fact that $\langle P_M x|y \rangle = \langle P_M^2 x|y \rangle = \langle P_M x|P_M y \rangle$ for all $x, y \in H$. ■

Exercise 14.4. Let $(H, \langle \cdot|\cdot \rangle)$ be a Hilbert space and suppose that $\{P_n\}_{n=1}^{\infty}$ is a sequence of orthogonal projection operators on H such that $P_n(H) \subset P_{n+1}(H)$ for all n. Let $M := \cup_{n=1}^{\infty} P_n(H)$ (a subspace of H) and let P denote orthonormal projection onto $\bar{M}$. Show $\lim_{n \to \infty} P_n x = Px$ for all $x \in H$. **Hint:** first prove the result for $x \in M^{\perp}$, then for $x \in M$ and then for $x \in \bar{M}$.

Definition 14.21 (Basis). *Let H be a Hilbert space. A* ***basis*** *β of H is a maximal orthonormal subset $\beta \subset H$.*

Proposition 14.22. *Every Hilbert space has an orthonormal basis.*

Proof. Let $\mathcal{F}$ be the collection of all orthonormal subsets of H ordered by inclusion. If $\Phi \subset \mathcal{F}$ is linearly ordered then $\cup\Phi$ is an upper bound. By Zorn's Lemma (see Theorem B.7) there exists a maximal element $\beta \in \mathcal{F}$. ∎

An orthonormal set $\beta \subset H$ is said to be **complete** if $\beta^{\perp} = \{0\}$. That is to say if $\langle x|u\rangle = 0$ for all $u \in \beta$ then $x = 0$.

Lemma 14.23. *Let β be an orthonormal subset of H then the following are equivalent:*

1. *β is a basis,*
2. *β is complete and*
3. *$\overline{\operatorname{span} \beta} = H$.*

Proof. (1. $\iff$ 2.) If β is not complete, then there exists a unit vector $x \in \beta^{\perp} \setminus \{0\}$. The set $\beta \cup \{x\}$ is an orthonormal set properly containing β, so β is not maximal. Conversely, if β is not maximal, there exists an orthonormal set $\beta_1 \subset H$ such that $\beta \subsetneqq \beta_1$. Then if $x \in \beta_1 \setminus \beta$, we have $\langle x|u\rangle = 0$ for all $u \in \beta$ showing β is not complete.

(2. $\iff$ 3.) If β is not complete and $x \in \beta^{\perp} \setminus \{0\}$, then $\overline{\operatorname{span} \beta} \subset x^{\perp}$ which is a proper subspace of H. Conversely if $\overline{\operatorname{span} \beta}$ is a proper subspace of H, $\beta^{\perp} = \overline{\operatorname{span} \beta}^{\perp}$ is a non-trivial subspace by Corollary 14.14 and β is not complete. ∎

Theorem 14.24. *Let $\beta \subset H$ be an orthonormal set. Then the following are equivalent:*

1. *β is complete, i.e. β is an orthonormal basis for H.*
2. *$x = \sum_{u \in \beta} \langle x|u\rangle u$ for all $x \in H$.*
3. *$\langle x|y\rangle = \sum_{u \in \beta} \langle x|u\rangle \, \langle u|y\rangle$ for all $x, y \in H$.*
4. *$\|x\|^2 = \sum_{u \in \beta} |\langle x|u\rangle|^2$ for all $x \in H$.*

Proof. Let $M = \overline{\operatorname{span} \beta}$ and $P = P_M$.

(1) $\Rightarrow$ (2) By Corollary 14.20, $\sum_{u \in \beta} \langle x|u\rangle u = P_M x$. Therefore

$$x - \sum_{u \in \beta} \langle x|u\rangle u = x - P_M x \in M^{\perp} = \beta^{\perp} = \{0\}.$$

(2) $\Rightarrow$ (3) is a consequence of Proposition 14.19.

(3) $\Rightarrow$ (4) is obvious, just take $y = x$.

(4) $\Rightarrow$ (1) If $x \in \beta^{\perp}$, then by 4), $\|x\| = 0$, i.e. $x = 0$. This shows that β is complete. ■

Suppose $\Gamma := \{u_n\}_{n=1}^{\infty}$ is a collection of vectors in an inner product space $(H, \langle\cdot|\cdot\rangle)$. The standard **Gram-Schmidt** process produces from Γ an orthonormal subset, $\beta = \{v_n\}_{n=1}^{\infty}$, such that every element $u_n \in \Gamma$ is a finite linear combination of elements from β. Recall the procedure is to define v_n inductively by setting

$$\tilde{v}_{n+1} := v_{n+1} - \sum_{j=1}^{n} \langle u_{n+1}|v_j\rangle v_j = v_{n+1} - P_n v_{n+1}$$

where P_n is orthogonal projection onto $M_n := \mathrm{span}(\{v_k\}_{k=1}^{n})$. If $v_{n+1} := 0$, let $\tilde{v}_{n+1} = 0$, otherwise set $v_{n+1} := \|\tilde{v}_{n+1}\|^{-1} \tilde{v}_{n+1}$. Finally re-index the resulting sequence so as to throw out those v_n with $v_n = 0$. The result is an orthonormal subset, $\beta \subset H$, with the desired properties.

Definition 14.25. *As subset, Γ, of a normed space X is said to be **total** if* $\mathrm{span}(\Gamma)$ *is a dense in X.*

Remark 14.26. Suppose that $\{u_n\}_{n=1}^{\infty}$ is a **total** subset of H. Let $\{v_n\}_{n=1}^{\infty}$ be the vectors found by performing Gram-Schmidt on the set $\{u_n\}_{n=1}^{\infty}$. Then $\beta := \{v_n\}_{n=1}^{\infty}$ is an orthonormal basis for H. Indeed, if $h \in H$ is orthogonal to β then h is orthogonal to $\{u_n\}_{n=1}^{\infty}$ and hence also $\overline{\mathrm{span}\,\{u_n\}_{n=1}^{\infty}} = H$. In particular h is orthogonal to itself and so $h = 0$.

Proposition 14.27. *A Hilbert space H is separable iff H has a countable orthonormal basis $\beta \subset H$. Moreover, if H is separable, all orthonormal bases of H are countable. (See Proposition 4.14 in Conway's, "A Course in Functional Analysis," for a more general version of this proposition.)*

Proof. Let $\mathbb{D} \subset H$ be a countable dense set $\mathbb{D} = \{u_n\}_{n=1}^{\infty}$. By Gram-Schmidt process there exists $\beta = \{v_n\}_{n=1}^{\infty}$ an orthonormal set such that $\mathrm{span}\{v_n : n = 1, 2\ldots, N\} \supseteq \mathrm{span}\{u_n : n = 1, 2\ldots, N\}$. So if $\langle x|v_n\rangle = 0$ for all n then $\langle x|u_n\rangle = 0$ for all n. Since $\mathbb{D} \subset H$ is dense we may choose $\{w_k\} \subset \mathbb{D}$ such that $x = \lim_{k\to\infty} w_k$ and therefore $\langle x|x\rangle = \lim_{k\to\infty}\langle x|w_k\rangle = 0$. That is to say $x = 0$ and β is complete. Conversely if $\beta \subset H$ is a countable orthonormal basis, then the countable set

$$\mathbb{D} = \left\{ \sum_{u\in\beta} a_u u : a_u \in \mathbb{Q} + i\mathbb{Q} : \#\{u : a_u \neq 0\} < \infty \right\}$$

is dense in H. Finally let $\beta = \{u_n\}_{n=1}^{\infty}$ be an orthonormal basis and $\beta_1 \subset H$ be another orthonormal basis. Then the sets

$$B_n = \{v \in \beta_1 : \langle v|u_n\rangle \neq 0\}$$

are countable for each $n \in \mathbb{N}$ and hence $B := \bigcup_{n=1}^{\infty} B_n$ is a countable subset of β_1. Suppose there exists $v \in \beta_1 \setminus B$, then $\langle v|u_n\rangle = 0$ for all n and since $\beta = \{u_n\}_{n=1}^{\infty}$ is an orthonormal basis, this implies $v = 0$ which is impossible since $\|v\| = 1$. Therefore $\beta_1 \setminus B = \emptyset$ and hence $\beta_1 = B$ is countable. ■

Proposition 14.28. *Suppose X and Y are sets and $\mu : X \to (0,\infty)$ and $\nu : Y \to (0,\infty)$ are give weight functions. For functions $f : X \to \mathbb{C}$ and $g : Y \to \mathbb{C}$ let $f \otimes g : X \times Y \to \mathbb{C}$ be defined by $f \otimes g\,(x,y) := f(x)\,g(y)$. If $\beta \subset \ell^2(\mu)$ and $\gamma \subset \ell^2(\nu)$ are orthonormal bases, then*

$$\beta \otimes \gamma := \{f \otimes g : f \in \beta \text{ and } g \in \gamma\}$$

is an orthonormal basis for $\ell^2(\mu \otimes \nu)$.

Proof. Let $f, f' \in \ell^2(\mu)$ and $g, g' \in \ell^2(\nu)$, then by the Tonelli's Theorem 4.22 for sums and Hölder's inequality,

$$\begin{aligned}\sum_{X\times Y} |f \otimes g \cdot f' \otimes g'|\, \mu \otimes \nu &= \sum_X |ff'|\, \mu \cdot \sum_Y |gg'|\, \nu \\ &\le \|f\|_{\ell^2(\mu)} \|f'\|_{\ell^2(\mu)} \|g\|_{\ell^2(\nu)} \|g'\|_{\ell^2(\nu)} = 1 < \infty.\end{aligned}$$

So by Fubini's Theorem 4.23 for sums,

$$\begin{aligned}\langle f \otimes g | f' \otimes g'\rangle_{\ell^2(\mu\otimes\nu)} &= \sum_X ff'\mu \cdot \sum_Y g\bar{g}'\nu \\ &= \langle f|f'\rangle_{\ell^2(\mu)} \langle g|g'\rangle_{\ell^2(\nu)} = \delta_{f,f'}\delta_{g,g'}.\end{aligned}$$

Therefore, $\beta \otimes \gamma$ is an orthonormal subset of $\ell^2(\mu \otimes \nu)$. So it only remains to show $\beta \otimes \gamma$ is complete. We will give two proofs of this fact. Let $F \in \ell^2(\mu \otimes \nu)$. In the first proof we will verify item 4. of Theorem 14.24 while in the second we will verify item 1 of Theorem 14.24.

First Proof. By Tonelli's Theorem,

$$\sum_{x\in X} \mu(x) \sum_{y\in Y} \nu(y)\, |F(x,y)|^2 = \|F\|^2_{\ell^2(\mu\otimes\nu)} < \infty$$

and since $\mu > 0$, it follows that

$$\sum_{y\in Y} |F(x,y)|^2\, \nu(y) < \infty \text{ for all } x \in X,$$

i.e. $F(x,\cdot) \in \ell^2(\nu)$ for all $x \in X$. By the completeness of γ,

$$\sum_Y |F(x,y)|^2\, \nu(y) = \langle F(x,\cdot)\, | F(x,\cdot)\rangle_{\ell^2(\nu)} = \sum_{g\in\gamma} \left|\langle F(x,\cdot)\, |g\rangle_{\ell^2(\nu)}\right|^2$$

and therefore,

$$\|F\|^2_{\ell^2(\mu\otimes\nu)} = \sum_{x\in X}\mu(x)\sum_{y\in Y}\nu(y)\left|F(x,y)\right|^2$$
$$= \sum_{x\in X}\sum_{g\in\gamma}\left|\langle F(x,\cdot)|g\rangle_{\ell^2(\nu)}\right|^2\mu(x). \tag{14.15}$$

and in particular, $x \to \langle F(x,\cdot)|g\rangle_{\ell^2(\nu)}$ is in $\ell^2(\mu)$. So by the completeness of β and the Fubini and Tonelli theorems, we find

$$\begin{aligned}\sum_X \left|\langle F(x,\cdot)|g\rangle_{\ell^2(\nu)}\right|^2\mu(x) &= \sum_{f\in\beta}\left|\sum_X \langle F(x,\cdot)|g\rangle_{\ell^2(\nu)} f(x)\,\mu(x)\right|^2\\ &= \sum_{f\in\beta}\left|\sum_X\left(\sum_Y F(x,y)\,g(y)\,\nu(y)\right)f(x)\,\mu(x)\right|^2\\ &= \sum_{f\in\beta}\left|\sum_{X\times Y}F(x,y)\,\overline{f\otimes g}(x,y)\,\mu\otimes\nu(x,y)\right|^2\\ &= \sum_{f\in\beta}\left|\langle F|f\otimes g\rangle_{\ell^2(\mu\otimes\nu)}\right|^2.\end{aligned}$$

Combining this result with Eq. (14.15) shows

$$\|F\|^2_{\ell^2(\mu\otimes\nu)} = \sum_{f\in\beta,\ g\in\gamma}\left|\langle F|f\otimes g\rangle_{\ell^2(\mu\otimes\nu)}\right|^2$$

as desired.

Second Proof. Suppose, for all $f \in \beta$ and $g \in \gamma$ that $\langle F|f\otimes g\rangle = 0$, i.e.

$$0 = \langle F|f\otimes g\rangle_{\ell^2(\mu\otimes\nu)} = \sum_{x\in X}\mu(x)\sum_{y\in Y}\nu(y)\,F(x,y)\bar f(x)g(y)$$
$$= \sum_{x\in X}\mu(x)\,\langle F(x,\cdot)|g\rangle_{\ell^2(\nu)}\bar f(x). \tag{14.16}$$

Since

$$\sum_{x\in X}\left|\langle F(x,\cdot)|g\rangle_{\ell^2(\nu)}\right|^2\mu(x) \le \sum_{x\in X}\mu(x)\sum_{y\in Y}\left|F(x,y)\right|^2\nu(y) < \infty, \tag{14.17}$$

it follows from Eq. (14.16) and the completeness of β that $\langle F(x,\cdot)|g\rangle_{\ell^2(\nu)} = 0$ for all $x \in X$. By the completeness of γ we conclude that $F(x,y) = 0$ for all $(x,y) \in X \times Y$. ∎

Definition 14.29. *A linear map $U : H \to K$ is an* ***isometry*** *if $\|Ux\|_K = \|x\|_H$ for all $x \in H$ and U is* ***unitary if*** *U is also surjective.*

Exercise 14.5. Let $U : H \to K$ be a linear map, show the following are equivalent:

1. $U : H \to K$ is an isometry,
2. $\langle Ux|Ux'\rangle_K = \langle x|x'\rangle_H$ for all $x, x' \in H$, (see Eq. (14.33) below)
3. $U^*U = id_H$.

Exercise 14.6. Let $U : H \to K$ be a linear map, show the following are equivalent:

1. $U : H \to K$ is unitary
2. $U^*U = id_H$ and $UU^* = id_K$.
3. U is invertible and $U^{-1} = U^*$.

Exercise 14.7. Let H be a Hilbert space. Use Theorem 14.24 to show there exists a set X and a unitary map $U : H \to \ell^2(X)$. Moreover, if H is separable and $\dim(H) = \infty$, then X can be taken to be $\mathbb{N}$ so that H is unitarily equivalent to $\ell^2 = \ell^2(\mathbb{N})$.

14.2 Some Spectral Theory

For this section let H and K be two Hilbert space over $\mathbb{C}$.

Exercise 14.8. Suppose $A : H \to H$ is a bounded self-adjoint operator. Show:

1. If λ is an eigenvalue of A, i.e. $Ax = \lambda x$ for some $x \in H \setminus \{0\}$, then $\lambda \in \mathbb{R}$.
2. If λ and μ are two distinct eigenvalues of A with eigenvectors x and y respectively, then $x \perp y$.

Unlike in finite dimensions, it is possible that an operator on a complex Hilbert space may have no eigenvalues, see Example 14.35 and Lemma 14.36 below for a couple of examples. For this reason it is useful to generalize the notion of an eigenvalue as follows.

Definition 14.30. *Suppose X is a Banach space over $\mathbb{F}$ ($\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$) and $A \in L(X)$. We say $\lambda \in \mathbb{F}$ is in the* ***spectrum*** *of A if $A - \lambda I$ does* ***not*** *have a bounded*[3] *inverse. The* ***spectrum*** *will be denoted by $\sigma(A) \subset \mathbb{F}$. The* ***resolvent set*** *for A is $\rho(A) := \mathbb{F} \setminus \sigma(A)$.*

Remark 14.31. If λ is an eigenvalue of A, then $A - \lambda I$ is not injective and hence not invertible. Therefore any eigenvalue of A is in the spectrum of A. If H is a Hilbert space ant $A \in L(H)$, it follows from item 5. of Proposition 14.16 that $\lambda \in \sigma(A)$ iff $\bar{\lambda} \in \sigma(A^*)$, i.e.

$$\sigma(A^*) = \left\{\bar{\lambda} : \lambda \in \sigma(A)\right\}.$$

[3] It will follow by the open mapping Theorem **??** or the closed graph Theorem **??** that the word bounded may be omitted from this definition.

Exercise 14.9. Suppose X is a complex Banach space and $A \in GL(X)$. Show

$$\sigma\left(A^{-1}\right) = \sigma(A)^{-1} := \left\{\lambda^{-1} : \lambda \in \sigma(A)\right\}.$$

If we further assume A is both invertible and isometric, i.e. $\|Ax\| = \|x\|$ for all $x \in X$, then show

$$\sigma(A) \subset S^1 := \{z \in \mathbb{C} : |z| = 1\}.$$

Hint: working formally,

$$\left(A^{-1} - \lambda^{-1}\right)^{-1} = \frac{1}{\frac{1}{A} - \frac{1}{\lambda}} = \frac{1}{\frac{\lambda - A}{A\lambda}} = \frac{A\lambda}{\lambda - A}$$

from which you might expect that $\left(A^{-1} - \lambda^{-1}\right)^{-1} = -\lambda A(A - \lambda)^{-1}$ if $\lambda \in \rho(A)$.

Exercise 14.10. Suppose X is a Banach space and $A \in L(X)$. Use Corollary 7.20 to show $\sigma(A)$ is a closed subset of $\left\{\lambda \in \mathbb{F} : |\lambda| \leq \|A\| := \|A\|_{L(X)}\right\}$.

Lemma 14.32. *Suppose that $A \in L(H)$ is a normal operator, i.e. $[A, A^*] = 0$. Then $\lambda \in \sigma(A)$ iff*

$$\inf_{\|\psi\|=1} \|(A - \lambda 1)\psi\| = 0. \tag{14.18}$$

In other words, $\lambda \in \sigma(A)$ iff there is an "approximate sequence of eigenvectors" for (A, λ), i.e. there exists $\psi_n \in H$ such that $\|\psi_n\| = 1$ and $A\psi_n - \lambda\psi_n \to 0$ as $n \to \infty$.

Proof. By replacing A by $A - \lambda I$ we may assume that $\lambda = 0$. If $0 \notin \sigma(A)$, then

$$\inf_{\|\psi\|=1} \|A\psi\| = \inf \frac{\|A\psi\|}{\|\psi\|} = \inf \frac{\|\psi\|}{\|A^{-1}\psi\|} = 1/\left\|A^{-1}\right\| > 0.$$

Now suppose that $\inf_{\|\psi\|=1} \|A\psi\| = \varepsilon > 0$ or equivalently we have

$$\|A\psi\| \geq \varepsilon \|\psi\|$$

for all $\psi \in H$. Because A is normal,

$$\|A\psi\|^2 = \langle A^*A\psi|\psi\rangle = \langle AA^*\psi|\psi\rangle = \langle A^*\psi|A^*\psi\rangle = \|A^*\psi\|^2.$$

Therefore we also have

$$\|A^*\psi\| = \|A\psi\| \geq \varepsilon \|\psi\| \;\; \forall\; \psi \in H. \tag{14.19}$$

This shows in particular that A and A^* are injective, $\mathrm{Ran}(A)$ is closed and hence by Lemma 14.17

$$\mathrm{Ran}(A) = \overline{\mathrm{Ran}(A)} = \mathrm{Nul}(A^*)^{\perp} = \{0\}^{\perp} = H.$$

Therefore A is algebraically invertible and the inverse is bounded by Eq. (14.19). ■

Lemma 14.33. *Suppose that $A \in L(H)$ is self-adjoint (i.e. $A = A^*$) then*

$$\sigma(A) \subset \left[-\|A\|_{op}, \|A\|_{op}\right] \subset \mathbb{R}.$$

Proof. Writting $\lambda = \alpha + i\beta$ with $\alpha, \beta \in \mathbb{R}$, then

$$\begin{aligned} \|(A + \alpha + i\beta)\psi\|^2 &= \|(A+\alpha)\psi\|^2 + |\beta|^2 \|\psi\|^2 + 2\operatorname{Re}((A+\alpha)\psi, i\beta\psi) \\ &= \|(A+\alpha)\psi\|^2 + |\beta|^2 \|\psi\|^2 \end{aligned} \tag{14.20}$$

wherein we have used

$$\operatorname{Re}\left[i\beta((A+\alpha)\psi,\psi)\right] = \beta\operatorname{Im}((A+\alpha)\psi,\psi) = 0$$

since

$$((A+\alpha)\psi,\psi) = (\psi,(A+\alpha)\psi) = \overline{((A+\alpha)\psi,\psi)}.$$

Eq. (14.20) along with Lemma 14.32 shows that $\lambda \notin \sigma(A)$ if $\beta \neq 0$, i.e. $\sigma(A) \subset \mathbb{R}$. The fact that $\sigma(A)$ is now contained in $\left[-\|A\|_{op}, \|A\|_{op}\right]$ is a consequence of Exercise 14.9. ∎

Remark 14.34. It is not true that $\sigma(A) \subset \mathbb{R}$ implies $A = A^*$. For example let $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ on $H = \mathbb{C}^2$, then $\sigma(A) = \{0\}$ yet $A \neq A^*$.

Example 14.35. Let $S \in L(H)$ be a (not necessarily) normal operator. The proof of Lemma 14.32 gives $\lambda \in \sigma(S)$ if Eq. (14.18) holds. However the converse is not always valid unless S is normal. For example, let $S : \ell^2 \to \ell^2$ be the shift, $S(\omega_1, \omega_2, \dots) = (0, \omega_1, \omega_2, \dots)$. Then for any $\lambda \in D := \{z \in \mathbb{C} : |z| < 1\}$,

$$\|(S-\lambda)\psi\| = \|S\psi - \lambda\psi\| \geq |\|S\psi\| - |\lambda|\|\psi\|| = (1-|\lambda|)\|\psi\|$$

and so there does not exists an approximate sequence of eigenvectors for (S, λ). However, as we will now show, $\sigma(S) = \bar{D}$.

To prove this it suffices to show by Remark 14.31 and Exercise 14.9 that $D \subset \sigma(S^*)$. For if this is the cae then $\bar{D} \subset \sigma(S^*) \subset \bar{D}$ and hence $\sigma(S) = \bar{D}$ since $\bar{D}$ is invariant under complex conjugation.

A simple computation shows,

$$S^*(\omega_1, \omega_2, \dots) = (\omega_2, \omega_3, \dots)$$

and $\omega = (\omega_1, \omega_2, \dots)$ is an eigenvector for S^* with eigenvalue $\lambda \in \mathbb{C}$ iff

$$0 = (S^* - \lambda I)(\omega_1, \omega_2, \dots) = (\omega_2 - \lambda\omega_1, \omega_3 - \lambda\omega_2, \dots).$$

Solving these equation shows

$$\omega_2 = \lambda\omega_1,\ \omega_3 = \lambda\omega_2 = \lambda^2\omega_1\ , \dots,\ \omega_n = \lambda^{n-1}\omega_1.$$

Hence if $\lambda \in D$, we may let $\omega_1 = 1$ above to find

$$S^*(1, \lambda, \lambda^2, \dots) = \lambda(1, \lambda, \lambda^2, \dots)$$

where $(1, \lambda, \lambda^2, \dots) \in \ell^2$. Thus we have shown λ is an eigenvalue for S^* for all $\lambda \in D$ and hence $D \subset \sigma(S^*)$.

Lemma 14.36. *Let $H = \ell^2(\mathbb{Z})$ and let $A : H \to H$ be defined by*

$$Af(k) = i(f(k+1) - f(k-1)) \text{ for all } k \in \mathbb{Z}.$$

Then:

1. *A is a bounded self-adjoint operator.*
2. *A has no eigenvalues.*
3. *$\sigma(A) = [-2, 2]$.*

Proof. For another (simpler) proof of this lemma, see Exercise 23.8 below.
1. Since

$$\|Af\|_2 \le \|f(\cdot + 1)\|_2 + \|f(\cdot - 1)\|_2 = 2\|f\|_2,$$

$\|A\|_{op} \le 2 < \infty$. Moreover, for $f, g \in \ell^2(\mathbb{Z})$,

$$\begin{aligned}\langle Af|g\rangle &= \sum_k i(f(k+1) - f(k-1))\bar{g}(k) \\ &= \sum_k if(k)\bar{g}(k-1) - \sum_k if(k)\bar{g}(k+1) \\ &= \sum_k f(k)\overline{Ag(k)} = \langle f|Ag\rangle,\end{aligned}$$

which shows $A = A^*$.

2. From Lemma 14.33, we know that $\sigma(A) \subset [-2, 2]$. If $\lambda \in [-2, 2]$ and $f \in H$ satisfies $Af = \lambda f$, then

$$f(k+1) = -i\lambda f(k) + f(k-1) \text{ for all } k \in \mathbb{Z}. \tag{14.21}$$

This is a second order difference equations which can be solved analogously to second order ordinary differential equations. The idea is to start by looking for a solution of the form $f(k) = \alpha^k$. Then Eq. (14.21) beocmes, $\alpha^{k+1} = -i\lambda\alpha^k + \alpha^{k-1}$ or equivalently that

$$\alpha^2 + i\lambda\alpha - 1 = 0.$$

So we will have a solution if $\alpha \in \{\alpha_\pm\}$ where

$$\alpha_\pm = \frac{-i\lambda \pm \sqrt{4 - \lambda^2}}{2}.$$

For $|\lambda| \neq 2$, there are two distinct roots and the general solution to Eq. (14.21) is of the form

$$f(k) = c_+ \alpha_+^k + c_- \alpha_-^k \tag{14.22}$$

for some constants $c_\pm \in \mathbb{C}$ and $|\lambda| = 2$, the general solution has the form

$$f(k) = c\alpha_+^k + dk\alpha_+^k \tag{14.23}$$

Since in all cases, $|\alpha_\pm| = \frac{1}{4}\left(\lambda^2 + 4 - \lambda^2\right) = 1$, it follows that neither of these functions, f, will be in $\ell^2(\mathbb{Z})$ unless they are identically zero. This shows that A has no eigenvalues.

3. The above argument suggest a method for constructing approximate eigenfucntions. Namely, let $\lambda \in [-2, 2]$ and define $f_n(k) := 1_{|k| \leq n} \alpha^k$ where $\alpha = \alpha_+$. Then a simple computation shows

$$\lim_{n\to\infty} \frac{\|(A - \lambda I) f_n\|_2}{\|f_n\|_2} = 0 \tag{14.24}$$

and therefore $\lambda \in \sigma(A)$. ■

Exercise 14.11. Verify Eq. (14.24). Also show by explicit computations that

$$\lim_{n\to\infty} \frac{\|(A - \lambda I) f_n\|_2}{\|f_n\|_2} \neq 0$$

if $\lambda \notin [-2, 2]$.

The next couple of results will be needed for the next section.

Theorem 14.37 (Rayleigh quotient). *Suppose $T \in L(H) := L(H, H)$ is a bounded self-adjoint operator, then*

$$\|T\| = \sup_{f \neq 0} \frac{|\langle f|Tf\rangle|}{\|f\|^2}.$$

Moreover if there exists a non-zero element $g \in H$ such that

$$\frac{|\langle Tg|g\rangle|}{\|g\|^2} = \|T\|,$$

then g is an eigenvector of T with $Tg = \lambda g$ and $\lambda \in \{\pm\|T\|\}$.

Proof. Let

$$M := \sup_{f \neq 0} \frac{|\langle f|Tf\rangle|}{\|f\|^2}.$$

We wish to show $M = \|T\|$. Since

$$|\langle f|Tf\rangle| \leq \|f\|\|Tf\| \leq \|T\|\|f\|^2,$$

we see $M \le \|T\|$. Conversely let $f, g \in H$ and compute

$$\begin{aligned}\langle f+g|T(f+g)\rangle &- (f-g|T(f-g)\rangle \\ &= \langle f|Tg\rangle + \langle g|Tf\rangle + \langle f|Tg\rangle + \langle g|Tf\rangle \\ &= 2[\langle f|Tg\rangle + \langle Tg|f\rangle] = 2[\langle f|Tg\rangle + \overline{\langle f|Tg\rangle}] \\ &= 4\mathrm{Re}\langle f|Tg\rangle.\end{aligned}$$

Therefore, if $\|f\| = \|g\| = 1$, it follows that

$$|\mathrm{Re}\langle f|Tg\rangle| \le \frac{M}{4}\left\{\|f+g\|^2 + \|f-g\|^2\right\} = \frac{M}{4}\left\{2\|f\|^2 + 2\|g\|^2\right\} = M.$$

By replacing f be $e^{i\theta} f$ where θ is chosen so that $e^{i\theta}\langle f|Tg\rangle$ is real, we find

$$|\langle f|Tg\rangle| \le M \text{ for all } \|f\| = \|g\| = 1.$$

Hence

$$\|T\| = \sup_{\|f\|=\|g\|=1} |\langle f|Tg\rangle| \le M.$$

If $g \in H \setminus \{0\}$ and $\|T\| = |\langle Tg|g\rangle|/\|g\|^2$ then, using the Cauchy Schwarz inequality,

$$\|T\| = \frac{|\langle Tg|g\rangle|}{\|g\|^2} \le \frac{\|Tg\|}{\|g\|} \le \|T\|. \tag{14.25}$$

This implies $|\langle Tg|g\rangle| = \|Tg\|\|g\|$ and forces equality in the Cauchy Schwarz inequality. So by Theorem 14.2, Tg and g are linearly dependent, i.e. $Tg = \lambda g$ for some $\lambda \in \mathbb{C}$. Substituting this into (14.25) shows that $|\lambda| = \|T\|$. Since T is self-adjoint,

$$\lambda\|g\|^2 = \langle \lambda g|g\rangle = \langle Tg|g\rangle = \langle g|Tg\rangle = \langle g|\lambda g\rangle = \bar{\lambda}\langle g|g\rangle,$$

which implies that $\lambda \in \mathbb{R}$ and therefore, $\lambda \in \{\pm\|T\|\}$. ■

Lemma 14.38 (Invariant subspaces). *Let $T : H \to H$ be a self-adjoint operator and M be a T – invariant subspace of H, i.e. $T(M) \subset M$. Then $M^\perp$ is also a T – invariant subspace, i.e. $T(M^\perp) \subset M^\perp$.*

Proof. Let $x \in M$ and $y \in M^\perp$, then $Tx \in M$ and hence

$$0 = \langle Tx|y\rangle = \langle x|Ty\rangle \text{ for all } x \in M.$$

Thus $Ty \in M^\perp$. ■

14.3 Compact Operators on a Hilbert Space

In this section let H and B be Hilbert spaces and $U := \{x \in H : \|x\| < 1\}$ be the **unit ball** in H. Recall from Definition 11.16 that a bounded operator, $K : H \to B$, is compact iff $\overline{K(U)}$ is compact in B. Equivalently, for all bounded sequences $\{x_n\}_{n=1}^{\infty} \subset H$, the sequence $\{Kx_n\}_{n=1}^{\infty}$ has a convergent subsequence in B. Because of Theorem 11.15, if $\dim(H) = \infty$ and $K : H \to B$ is invertible, then K is **not** compact.

Definition 14.39. *$K : H \to B$ is said to have **finite rank** if $\mathrm{Ran}(K) \subset B$ is finite dimensional.*

The following result is a simple consequence of Corollaries 11.13 and 11.14.

Corollary 14.40. *If $K : H \to B$ is a finite rank operator, then K is compact. In particular if either $\dim(H) < \infty$ or $\dim(B) < \infty$ then any bounded operator $K : H \to B$ is finite rank and hence compact.*

Lemma 14.41. *Let $\mathcal{K} := \mathcal{K}(H, B)$ denote the compact operators from H to B. Then $\mathcal{K}(H, B)$ is a norm closed subspace of $L(H, B)$.*

Proof. The fact that $\mathcal{K}$ is a vector subspace of $L(H, B)$ will be left to the reader. To finish the proof, we must show that $K \in L(H, B)$ is compact if there exists $K_n \in \mathcal{K}(H, B)$ such that $\lim_{n\to\infty} \|K_n - K\|_{op} = 0$.

First Proof. Given $\varepsilon > 0$, choose $N = N(\varepsilon)$ such that $\|K_N - K\| < \varepsilon$. Using the fact that $K_N U$ is precompact, choose a finite subset $\Lambda \subset U$ such that $\min_{x\in\Lambda} \|y - K_N x\| < \varepsilon$ for all $y \in K_N(U)$. Then for $z = Kx_0 \in K(U)$ and $x \in \Lambda$,

$$\begin{aligned}\|z - Kx\| &= \|(K - K_N)x_0 + K_N(x_0 - x) + (K_N - K)x\| \\ &\le 2\varepsilon + \|K_N x_0 - K_N x\|.\end{aligned}$$

Therefore $\min_{x\in\Lambda} \|z - K_N x\| < 3\varepsilon$, which shows $K(U)$ is 3ε bounded for all $\varepsilon > 0$, $K(U)$ is totally bounded and hence precompact.

Second Proof. Suppose $\{x_n\}_{n=1}^{\infty}$ is a bounded sequence in H. By compactness, there is a subsequence $\left\{x_n^1\right\}_{n=1}^{\infty}$ of $\{x_n\}_{n=1}^{\infty}$ such that $\left\{K_1 x_n^1\right\}_{n=1}^{\infty}$ is convergent in B. Working inductively, we may construct subsequences

$$\{x_n\}_{n=1}^{\infty} \supset \left\{x_n^1\right\}_{n=1}^{\infty} \supset \left\{x_n^2\right\}_{n=1}^{\infty} \cdots \supset \{x_n^m\}_{n=1}^{\infty} \supset \dots$$

such that $\{K_m x_n^m\}_{n=1}^{\infty}$ is convergent in B for each m. By the usual Cantor's diagonalization procedure, let $y_n := x_n^n$, then $\{y_n\}_{n=1}^{\infty}$ is a subsequence of $\{x_n\}_{n=1}^{\infty}$ such that $\{K_m y_n\}_{n=1}^{\infty}$ is convergent for all m. Since

$$\begin{aligned}\|Ky_n - Ky_l\| &\le \|(K - K_m)\, y_n\| + \|K_m(y_n - y_l)\| + \|(K_m - K)\, y_l)\| \\ &\le 2\,\|K - K_m\| + \|K_m(y_n - y_l)\|\,,\end{aligned}$$

$$\lim \sup_{n,l\to\infty} \|Ky_n - Ky_l\| \le 2\,\|K - K_m\| \to 0 \text{ as } m \to \infty,$$

which shows $\{Ky_n\}_{n=1}^{\infty}$ is Cauchy and hence convergent. ∎

Proposition 14.42. *A bounded operator $K : H \to B$ is compact iff there exists finite rank operators, $K_n : H \to B$, such that $\|K - K_n\| \to 0$ as $n \to \infty$.*

Proof. Since $\overline{K(U)}$ is compact it contains a countable dense subset and from this it follows that $\overline{K(H)}$ is a separable subspace of B. Let $\{\phi_n\}$ be an orthonormal basis for $\overline{K(H)} \subset B$ and

$$P_N y = \sum_{n=1}^{N} \langle y|\phi_n\rangle \phi_n$$

be the orthogonal projection of y onto $\operatorname{span}\{\phi_n\}_{n=1}^N$. Then $\lim_{N\to\infty} \|P_N y - y\| = 0$ for all $y \in K(H)$. Define $K_n := P_n K$ – a finite rank operator on H. For sake of contradiction suppose that

$$\limsup_{n\to\infty} \|K - K_n\| = \varepsilon > 0,$$

in which case there exists $x_{n_k} \in U$ such that $\|(K - K_{n_k})x_{n_k}\| \geq \varepsilon$ for all n_k. Since K is compact, by passing to a subsequence if necessary, we may assume $\{Kx_{n_k}\}_{n_k=1}^\infty$ is convergent in B. Letting $y := \lim_{k\to\infty} Kx_{n_k}$,

$$\begin{aligned}
\|(K - K_{n_k})x_{n_k}\| &= \|(1 - P_{n_k})Kx_{n_k}\| \\
&\leq \|(1 - P_{n_k})(Kx_{n_k} - y)\| + \|(1 - P_{n_k})y\| \\
&\leq \|Kx_{n_k} - y\| + \|(1 - P_{n_k})y\| \to 0 \text{ as } k \to \infty.
\end{aligned}$$

But this contradicts the assumption that ε is positive and hence we must have $\lim_{n\to\infty} \|K - K_n\| = 0$, i.e. K is an operator norm limit of finite rank operators. The converse direction follows from Corollary 14.40 and Lemma 14.41. ∎

Corollary 14.43. *If K is compact then so is K^*.*

Proof. First Proof. Let $K_n = P_n K$ be as in the proof of Proposition 14.42, then $K_n^* = K^* P_n$ is still finite rank. Furthermore, using Proposition 14.16,

$$\|K^* - K_n^*\| = \|K - K_n\| \to 0 \text{ as } n \to \infty$$

showing K^* is a limit of finite rank operators and hence compact.

Second Proof. Let $\{x_n\}_{n=1}^\infty$ be a bounded sequence in B, then

$$\|K^* x_n - K^* x_m\|^2 = (x_n - x_m, KK^*(x_n - x_m)) \leq 2C \|KK^*(x_n - x_m)\| \tag{14.26}$$

where C is a bound on the norms of the x_n. Since $\{K^* x_n\}_{n=1}^\infty$ is also a bounded sequence, by the compactness of K there is a subsequence $\{x_n'\}$ of the $\{x_n\}$ such that $KK^* x_n'$ is convergent and hence by Eq. (14.26), so is the sequence $\{K^* x_n'\}$. ∎

14.3.1 The Spectral Theorem for Self Adjoint Compact Operators

For the rest of this section, $K \in \mathcal{K}(H) := \mathcal{K}(H,H)$ will be a self-adjoint compact operator or **S.A.C.O.** for short. Because of Proposition 14.42, we might expect compact operators to behave very much like finite dimensional matrices. This is typically the case as we will see below.

Example 14.44 (Model S.A.C.O.). Let $H = \ell_2$ and K be the diagonal matrix

$$K = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots \\ 0 & \lambda_2 & 0 & \cdots \\ 0 & 0 & \lambda_3 & \cdots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix},$$

where $\lim_{n\to\infty} |\lambda_n| = 0$ and $\lambda_n \in \mathbb{R}$. Then K is a self-adjoint compact operator. This assertion was proved in Example 11.17 above.

The main theorem (Theorem 14.46) of this subsection states that up to unitary equivalence, Example 14.44 is essentially the most general example of an S.A.C.O.

Theorem 14.45. *Let K be a S.A.C.O., then either $\lambda = \|K\|$ or $\lambda = -\|K\|$ is an eigenvalue of K.*

Proof. Without loss of generality we may assume that K is non-zero since otherwise the result is trivial. By Theorem 14.37, there exists $f_n \in H$ such that $\|f_n\| = 1$ and

$$\frac{|\langle f_n|Kf_n\rangle|}{\|f_n\|^2} = |\langle f_n|Kf_n\rangle| \longrightarrow \|K\| \text{ as } n\to\infty. \tag{14.27}$$

By passing to a subsequence if necessary, we may assume that $\lambda := \lim_{n\to\infty}\langle f_n|Kf_n\rangle$ exists and $\lambda \in \{\pm\|K\|\}$. By passing to a further subsequence if necessary, we may assume, using the compactness of K, that Kf_n is convergent as well. We now compute:

$$\begin{aligned} 0 \le \|Kf_n - \lambda f_n\|^2 &= \|Kf_n\|^2 - 2\lambda\langle Kf_n|f_n\rangle + \lambda^2 \\ &\le \lambda^2 - 2\lambda\langle Kf_n|f_n\rangle + \lambda^2 \\ &\to \lambda^2 - 2\lambda^2 + \lambda^2 = 0 \text{ as } n\to\infty. \end{aligned}$$

Hence

$$Kf_n - \lambda f_n \to 0 \text{ as } n\to\infty \tag{14.28}$$

and therefore

$$f := \lim_{n\to\infty} f_n = \frac{1}{\lambda}\lim_{n\to\infty} Kf_n$$

exists. By the continuity of the inner product, $\|f\| = 1 \neq 0$. By passing to the limit in Eq. (14.28) we find that $Kf = \lambda f$. ∎

Theorem 14.46 (Compact Operator Spectral Theorem). *Suppose that $K : H \to H$ is a non-zero S.A.C.O., then*

1. *there exists at least one eigenvalue $\lambda \in \{\pm\|K\|\}$.*
2. *There are at most countable many* ***non-zero*** *eigenvalues, $\{\lambda_n\}_{n=1}^N$, where $N = \infty$ is allowed. (Unless K is finite rank (i.e. $\dim \mathrm{Ran}\,(K) < \infty$), N will be infinite.)*
3. *The λ_n's (including multiplicities) may be arranged so that $|\lambda_n| \geq |\lambda_{n+1}|$ for all n. If $N = \infty$ then $\lim_{n\to\infty} |\lambda_n| = 0$. (In particular any eigenspace for K with* ***non-zero*** *eigenvalue is finite dimensional.)*
4. *The eigenvectors $\{\phi_n\}_{n=1}^N$ can be chosen to be an O.N. set such that $H = \overline{span\{\phi_n\}} \oplus \mathrm{Nul}(K)$.*
5. *Using the $\{\phi_n\}_{n=1}^N$ above,*
$$K\psi = \sum_{n=1}^N \lambda_n \langle\psi|\phi_n\rangle\phi_n \text{ for all } \psi \in H.$$
6. *The spectrum of K is, $\sigma(K) = \{0\} \cup \{\lambda_n : n < N+1\}$.*

Proof. We will find λ_n's and ϕ_n's recursively. Let $\lambda_1 \in \{\pm\|K\|\}$ and $\phi_1 \in H$ such that $K\phi_1 = \lambda_1\phi_1$ as in Theorem 14.45. Take $M_1 = \mathrm{span}(\phi_1)$ so $K(M_1) \subset M_1$. By Lemma 14.38, $KM_1^\perp \subset M_1^\perp$. Define $K_1 : M_1^\perp \to M_1^\perp$ via $K_1 = K|_{M_1^\perp}$. Then K_1 is again a compact operator. If $K_1 = 0$, we are done. If $K_1 \neq 0$, by Theorem 14.45 there exists $\lambda_2 \in \{\pm\|K\|_1\}$ and $\phi_2 \in M_1^\perp$ such that $\|\phi_2\| = 1$ and $K_1\phi_2 = K\phi_2 = \lambda_2\phi_2$. Let $M_2 := \mathrm{span}(\phi_1, \phi_2)$. Again $K(M_2) \subset M_2$ and hence $K_2 := K|_{M_2^\perp} : M_2^\perp \to M_2^\perp$ is compact. Again if $K_2 = 0$ we are done. If $K_2 \neq 0$. Then by Theorem 14.45 there exists $\lambda_3 \in \{\pm\|K\|_2\}$ and $\phi_3 \in M_2^\perp$ such that $\|\phi_3\| = 1$ and $K_2\phi_3 = K\phi_3 = \lambda_3\phi_3$. Continuing this way indefinitely or until we reach a point where $K_n = 0$, we construct a sequence $\{\lambda_n\}_{n=1}^N$ of eigenvalues and orthonormal eigenvectors $\{\phi_n\}_{n=1}^N$ such that $|\lambda_i| \geq |\lambda_{i+1}|$ with the further property that

$$|\lambda_i| = \sup_{\phi\perp\{\phi_1,\phi_2,\ldots\phi_{i-1}\}} \frac{\|K\phi\|}{\|\phi\|} \tag{14.29}$$

If $N = \infty$ then $\lim_{i\to\infty} |\lambda_i| = 0$ for if not there would exist $\varepsilon > 0$ such that $|\lambda_i| \geq \varepsilon > 0$ for all i. In this case $\{\phi_i/\lambda_i\}_{i=1}^\infty$ is sequence in H bounded by ε^{-1}. By compactness of K, there exists a subsequence i_k such that $\phi_{i_k} = K\phi_{i_k}/\lambda_{i_k}$ is convergent. But this is impossible since $\{\phi_{i_k}\}$ is an orthonormal set. Hence we must have that $\varepsilon = 0$. Let $M := \mathrm{span}\{\phi_i\}_{i=1}^N$ with $N = \infty$ **possible**. Then $K(M) \subset M$ and hence $K(M^\perp) \subset M^\perp$. Using Eq. (14.29),

$$\|K|_{M^\perp}\| \leq \|K|_{M_n^\perp}\| = |\lambda_n| \longrightarrow 0 \text{ as } n \to \infty$$

showing $K|M^\perp \equiv 0$. Define P_0 to be orthogonal projection onto $M^\perp$. Then for $\psi \in H$,

$$\psi = P_0\psi + (1 - P_0)\psi = P_0\psi + \sum_{i=1}^{N} \langle\psi|\phi_i\rangle\phi_i$$

and

$$K\psi = KP_0\psi + K\sum_{i=1}^{N} \langle\psi|\phi_i\rangle\phi_i = \sum_{i=1}^{N} \lambda_i\langle\psi|\phi_i\rangle\phi_i.$$

Since $\{\lambda_n\} \subset \sigma(K)$ and $\sigma(K)$ is closed, it follows that $0 \in \sigma(K)$ and hence $\{\lambda_n\}_{n=1}^{\infty} \cup \{0\} \subset \sigma(K)$. Suppose that $z \notin \{\lambda_n\}_{n=1}^{\infty} \cup \{0\}$ and let d be the distance between z and $\{\lambda_n\}_{n=1}^{\infty} \cup \{0\}$. Notice that $d > 0$ because $\lim_{n\to\infty} \lambda_n = 0$. A few simple computations show that:

$$(K - zI)\psi = \sum_{i=1}^{N} \langle\psi|\phi_i\rangle(\lambda_i - z)\phi_i - zP_0\psi,$$

$(K - z)^{-1}$ exists,

$$(K - zI)^{-1}\psi = \sum_{i=1}^{N} \langle\psi|\phi_i\rangle(\lambda_i - z)^{-1}\phi_i - z^{-1}P_0\psi,$$

and

$$\begin{aligned}\|(K - zI)^{-1}\psi\|^2 &= \sum_{i=1}^{N} |\langle\psi|\phi_i\rangle|^2 \frac{1}{|\lambda_i - z|^2} + \frac{1}{|z|^2}\|P_0\psi\|^2 \\ &\le \left(\frac{1}{d}\right)^2 \left(\sum_{i=1}^{N} |\langle\psi|\phi_i\rangle|^2 + \|P_0\psi\|^2\right) = \frac{1}{d^2}\|\psi\|^2.\end{aligned}$$

We have thus shown that $(K - zI)^{-1}$ exists, $\|(K - zI)^{-1}\| \le d^{-1} < \infty$ and hence $z \notin \sigma(K)$. ∎

Theorem 14.47 (Structure of Compact Operators). *Let $K : H \to B$ be a compact operator. Then there exists $N \in \mathbb{N} \cup \{\infty\}$, orthonormal subsets $\{\phi_n\}_{n=1}^{N} \subset H$ and $\{\psi_n\}_{n=1}^{N} \subset B$ and a sequences $\{\alpha_n\}_{n=1}^{N} \subset \mathbb{R}_+$ such that $\lambda_1 \ge \lambda_2 \ge \dots$, $\lim_{n\to\infty} \alpha_n = 0$ if $N = \infty$, $\|\psi_n\| \le 1$ for all n and*

$$Kf = \sum_{n=1}^{N} \alpha_n\langle f|\phi_n\rangle\psi_n \text{ for all } f \in H. \tag{14.30}$$

Proof. Since K^*K is a selfadjoint compact operator, Theorem 14.46 implies there exists an orthonormal set $\{\phi_n\}_{n=1}^{N} \subset H$ and positive numbers $\{\lambda_n\}_{n=1}^{N}$ such that

$$K^*K\psi = \sum_{n=1}^{N} \lambda_n\langle\psi|\phi_n\rangle\phi_n \text{ for all } \psi \in H.$$

Let A be the positive square root of K^*K defined by

$$A\psi := \sum_{n=1}^{N} \sqrt{\lambda_n} \langle \psi | \phi_n \rangle \phi_n \text{ for all } \psi \in H.$$

A simple computation shows, $A^2 = K^*K$, and therefore,

$$\begin{aligned} \|A\psi\|^2 &= \langle A\psi | A\psi \rangle = \langle \psi | A^2 \psi \rangle \\ &= \langle \psi | K^*K\psi \rangle = \langle K\psi | K\psi \rangle = \|K\psi\|^2 \end{aligned}$$

for all $\psi \in H$. Hence we may define a unitary operator, $u : \overline{\mathrm{Ran}(A)} \to \overline{\mathrm{Ran}(K)}$ by the formula

$$uA\psi = K\psi \text{ for all } \psi \in H.$$

We then have

$$K\psi = uA\psi = \sum_{n=1}^{N} \sqrt{\lambda_n} \langle \psi | \phi_n \rangle u\phi_n \tag{14.31}$$

which proves the result with $\psi_n := u\phi_n$ and $\alpha_n = \sqrt{\lambda_n}$.

It is instructive to find ψ_n explicitly and to verify Eq. (14.31) by bruit force. Since $\phi_n = \lambda_n^{-1/2} A\phi_n$,

$$\psi_n = \lambda_n^{-1/2} uA\phi_n = \lambda_n^{-1/2} uA\phi_n = \lambda_n^{-1/2} K\phi_n$$

and

$$\langle K\phi_n | K\phi_m \rangle = \langle \phi_n | K^*K\phi_m \rangle = \lambda_n \delta_{mn}.$$

This verifies that $\{\psi_n\}_{n=1}^N$ is an orthonormal set. Moreover,

$$\begin{aligned} \sum_{n=1}^{N} \sqrt{\lambda_n} \langle \psi | \phi_n \rangle \psi_n &= \sum_{n=1}^{N} \sqrt{\lambda_n} \langle \psi | \phi_n \rangle \lambda_n^{-1/2} K\phi_n \\ &= K \sum_{n=1}^{N} \langle \psi | \phi_n \rangle \phi_n = K\psi \end{aligned}$$

since $\sum_{n=1}^N \langle \psi | \phi_n \rangle \phi_n = P\psi$ where P is orthogonal projection onto $\mathrm{Nul}(K)^\perp$.

Second Proof. Let $K = u\,|K|$ be the polar decomposition of K. Then $|K|$ is self-adjoint and compact, by Corollary **??**, and hence by Theorem 14.46 there exists an orthonormal basis $\{\phi_n\}_{n=1}^N$ for $\mathrm{Nul}(|K|)^\perp = \mathrm{Nul}(K)^\perp$ such that $|K|\,\phi_n = \lambda_n \phi_n$, $\lambda_1 \geq \lambda_2 \geq \dots$ and $\lim_{n\to\infty} \lambda_n = 0$ if $N = \infty$. For $f \in H$,

$$Kf = u\,|K| \sum_{n=1}^{N} \langle f | \phi_n \rangle \phi_n = \sum_{n=1}^{N} \langle f | \phi_n \rangle u\,|K|\,\phi_n = \sum_{n=1}^{N} \lambda_n \langle f | \phi_n \rangle u\phi_n$$

which is Eq. (14.30) with $\psi_n := u\phi_n$. ∎

14.4 Weak Convergence

Suppose H is an infinite dimensional Hilbert space and $\{x_n\}_{n=1}^{\infty}$ is an orthonormal subset of H. Then, by Eq. (14.1), $\|x_n - x_m\|^2 = 2$ for all $m \neq n$ and in particular, $\{x_n\}_{n=1}^{\infty}$ has no convergent subsequences. From this we conclude that $C := \{x \in H : \|x\| \leq 1\}$, the closed unit ball in H, is not compact. To overcome this problems it is sometimes useful to introduce a weaker topology on X having the property that C is compact.

Definition 14.48. *Let $(X, \|\cdot\|)$ be a Banach space and X^* be its continuous dual. The weak topology, τ_w, on X is the topology generated by X^*. If $\{x_n\}_{n=1}^{\infty} \subset X$ is a sequence we will write $x_n \xrightarrow{w} x$ as $n \to \infty$ to mean that $x_n \to x$ in the weak topology.*

Because $\tau_w = \tau(X^*) \subset \tau_{\|\cdot\|} := \tau(\{\|x - \cdot\| : x \in X\}$, it is harder for a function $f : X \to \mathbb{F}$ to be continuous in the τ_w – topology than in the norm topology, $\tau_{\|\cdot\|}$. In particular if $\phi : X \to \mathbb{F}$ is a linear functional which is τ_w – continuous, then ϕ is $\tau_{\|\cdot\|}$ – continuous and hence $\phi \in X^*$.

Exercise 14.12. Show the vector space operations of X are continuous in the weak topology, i.e. show:

1. $(x, y) \in X \times X \to x + y \in X$ is $(\tau_w \otimes \tau_w, \tau_w)$ – continuous and
2. $(\lambda, x) \in \mathbb{F} \times X \to \lambda x \in X$ is $(\tau_{\mathbb{F}} \otimes \tau_w, \tau_w)$ – continuous.

Proposition 14.49. *Let $\{x_n\}_{n=1}^{\infty} \subset X$ be a sequence, then $x_n \xrightarrow{w} x \in X$ as $n \to \infty$ iff $\phi(x) = \lim_{n\to\infty} \phi(x_n)$ for all $\phi \in X^*$.*

Proof. By definition of τ_w, we have $x_n \xrightarrow{w} x \in X$ iff for all $\Gamma \subset\subset X^*$ and $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that $|\phi(x) - \phi(x_n)| < \varepsilon$ for all $n \geq N$ and $\phi \in \Gamma$. This later condition is easily seen to be equivalent to $\phi(x) = \lim_{n\to\infty} \phi(x_n)$ for all $\phi \in X^*$. ■

The topological space (X, τ_w) is still Hausdorff as follows from the Hahn Banach Theorem, see Theorem **??**. For the moment we will concentrate on the special case where $X = H$ is a Hilbert space in which case $H^* = \{\phi_z := \langle \cdot | z \rangle : z \in H\}$, see Theorem 14.15. If $x, y \in H$ and $z := y - x \neq 0$, then

$$0 < \varepsilon := \|z\|^2 = \phi_z(z) = \phi_z(y) - \phi_z(x).$$

Thus

$$V_x := \{w \in H : |\phi_z(x) - \phi_z(w)| < \varepsilon/2\} \text{ and}$$
$$V_y := \{w \in H : |\phi_z(y) - \phi_z(w)| < \varepsilon/2\}$$

are disjoint sets from τ_w which contain x and y respectively. This shows that (H, τ_w) is a Hausdorff space. In particular, this shows that weak limits are unique if they exist.

Remark 14.50. Suppose that H is an infinite dimensional Hilbert space $\{x_n\}_{n=1}^\infty$ is an orthonormal subset of H. Then Bessel's inequality (Proposition 14.18) implies $x_n \overset{w}{\to} 0 \in H$ as $n \to \infty$. This points out the fact that if $x_n \overset{w}{\to} x \in H$ as $n \to \infty$, it is no longer necessarily true that $\|x\| = \lim_{n\to\infty} \|x_n\|$. However we do always have $\|x\| \le \liminf_{n\to\infty} \|x_n\|$ because,

$$\|x\|^2 = \lim_{n\to\infty} \langle x_n|x\rangle \le \liminf_{n\to\infty} [\|x_n\| \|x\|] = \|x\| \liminf_{n\to\infty} \|x_n\|.$$

Proposition 14.51. *Let H be a Hilbert space, $\beta \subset H$ be an orthonormal basis for H and $\{x_n\}_{n=1}^\infty \subset H$ be a bounded sequence, then the following are equivalent:*

1. $x_n \overset{w}{\to} x \in H$ *as* $n \to \infty$.
2. $\langle x|y\rangle = \lim_{n\to\infty} \langle x_n|y\rangle$ *for all* $y \in H$.
3. $\langle x|y\rangle = \lim_{n\to\infty} \langle x_n|y\rangle$ *for all* $y \in \beta$.

Moreover, if $c_y := \lim_{n\to\infty} \langle x_n|y\rangle$ *exists for all* $y \in \beta$, *then* $\sum_{y\in\beta} |c_y|^2 < \infty$ *and* $x_n \overset{w}{\to} x := \sum_{y\in\beta} c_y y \in H$ *as* $n \to \infty$.

Proof. 1. $\Longrightarrow$ 2. This is a consequence of Theorem 14.15 and Proposition 14.49. 2. $\Longrightarrow$ 3. is trivial. 3. $\Longrightarrow$ 1. Let $M := \sup_n \|x_n\|$ and H_0 denote the algebraic span of β. Then for $y \in H$ and $z \in H_0$,

$$|\langle x - x_n|y\rangle| \le |\langle x - x_n|z\rangle| + |\langle x - x_n|y - z\rangle| \le |\langle x - x_n|z\rangle| + 2M \|y - z\|.$$

Passing to the limit in this equation implies $\limsup_{n\to\infty} |\langle x - x_n|y\rangle| \le 2M \|y - z\|$ which shows $\limsup_{n\to\infty} |\langle x - x_n|y\rangle| = 0$ since H_0 is dense in H. To prove the last assertion, let $\Gamma \subset\subset \beta$. Then by Bessel's inequality (Proposition 14.18),

$$\sum_{y\in\Gamma} |c_y|^2 = \lim_{n\to\infty} \sum_{y\in\Gamma} |\langle x_n|y\rangle|^2 \le \liminf_{n\to\infty} \|x_n\|^2 \le M^2.$$

Since $\Gamma \subset\subset \beta$ was arbitrary, we conclude that $\sum_{y\in\beta} |c_y|^2 \le M < \infty$ and hence we may define $x := \sum_{y\in\beta} c_y y$. By construction we have

$$\langle x|y\rangle = c_y = \lim_{n\to\infty} \langle x_n|y\rangle \text{ for all } y \in \beta$$

and hence $x_n \overset{w}{\to} x \in H$ as $n \to \infty$ by what we have just proved. ∎

Theorem 14.52. *Suppose $\{x_n\}_{n=1}^\infty$ is a bounded sequence in a Hilbert space, H. Then there exists a subsequence $y_k := x_{n_k}$ of $\{x_n\}_{n=1}^\infty$ and $x \in X$ such that $y_k \overset{w}{\to} x$ as $k \to \infty$.*

Proof. This is a consequence of Proposition 14.51 and a Cantor's diagonalization argument which is left to the reader, see Exercise 14.13. ∎

Theorem 14.53 (Alaoglu's Theorem for Hilbert Spaces). *Suppose that H is a separable Hilbert space, $C := \{x \in H : \|x\| \leq 1\}$ is the closed unit ball in H and $\{e_n\}_{n=1}^{\infty}$ is an orthonormal basis for H. Then*

$$\rho(x, y) := \sum_{n=1}^{\infty} \frac{1}{2^n} |\langle x - y | e_n \rangle| \tag{14.32}$$

*defines a metric on C which is compatible with the weak topology on C, $\tau_C := (\tau_w)_C = \{V \cap C : V \in \tau_w\}$. Moreover (C, ρ) is a compact metric space. (This theorem will be extended to Banach spaces, see Theorems **??** and **??** below.)*

Proof. The routine check that ρ is a metric is left to the reader. Let τ_ρ be the topology on C induced by ρ. For any $y \in H$ and $n \in \mathbb{N}$, the map $x \in H \to \langle x - y | e_n \rangle = \langle x | e_n \rangle - \langle y | e_n \rangle$ is τ_w continuous and since the sum in Eq. (14.32) is uniformly convergent for $x, y \in C$, it follows that $x \to \rho(x, y)$ is τ_C – continuous. This implies the open balls relative to ρ are contained in τ_C and therefore $\tau_\rho \subset \tau_C$. For the converse inclusion, let $z \in H$, $x \to \phi_z(x) = \langle x | z \rangle$ be an element of H^*, and for $N \in \mathbb{N}$ let $z_N := \sum_{n=1}^{N} \langle z | e_n \rangle e_n$. Then $\phi_{z_N} = \sum_{n=1}^{N} \overline{\langle z | e_n \rangle} \phi_{e_n}$ is ρ continuous, being a finite linear combination of the ϕ_{e_n} which are easily seen to be ρ – continuous. Because $z_N \to z$ as $N \to \infty$ it follows that

$$\sup_{x \in C} |\phi_z(x) - \phi_{z_N}(x)| = \|z - z_N\| \to 0 \text{ as } N \to \infty.$$

Therefore $\phi_z|_C$ is ρ – continuous as well and hence $\tau_C = \tau(\phi_z|_C : z \in H) \subset \tau_\rho$. The last assertion follows directly from Theorem 14.52 and the fact that sequential compactness is equivalent to compactness for metric spaces. ∎

14.5 Supplement 1: Converse of the Parallelogram Law

Proposition 14.54 (Parallelogram Law Converse). *If $(X, \|\cdot\|)$ is a normed space such that Eq. (14.2) holds for all $x, y \in X$, then there exists a unique inner product on $\langle \cdot | \cdot \rangle$ such that $\|x\| := \sqrt{\langle x | x \rangle}$ for all $x \in X$. In this case we say that $\|\cdot\|$ is a Hilbertian norm.*

Proof. If $\|\cdot\|$ is going to come from an inner product $\langle \cdot | \cdot \rangle$, it follows from Eq. (14.1) that

$$2\mathrm{Re}\langle x | y \rangle = \|x + y\|^2 - \|x\|^2 - \|y\|^2$$

and

$$-2\mathrm{Re}\langle x | y \rangle = \|x - y\|^2 - \|x\|^2 - \|y\|^2.$$

Subtracting these two equations gives the "polarization identity,"

$$4\mathrm{Re}\langle x | y \rangle = \|x + y\|^2 - \|x - y\|^2.$$

Replacing y by iy in this equation then implies that

$$4\mathrm{Im}\langle x|y\rangle = \|x+iy\|^2 - \|x-iy\|^2$$

from which we find

$$\langle x|y\rangle = \frac{1}{4}\sum_{\varepsilon\in G} \varepsilon\|x+\varepsilon y\|^2 \tag{14.33}$$

where $G = \{\pm 1, \pm i\}$ – a cyclic subgroup of $S^1 \subset \mathbb{C}$. Hence if $\langle\cdot|\cdot\rangle$ is going to exists we must define it by Eq. (14.33). Notice that

$$\begin{aligned}\langle x|x\rangle &= \frac{1}{4}\sum_{\varepsilon\in G} \varepsilon\|x+\varepsilon x\|^2 = \|x\|^2 + i\|x+ix\|^2 - i\|x-ix\|^2 \\ &= \|x\|^2 + i\left|1+i\right|^2\left|\|x\|^2 - i\left|1-i\right|^2\right|\|x\|^2 = \|x\|^2.\end{aligned}$$

So to finish the proof of (4) we must show that $\langle x|y\rangle$ in Eq. (14.33) is an inner product. Since

$$\begin{aligned}4\langle y|x\rangle &= \sum_{\varepsilon\in G} \varepsilon\|y+\varepsilon x\|^2 = \sum_{\varepsilon\in G} \varepsilon\|\varepsilon\left(y+\varepsilon x\right)\|^2 \\ &= \sum_{\varepsilon\in G} \varepsilon\|\varepsilon y + \varepsilon^2 x\|^2 \\ &= \|y+x\|^2 + \|-y+x\|^2 + i\|iy-x\|^2 - i\|-iy-x\|^2 \\ &= \|x+y\|^2 + \|x-y\|^2 + i\|x-iy\|^2 - i\|x+iy\|^2 \\ &= 4\overline{\langle x|y\rangle}\end{aligned}$$

it suffices to show $x \to \langle x|y\rangle$ is linear for all $y \in H$. (The rest of this proof may safely be skipped by the reader.) For this we will need to derive an identity from Eq. (14.2). To do this we make use of Eq. (14.2) three times to find

$$\begin{aligned}\|x+y+z\|^2 &= -\|x+y-z\|^2 + 2\|x+y\|^2 + 2\|z\|^2 \\ &= \|x-y-z\|^2 - 2\|x-z\|^2 - 2\|y\|^2 + 2\|x+y\|^2 + 2\|z\|^2 \\ &= \|y+z-x\|^2 - 2\|x-z\|^2 - 2\|y\|^2 + 2\|x+y\|^2 + 2\|z\|^2 \\ &= -\|y+z+x\|^2 + 2\|y+z\|^2 + 2\|x\|^2 \\ &\qquad - 2\|x-z\|^2 - 2\|y\|^2 + 2\|x+y\|^2 + 2\|z\|^2.\end{aligned}$$

Solving this equation for $\|x+y+z\|^2$ gives

$$\|x+y+z\|^2 = \|y+z\|^2 + \|x+y\|^2 - \|x-z\|^2 + \|x\|^2 + \|z\|^2 - \|y\|^2. \tag{14.34}$$

Using Eq. (14.34), for $x, y, z \in H$,

$$\begin{aligned}4\,\mathrm{Re}\langle x+z|y\rangle &= \|x+z+y\|^2 - \|x+z-y\|^2 \\ &= \|y+z\|^2 + \|x+y\|^2 - \|x-z\|^2 + \|x\|^2 + \|z\|^2 - \|y\|^2 \\ &\quad - \left(\|z-y\|^2 + \|x-y\|^2 - \|x-z\|^2 + \|x\|^2 + \|z\|^2 - \|y\|^2\right) \\ &= \|z+y\|^2 - \|z-y\|^2 + \|x+y\|^2 - \|x-y\|^2 \\ &= 4\,\mathrm{Re}\langle x|y\rangle + 4\,\mathrm{Re}\langle z|y\rangle.\end{aligned} \tag{14.35}$$

Now suppose that $\delta \in G$, then since $|\delta| = 1$,

$$\begin{aligned} 4\langle \delta x|y\rangle &= \frac{1}{4}\sum_{\varepsilon\in G} \varepsilon\|\delta x + \varepsilon y\|^2 = \frac{1}{4}\sum_{\varepsilon\in G} \varepsilon\|x + \delta^{-1}\varepsilon y\|^2 \\ &= \frac{1}{4}\sum_{\varepsilon\in G} \varepsilon\delta\|x + \delta\varepsilon y\|^2 = 4\delta\langle x|y\rangle \end{aligned} \tag{14.36}$$

where in the third inequality, the substitution $\varepsilon \to \varepsilon\delta$ was made in the sum. So Eq. (14.36) says $\langle \pm ix|y\rangle = \pm i\langle ix|y\rangle$ and $\langle -x|y\rangle = -\langle x|y\rangle$. Therefore

$$\mathrm{Im}\langle x|y\rangle = \mathrm{Re}\,(-i\langle x|y\rangle) = \mathrm{Re}\langle -ix|y\rangle$$

which combined with Eq. (14.35) shows

$$\begin{aligned} \mathrm{Im}\langle x+z|y\rangle &= \mathrm{Re}\langle -ix - iz|y\rangle = \mathrm{Re}\langle -ix|y\rangle + \mathrm{Re}\langle -iz|y\rangle \\ &= \mathrm{Im}\langle x|y\rangle + \mathrm{Im}\langle z|y\rangle \end{aligned}$$

and therefore (again in combination with Eq. (14.35)),

$$\langle x+z|y\rangle = \langle x|y\rangle + \langle z|y\rangle \text{ for all } x, y \in H.$$

Because of this equation and Eq. (14.36) to finish the proof that $x \to \langle x|y\rangle$ is linear, it suffices to show $\langle \lambda x|y\rangle = \lambda\langle x|y\rangle$ for all $\lambda > 0$. Now if $\lambda = m \in \mathbb{N}$, then

$$\langle mx|y\rangle = \langle x + (m-1)x|y\rangle = \langle x|y\rangle + \langle (m-1)x|y\rangle$$

so that by induction $\langle mx|y\rangle = m\langle x|y\rangle$. Replacing x by x/m then shows that $\langle x|y\rangle = m\langle m^{-1}x|y\rangle$ so that $\langle m^{-1}x|y\rangle = m^{-1}\langle x|y\rangle$ and so if $m, n \in \mathbb{N}$, we find

$$\langle \frac{n}{m}x|y\rangle = n\langle \frac{1}{m}x|y\rangle = \frac{n}{m}\langle x|y\rangle$$

so that $\langle \lambda x|y\rangle = \lambda\langle x|y\rangle$ for all $\lambda > 0$ and $\lambda \in \mathbb{Q}$. By continuity, it now follows that $\langle \lambda x|y\rangle = \lambda\langle x|y\rangle$ for all $\lambda > 0$. ∎

14.6 Supplement 2. Non-complete inner product spaces

Part of Theorem 14.24 goes through when H is a not necessarily complete inner product space. We have the following proposition.

Proposition 14.55. *Let $(H, \langle\cdot|\cdot\rangle)$ be a not necessarily complete inner product space and $\beta \subset H$ be an orthonormal set. Then the following two conditions are equivalent:*

1. $x = \sum_{u\in\beta} \langle x|u\rangle u$ *for all* $x \in H$.

2. $\|x\|^2 = \sum_{u\in\beta} |\langle x|u\rangle|^2$ *for all* $x \in H$.

Moreover, either of these two conditions implies that $\beta \subset H$ *is a maximal orthonormal set. However* $\beta \subset H$ *being a maximal orthonormal set is not sufficient to conditions for 1) and 2) hold!*

Proof. As in the proof of Theorem 14.24, 1) implies 2). For 2) implies 1) let $\Lambda \subset\subset \beta$ and consider

$$\left\|x - \sum_{u\in\Lambda}\langle x|u\rangle u\right\|^2 = \|x\|^2 - 2\sum_{u\in\Lambda}|\langle x|u\rangle|^2 + \sum_{u\in\Lambda}|\langle x|u\rangle|^2$$
$$= \|x\|^2 - \sum_{u\in\Lambda}|\langle x|u\rangle|^2 .$$

Since $\|x\|^2 = \sum_{u\in\beta} |\langle x|u\rangle|^2$, it follows that for every $\varepsilon > 0$ there exists $\Lambda_\varepsilon \subset\subset \beta$ such that for all $\Lambda \subset\subset \beta$ such that $\Lambda_\varepsilon \subset \Lambda$,

$$\left\|x - \sum_{u\in\Lambda}\langle x|u\rangle u\right\|^2 = \|x\|^2 - \sum_{u\in\Lambda}|\langle x|u\rangle|^2 < \varepsilon$$

showing that $x = \sum_{u\in\beta}\langle x|u\rangle u$. Suppose $x = (x_1, x_2, \dots, x_n, \dots) \in \beta^\perp$. If 2) is valid then $\|x\|^2 = 0$, i.e. $x = 0$. So β is maximal. Let us now construct a counter example to prove the last assertion. Take $H = \mathrm{Span}\{e_i\}_{i=1}^\infty \subset \ell^2$ and let $\tilde{u}_n = e_1 - (n+1)e_{n+1}$ for $n = 1, 2 \dots$. Applying Gramn-Schmidt to $\{\tilde{u}_n\}_{n=1}^\infty$ we construct an orthonormal set $\beta = \{u_n\}_{n=1}^\infty \subset H$. I now claim that $\beta \subset H$ is maximal. Indeed if $x = (x_1, x_2, \dots, x_n, \dots) \in \beta^\perp$ then $x \perp u_n$ for all n, i.e.

$$0 = \langle x|\tilde{u}_n\rangle = x_1 - (n+1)x_{n+1}.$$

Therefore $x_{n+1} = (n+1)^{-1} x_1$ for all n. Since $x \in \mathrm{Span}\{e_i\}_{i=1}^\infty$, $x_N = 0$ for some N sufficiently large and therefore $x_1 = 0$ which in turn implies that $x_n = 0$ for all n. So $x = 0$ and hence β is maximal in H. On the other hand, β is not maximal in ℓ^2. In fact the above argument shows that $\beta^\perp$ in ℓ^2 is given by the span of $v = (1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \dots)$. Let P be the orthogonal projection of ℓ^2 onto the $\mathrm{Span}(\beta) = v^\perp$. Then

$$\sum_{i=1}^\infty \langle x|u_n\rangle u_n = Px = x - \frac{\langle x|v\rangle}{\|v\|^2}v,$$

so that $\sum_{i=1}^\infty \langle x|u_n\rangle u_n = x$ iff $x \in \mathrm{Span}(\beta) = v^\perp \subset \ell^2$. For example if $x = (1, 0, 0, \dots) \in H$ (or more generally for $x = e_i$ for any i), $x \notin v^\perp$ and hence $\sum_{i=1}^\infty \langle x|u_n\rangle u_n \neq x$. ∎

14.7 Exercises

Exercise 14.13. Prove Theorem 14.52. **Hint:** Let $H_0 := \overline{\mathrm{span}\{x_n : n \in \mathbb{N}\}}$ – a separable Hilbert subspace of H. Let $\{\lambda_m\}_{m=1}^\infty \subset H_0$ be an orthonormal basis and use Cantor's diagonalization argument to find a subsequence $y_k := x_{n_k}$ such that $c_m := \lim_{k\to\infty} \langle y_k | \lambda_m \rangle$ exists for all $m \in \mathbb{N}$. Finish the proof by appealing to Proposition 14.51.

Exercise 14.14. Suppose that $\{x_n\}_{n=1}^\infty \subset H$ and $x_n \overset{w}{\to} x \in H$ as $n \to \infty$. Show $x_n \to x$ as $n \to \infty$ (i.e. $\lim_{n\to\infty} \|x - x_n\| = 0$) iff $\lim_{n\to\infty} \|x_n\| = \|x\|$.

Exercise 14.15 (Banach-Saks). Suppose that $\{x_n\}_{n=1}^\infty \subset H$, $x_n \overset{w}{\to} x \in H$ as $n \to \infty$, and $c := \sup_n \|x_n\| < \infty$.[4] Show there exists a subsequence, $y_k = x_{n_k}$ such that

$$\lim_{N\to\infty} \left\| x - \frac{1}{N} \sum_{k=1}^N y_k \right\| = 0,$$

i.e. $\frac{1}{N}\sum_{k=1}^N y_k \to x$ as $N \to \infty$. **Hints:** 1. show it suffices to assume $x = 0$ and then choose $\{y_k\}_{k=1}^\infty$ so that $|\langle y_k | y_l \rangle| \le l^{-1}$ (or even smaller if you like) for all $k \le l$.

Exercise 14.16 (The Mean Ergodic Theorem). Let $U : H \to H$ be a unitary operator on a Hilbert space H, $M = \mathrm{Nul}(U - I)$, $P = P_M$ be orthogonal projection onto M, and $S_n = \frac{1}{n}\sum_{k=0}^{n-1} U^k$. Show $S_n \to P_M$ **strongly**, i.e. $\lim_{n\to\infty} S_n x = P_M x$ for all $x \in H$. **Hints:** 1. verify the result for $x \in \mathrm{Nul}(U - I)$ and $x \in \mathrm{Ran}(U - I)$, 2. show $\mathrm{Nul}(U^* - I) = \mathrm{Nul}(U - I)$, 3. finish the result with a limiting argument making use of items 1. and 2. and Lemma 14.17.

[4] The assumption that $c < \infty$ is superfluous because of the "uniform boundedness principle," see Theorem **??** below.

Part IV

Calculus and Ordinary Differential Equations in Banach Spaces

15

Ordinary Differential Equations in a Banach Space

Let X be a Banach space, $U \subset_o X$, $J = (a, b) \ni 0$ and $Z \in C(J \times U, X)$ – Z is to be interpreted as a time dependent vector-field on $U \subset X$. In this section we will consider the ordinary differential equation (ODE for short)

$$\dot{y}(t) = Z(t, y(t)) \text{ with } y(0) = x \in U. \tag{15.1}$$

The reader should check that any solution $y \in C^1(J, U)$ to Eq. (15.1) gives a solution $y \in C(J, U)$ to the integral equation:

$$y(t) = x + \int_0^t Z(\tau, y(\tau))d\tau \tag{15.2}$$

and conversely if $y \in C(J, U)$ solves Eq. (15.2) then $y \in C^1(J, U)$ and y solves Eq. (15.1).

Remark 15.1. For notational simplicity we have assumed that the initial condition for the ODE in Eq. (15.1) is taken at $t = 0$. There is no loss in generality in doing this since if $\tilde{y}$ solves

$$\frac{d\tilde{y}}{dt}(t) = \tilde{Z}(t, \tilde{y}(t)) \text{ with } \tilde{y}(t_0) = x \in U$$

iff $y(t) := \tilde{y}(t + t_0)$ solves Eq. (15.1) with $Z(t, x) = \tilde{Z}(t + t_0, x)$.

15.1 Examples

Let $X = \mathbb{R}$, $Z(x) = x^n$ with $n \in \mathbb{N}$ and consider the ordinary differential equation

$$\dot{y}(t) = Z(y(t)) = y^n(t) \text{ with } y(0) = x \in \mathbb{R}. \tag{15.3}$$

If y solves Eq. (15.3) with $x \neq 0$, then $y(t)$ is not zero for t near 0. Therefore up to the first time y possibly hits 0, we must have

$$t = \int_0^t \frac{\dot{y}(\tau)}{y(\tau)^n} d\tau = \int_0^{y(t)} u^{-n} du = \begin{cases} \frac{[y(t)]^{1-n} - x^{1-n}}{1-n} & \text{if } n > 1 \\ \ln \left| \frac{y(t)}{x} \right| & \text{if } n = 1 \end{cases}$$

and solving these equations for $y(t)$ implies

$$y(t) = y(t, x) = \begin{cases} \frac{x}{\sqrt[n-1]{1-(n-1)tx^{n-1}}} & \text{if } n > 1 \\ e^t x & \text{if } n = 1. \end{cases} \tag{15.4}$$

The reader should verify by direct calculation that $y(t, x)$ defined above does indeed solve Eq. (15.3). The above argument shows that these are the only possible solutions to the Equations in (15.3).

Notice that when $n = 1$, the solution exists for all time while for $n > 1$, we must require

$$1 - (n-1)tx^{n-1} > 0$$

or equivalently that

$$t < \frac{1}{(1-n)x^{n-1}} \text{ if } x^{n-1} > 0 \text{ and}$$
$$t > -\frac{1}{(1-n)\left|x\right|^{n-1}} \text{ if } x^{n-1} < 0.$$

Moreover for $n > 1$, $y(t, x)$ blows up as t approaches the value for which $1 - (n-1)tx^{n-1} = 0$. The reader should also observe that, at least for s and t close to 0,

$$y(t, y(s, x)) = y(t + s, x) \tag{15.5}$$

for each of the solutions above. Indeed, if $n = 1$ Eq. (15.5) is equivalent to the well know identity, $e^t e^s = e^{t+s}$ and for $n > 1$,

$$\begin{aligned} y(t, y(s, x)) &= \frac{y(s, x)}{\sqrt[n-1]{1 - (n-1)ty(s, x)^{n-1}}} \\ &= \frac{\frac{x}{\sqrt[n-1]{1-(n-1)sx^{n-1}}}}{\sqrt[n-1]{1 - (n-1)t\left[\frac{x}{\sqrt[n-1]{1-(n-1)sx^{n-1}}}\right]^{n-1}}} \\ &= \frac{\frac{x}{\sqrt[n-1]{1-(n-1)sx^{n-1}}}}{\sqrt[n-1]{1 - (n-1)t\frac{x^{n-1}}{1-(n-1)sx^{n-1}}}} \\ &= \frac{x}{\sqrt[n-1]{1 - (n-1)sx^{n-1} - (n-1)tx^{n-1}}} \\ &= \frac{x}{\sqrt[n-1]{1 - (n-1)(s+t)x^{n-1}}} = y(t + s, x). \end{aligned}$$

Now suppose $Z(x) = |x|^\alpha$ with $0 < \alpha < 1$ and we now consider the ordinary differential equation

$$\dot{y}(t) = Z(y(t)) = |y(t)|^\alpha \text{ with } y(0) = x \in \mathbb{R}. \tag{15.6}$$

Working as above we find, if $x \neq 0$ that

$$t = \int_0^t \frac{\dot{y}(\tau)}{|y(t)|^\alpha} d\tau = \int_0^{y(t)} |u|^{-\alpha}\, du = \frac{[y(t)]^{1-\alpha} - x^{1-\alpha}}{1-\alpha},$$

where $u^{1-\alpha} := |u|^{1-\alpha} \operatorname{sgn}(u)$. Since $\operatorname{sgn}(y(t)) = \operatorname{sgn}(x)$ the previous equation implies

$$\begin{aligned} \operatorname{sgn}(x)(1-\alpha)t &= \operatorname{sgn}(x)\left[\operatorname{sgn}(y(t))\,|y(t)|^{1-\alpha} - \operatorname{sgn}(x)\,|x|^{1-\alpha}\right] \\ &= |y(t)|^{1-\alpha} - |x|^{1-\alpha} \end{aligned}$$

and therefore,

$$y(t,x) = \operatorname{sgn}(x)\left(|x|^{1-\alpha} + \operatorname{sgn}(x)(1-\alpha)t\right)^{\frac{1}{1-\alpha}} \tag{15.7}$$

is uniquely determined by this formula until the first time t where $|x|^{1-\alpha} + \operatorname{sgn}(x)(1-\alpha)t = 0$. As before $y(t) = 0$ is a solution to Eq. (15.6), however it is far from being the unique solution. For example letting $x \downarrow 0$ in Eq. (15.7) gives a function

$$y(t, 0+) = ((1-\alpha)t)^{\frac{1}{1-\alpha}}$$

which solves Eq. (15.6) for $t > 0$. Moreover if we define

$$y(t) := \begin{cases} ((1-\alpha)t)^{\frac{1}{1-\alpha}} & \text{if } t > 0 \\ 0 & \text{if } t \leq 0 \end{cases},$$

(for example if $\alpha = 1/2$ then $y(t) = \frac{1}{4}t^2 1_{t \geq 0}$) then the reader may easily check y also solve Eq. (15.6). Furthermore, $y_a(t) := y(t-a)$ also solves Eq. (15.6) for all $a \geq 0$, see Figure 15.1 below.

With these examples in mind, let us now go to the general theory. The case of linear ODE's has already been studied in Section 8.3 above.

15.2 Uniqueness Theorem and Continuous Dependence on Initial Data

Lemma 15.2. ***Gronwall's Lemma.*** *Suppose that $f, \varepsilon,$ and k are non-negative functions of a real variable t such that*

$$f(t) \leq \varepsilon(t) + \left| \int_0^t k(\tau) f(\tau) d\tau \right|. \tag{15.8}$$

Then

Fig. 15.1. Three different solutions to the ODE $\dot{y}(t) = |y(t)|^{1/2}$ with $y(0) = 0$.

$$f(t) \leq \varepsilon(t) + \left| \int_0^t k(\tau)\varepsilon(\tau) e^{\left|\int_\tau^t k(s)ds\right|} d\tau \right|, \tag{15.9}$$

and in particular if ε and k are constants we find that

$$f(t) \leq \varepsilon e^{k|t|}. \tag{15.10}$$

Proof. I will only prove the case $t \geq 0$. The case $t \leq 0$ can be derived by applying the $t \geq 0$ to $\tilde{f}(t) = f(-t)$, $\tilde{k}(t) = k(-t)$ and $\varepsilon(t) = \varepsilon(-t)$. Set $F(t) = \int_0^t k(\tau) f(\tau) d\tau$. Then by (15.8),

$$\dot{F} = kf \leq k\varepsilon + kF.$$

Hence,

$$\frac{d}{dt}(e^{-\int_0^t k(s)ds} F) = e^{-\int_0^t k(s)ds}(\dot{F} - kF) \leq k\varepsilon e^{-\int_0^t k(s)ds}.$$

Integrating this last inequality from 0 to t and then solving for F yields:

$$F(t) \leq e^{\int_0^t k(s)ds} \cdot \int_0^t d\tau k(\tau)\varepsilon(\tau) e^{-\int_0^\tau k(s)ds} = \int_0^t d\tau k(\tau)\varepsilon(\tau) e^{\int_\tau^t k(s)ds}.$$

But by the definition of F we have that

$$f \leq \varepsilon + F,$$

and hence the last two displayed equations imply (15.9). Equation (15.10) follows from (15.9) by a simple integration. ∎

Corollary 15.3 (Continuous Dependence on Initial Data). *Let $U \subset_o X$, $0 \in (a,b)$ and $Z : (a,b) \times U \to X$ be a continuous function which is K–Lipschitz function on U, i.e. $\|Z(t,x) - Z(t,x')\| \leq K\|x - x'\|$ for all x and x' in U. Suppose $y_1, y_2 : (a,b) \to U$ solve*

$$\frac{dy_i(t)}{dt} = Z(t, y_i(t)) \quad \text{with } y_i(0) = x_i \quad \text{for } i = 1, 2. \tag{15.11}$$

Then

$$\|y_2(t) - y_1(t)\| \le \|x_2 - x_1\| e^{K|t|} \text{ for } t \in (a, b) \tag{15.12}$$

and in particular, there is at most one solution to Eq. (15.1) under the above Lipschitz assumption on Z.

Proof. Let $f(t) := \|y_2(t) - y_1(t)\|$. Then by the fundamental theorem of calculus,

$$\begin{aligned} f(t) &= \|y_2(0) - y_1(0) + \int_0^t (\dot{y}_2(\tau) - \dot{y}_1(\tau))\, d\tau\| \\ &\le f(0) + \left| \int_0^t \|Z(\tau, y_2(\tau)) - Z(\tau, y_1(\tau))\|\, d\tau \right| \\ &= \|x_2 - x_1\| + K \left| \int_0^t f(\tau)\, d\tau \right|. \end{aligned}$$

Therefore by Gronwall's inequality we have,

$$\|y_2(t) - y_1(t)\| = f(t) \le \|x_2 - x_1\| e^{K|t|}.$$

■

15.3 Local Existence (Non-Linear ODE)

We now show that Eq. (15.1) under a Lipschitz condition on Z. Another existence theorem was given in Exercise 11.19.

Theorem 15.4 (Local Existence). *Let $T > 0$, $J = (-T, T)$, $x_0 \in X$, $r > 0$ and*

$$C(x_0, r) := \{x \in X : \|x - x_0\| \le r\}$$

be the closed r – ball centered at $x_0 \in X$. Assume

$$M = \sup \{\|Z(t, x)\| : (t, x) \in J \times C(x_0, r)\} < \infty \tag{15.13}$$

and there exists $K < \infty$ such that

$$\|Z(t, x) - Z(t, y)\| \le K \|x - y\| \text{ for all } x, y \in C(x_0, r) \text{ and } t \in J. \tag{15.14}$$

Let $T_0 < \min\{r/M, T\}$ and $J_0 := (-T_0, T_0)$, then for each $x \in B(x_0, r - MT_0)$ there exists a unique solution $y(t) = y(t, x)$ to Eq. (15.2) in $C(J_0, C(x_0, r))$. Moreover $y(t, x)$ is jointly continuous in (t, x), $y(t, x)$ is differentiable in t, $\dot{y}(t, x)$ is jointly continuous for all $(t, x) \in J_0 \times B(x_0, r - MT_0)$ and satisfies Eq. (15.1).

Proof. The uniqueness assertion has already been proved in Corollary 15.3. To prove existence, let $C_r := C(x_0, r)$, $Y := C(J_0, C(x_0, r))$ and

$$S_x(y)(t) := x + \int_0^t Z(\tau, y(\tau))d\tau. \tag{15.15}$$

With this notation, Eq. (15.2) becomes $y = S_x(y)$, i.e. we are looking for a fixed point of S_x. If $y \in Y$, then

$$\begin{aligned}\|S_x(y)(t) - x_0\| &\le \|x - x_0\| + \left|\int_0^t \|Z(\tau, y(\tau))\| \, d\tau\right| \le \|x - x_0\| + M|t| \\ &\le \|x - x_0\| + MT_0 \le r - MT_0 + MT_0 = r,\end{aligned}$$

showing $S_x(Y) \subset Y$ for all $x \in B(x_0, r - MT_0)$. Moreover if $y, z \in Y$,

$$\begin{aligned}\|S_x(y)(t) - S_x(z)(t)\| &= \left\|\int_0^t [Z(\tau, y(\tau)) - Z(\tau, z(\tau))] \, d\tau\right\| \\ &\le \left|\int_0^t \|Z(\tau, y(\tau)) - Z(\tau, z(\tau))\| \, d\tau\right| \\ &\le K\left|\int_0^t \|y(\tau) - z(\tau)\| \, d\tau\right|. \end{aligned} \tag{15.16}$$

Let $y_0(t, x) = x$ and $y_n(\cdot, x) \in Y$ defined inductively by

$$y_n(\cdot, x) := S_x(y_{n-1}(\cdot, x)) = x + \int_0^t Z(\tau, y_{n-1}(\tau, x))d\tau. \tag{15.17}$$

Using the estimate in Eq. (15.16) repeatedly we find

$$\begin{aligned}&\| \, y_{n+1}(t) - y_n(t) \, \| \\ &\qquad \le K\left|\int_0^t \|y_n(\tau) - y_{n-1}(\tau)\| \, d\tau\right| \\ &\qquad \le K^2\left|\int_0^t dt_1 \left|\int_0^{t_1} dt_2 \, \|y_{n-1}(t_2) - y_{n-2}(t_2)\|\right|\right| \\ &\qquad \vdots \\ &\qquad \le K^n\left|\int_0^t dt_1 \left|\int_0^{t_1} dt_2 \dots \left|\int_0^{t_{n-1}} dt_n \, \|y_1(t_n) - y_0(t_n)\|\right| \dots\right|\right| \\ &\qquad \le K^n \|y_1(\cdot, x) - y_0(\cdot, x)\|_\infty \int_{\Delta_n(t)} d\tau \\ &\qquad = \frac{K^n |t|^n}{n!} \|y_1(\cdot, x) - y_0(\cdot, x)\|_\infty \le 2r\frac{K^n |t|^n}{n!} \end{aligned} \tag{15.18}$$

wherein we have also made use of Lemma 8.19. Combining this estimate with

$$\|y_1(t,x) - y_0(t,x)\| = \left\| \int_0^t Z(\tau,x)d\tau \right\| \le \left| \int_0^t \|Z(\tau,x)\|\, d\tau \right| \le M_0,$$

where

$$M_0 = T_0 \max\left\{ \int_0^{T_0} \|Z(\tau,x)\|\, d\tau, \int_{-T_0}^0 \|Z(\tau,x)\|\, d\tau \right\} \le MT_0,$$

shows

$$\|y_{n+1}(t,x) - y_n(t,x)\| \le M_0 \frac{K^n\, |t|^n}{n!} \le M_0 \frac{K^n T_0^n}{n!}$$

and this implies

$$\sum_{n=0}^{\infty} \sup\{\, \|y_{n+1}(\cdot,x) - y_n(\cdot,x)\|_{\infty,J_0} : t \in J_0\}$$
$$\le \sum_{n=0}^{\infty} M_0 \frac{K^n T_0^n}{n!} = M_0 e^{KT_0} < \infty$$

where

$$\|y_{n+1}(\cdot,x) - y_n(\cdot,x)\|_{\infty,J_0} := \sup\left\{\|y_{n+1}(t,x) - y_n(t,x)\| : t \in J_0\right\}.$$

So $y(t,x) := \lim_{n\to\infty} y_n(t,x)$ exists uniformly for $t \in J$ and using Eq. (15.14) we also have

$$\sup\{\, \|Z(t,y(t)) - Z(t,y_{n-1}(t))\| : t \in J_0\}$$
$$\le K\, \|y(\cdot,x) - y_{n-1}(\cdot,x)\|_{\infty,J_0} \to 0 \text{ as } n \to \infty.$$

Now passing to the limit in Eq. (15.17) shows y solves Eq. (15.2). From this equation it follows that $y(t,x)$ is differentiable in t and y satisfies Eq. (15.1). The continuity of $y(t,x)$ follows from Corollary 15.3 and mean value inequality (Corollary 8.14):

$$\begin{aligned} \|y(t,x) - y(t',x')\| &\le \|y(t,x) - y(t,x')\| + \|y(t,x') - y(t',x')\| \\ &= \|y(t,x) - y(t,x')\| + \left\| \int_{t'}^t Z(\tau, y(\tau,x'))d\tau \right\| \\ &\le \|y(t,x) - y(t,x')\| + \left| \int_{t'}^t \|Z(\tau, y(\tau,x'))\|\, d\tau \right| \\ &\le \|x - x'\| e^{KT} + \left| \int_{t'}^t \|Z(\tau, y(\tau,x'))\|\, d\tau \right| \qquad (15.19) \\ &\le \|x - x'\| e^{KT} + M\, |t - t'|\,. \end{aligned}$$

The continuity of $\dot{y}(t,x)$ is now a consequence Eq. (15.1) and the continuity of y and Z. ■

Corollary 15.5. *Let $J = (a, b) \ni 0$ and suppose $Z \in C(J \times X, X)$ satisfies*

$$\|Z(t, x) - Z(t, y)\| \le K \|x - y\| \text{ for all } x, y \in X \text{ and } t \in J. \tag{15.20}$$

Then for all $x \in X$, there is a unique solution $y(t, x)$ (for $t \in J$) to Eq. (15.1). Moreover $y(t, x)$ and $\dot{y}(t, x)$ are jointly continuous in (t, x).

Proof. Let $J_0 = (a_0, b_0) \ni 0$ be a precompact subinterval of J and $Y := BC(J_0, X)$. By compactness, $M := \sup_{t\in \bar{J}_0} \|Z(t, 0)\| < \infty$ which combined with Eq. (15.20) implies

$$\sup_{t\in \bar{J}_0} \|Z(t, x)\| \le M + K \|x\| \text{ for all } x \in X.$$

Using this estimate and Lemma 8.7 one easily shows $S_x(Y) \subset Y$ for all $x \in X$. The proof of Theorem 15.4 now goes through without any further change. ■

15.4 Global Properties

Definition 15.6 (Local Lipschitz Functions). *Let $U \subset_o X$, J be an open interval and $Z \in C(J \times U, X)$. The function Z is said to be locally Lipschitz in x if for all $x \in U$ and all compact intervals $I \subset J$ there exists $K = K(x, I) < \infty$ and $\varepsilon = \varepsilon(x, I) > 0$ such that $B(x, \varepsilon(x, I)) \subset U$ and*

$$\|Z(t, x_1) - Z(t, x_0)\| \le K(x, I)\|x_1 - x_0\| \ \forall\ x_0, x_1 \in B(x, \varepsilon(x, I)) \ \&\ t \in I. \tag{15.21}$$

For the rest of this section, we will assume J is an open interval containing 0, U is an open subset of X and $Z \in C(J\times U, X)$ is a locally Lipschitz function.

Lemma 15.7. *Let $Z \in C(J \times U, X)$ be a locally Lipschitz function in X and E be a compact subset of U and I be a compact subset of J. Then there exists $\varepsilon > 0$ such that $Z(t, x)$ is bounded for $(t, x) \in I \times E_\varepsilon$ and and $Z(t, x)$ is K – Lipschitz on E_ε for all $t \in I$, where*

$$E_\varepsilon := \{x \in U : \operatorname{dist}(x, E) < \varepsilon\}.$$

Proof. Let $\varepsilon(x, I)$ and $K(x, I)$ be as in Definition 15.6 above. Since E is compact, there exists a finite subset $\Lambda \subset E$ such that $E \subset V := \cup_{x\in\Lambda} B(x, \varepsilon(x, I)/2)$. If $y \in V$, there exists $x \in \Lambda$ such that $\|y - x\| < \varepsilon(x, I)/2$ and therefore

$$\begin{aligned}\|Z(t, y)\| &\le \|Z(t, x)\| + K(x, I) \|y - x\| \le \|Z(t, x)\| + K(x, I)\varepsilon(x, I)/2 \\ &\le \sup_{x\in\Lambda, t\in I} \{\|Z(t, x)\| + K(x, I)\varepsilon(x, I)/2\} =: M < \infty.\end{aligned}$$

This shows Z is bounded on $I \times V$. Let

$$\varepsilon := d(E, V^c) \le \frac{1}{2}\min_{x\in\Lambda} \varepsilon(x, I)$$

and notice that $\varepsilon > 0$ since E is compact, V^c is closed and $E \cap V^c = \emptyset$. If $y, z \in E_\varepsilon$ and $\|y - z\| < \varepsilon$, then as before there exists $x \in \Lambda$ such that $\|y - x\| < \varepsilon(x, I)/2$. Therefore

$$\|z - x\| \le \|z - y\| + \|y - x\| < \varepsilon + \varepsilon(x, I)/2 \le \varepsilon(x, I)$$

and since $y, z \in B(x, \varepsilon(x, I))$, it follows that

$$\|Z(t, y) - Z(t, z)\| \le K(x, I)\|y - z\| \le K_0\|y - z\|$$

where $K_0 := \max_{x\in\Lambda} K(x, I) < \infty$. On the other hand if $y, z \in E_\varepsilon$ and $\|y - z\| \ge \varepsilon$, then

$$\|Z(t, y) - Z(t, z)\| \le 2M \le \frac{2M}{\varepsilon}\|y - z\|.$$

Thus if we let $K := \max\{2M/\varepsilon, K_0\}$, we have shown

$$\|Z(t, y) - Z(t, z)\| \le K\|y - z\| \text{ for all } y, z \in E_\varepsilon \text{ and } t \in I.$$

■

Proposition 15.8 (Maximal Solutions). *Let $Z \in C(J \times U, X)$ be a locally Lipschitz function in x and let $x \in U$ be fixed. Then there is an interval $J_x = (a(x), b(x))$ with $a \in [-\infty, 0)$ and $b \in (0, \infty]$ and a C^1–function $y : J \to U$ with the following properties:*

1. *y solves ODE in Eq. (15.1).*
2. *If $\tilde{y} : \tilde{J} = (\tilde{a}, \tilde{b}) \to U$ is another solution of Eq. (15.1) (we assume that $0 \in \tilde{J}$) then $\tilde{J} \subset J$ and $\tilde{y} = y|_{\tilde{J}}$.*

The function $y : J \to U$ is called the maximal solution to Eq. (15.1).

Proof. Suppose that $y_i : J_i = (a_i, b_i) \to U$, $i = 1, 2$, are two solutions to Eq. (15.1). We will start by showing the $y_1 = y_2$ on $J_1 \cap J_2$. To do this[1] let

[1] Here is an alternate proof of the uniqueness. Let

$$T \equiv \sup\{t \in [0, \min\{b_1, b_2\}) : y_1 = y_2 \text{ on } [0, t]\}.$$

(T is the first positive time after which y_1 and y_2 disagree.

Suppose, for sake of contradiction, that $T < \min\{b_1, b_2\}$. Notice that $y_1(T) = y_2(T) =: x'$. Applying the local uniqueness theorem to $y_1(\cdot - T)$ and $y_2(\cdot - T)$ thought as function from $(-\delta, \delta) \to B(x', \epsilon(x'))$ for some δ sufficiently small, we learn that $y_1(\cdot - T) = y_2(\cdot - T)$ on $(-\delta, \delta)$. But this shows that $y_1 = y_2$ on $[0, T + \delta)$ which contradicts the definition of T. Hence we must have the $T = \min\{b_1, b_2\}$, i.e. $y_1 = y_2$ on $J_1 \cap J_2 \cap [0, \infty)$. A similar argument shows that $y_1 = y_2$ on $J_1 \cap J_2 \cap (-\infty, 0]$ as well.

$J_0 = (a_0, b_0)$ be chosen so that $0 \in J_0 \subset J_1 \cap J_2$, and let $E := y_1(J_0) \cup y_2(J_0)$ – a compact subset of X. Choose $\varepsilon > 0$ as in Lemma 15.7 so that Z is Lipschitz on E_ε. Then $y_1|_{J_0}, y_2|_{J_0} : J_0 \to E_\varepsilon$ both solve Eq. (15.1) and therefore are equal by Corollary 15.3. Since $J_0 = (a_0, b_0)$ was chosen arbitrarily so that $[a, b] \subset J_1 \cap J_2$, we may conclude that $y_1 = y_2$ on $J_1 \cap J_2$. Let $(y_\alpha, J_\alpha = (a_\alpha, b_\alpha))_{\alpha \in A}$ denote the possible solutions to (15.1) such that $0 \in J_\alpha$. Define $J_x = \cup J_\alpha$ and set $y = y_\alpha$ on J_α. We have just checked that y is well defined and the reader may easily check that this function $y : J_x \to U$ satisfies all the conclusions of the theorem. ∎

Notation 15.9 *For each $x \in U$, let $J_x = (a(x), b(x))$ be the maximal interval on which Eq. (15.1) may be solved, see Proposition 15.8. Set $\mathcal{D}(Z) := \cup_{x \in U} (J_x \times \{x\}) \subset J \times U$ and let $\phi : \mathcal{D}(Z) \to U$ be defined by $\phi(t, x) = y(t)$ where y is the maximal solution to Eq. (15.1). (So for each $x \in U$, $\phi(\cdot, x)$ is the maximal solution to Eq. (15.1).)*

Proposition 15.10. *Let $Z \in C(J \times U, X)$ be a locally Lipschitz function in x and $y : J_x = (a(x), b(x)) \to U$ be the maximal solution to Eq. (15.1). If $b(x) < b$, then either $\limsup_{t \uparrow b(x)} \|Z(t, y(t))\| = \infty$ or $y(b(x)-) := \lim_{t \uparrow b(x)} y(t)$ exists and $y(b(x)-) \notin U$. Similarly, if $a > a(x)$, then either $\limsup_{t \downarrow a(x)} \|y(t)\| = \infty$ or $y(a(x)+) := \lim_{t \downarrow a} y(t)$ exists and $y(a(x)+) \notin U$.*

Proof. Suppose that $b < b(x)$ and $M := \limsup_{t \uparrow b(x)} \|Z(t, y(t))\| < \infty$. Then there is a $b_0 \in (0, b(x))$ such that $\|Z(t, y(t))\| \leq 2M$ for all $t \in (b_0, b(x))$. Thus, by the usual fundamental theorem of calculus argument,

$$\|y(t) - y(t')\| \leq \left| \int_t^{t'} \|Z(t, y(\tau))\| \, d\tau \right| \leq 2M|t - t'|$$

for all $t, t' \in (b_0, b(x))$. From this it is easy to conclude that $y(b(x)-) = \lim_{t \uparrow b(x)} y(t)$ exists. If $y(b(x)-) \in U$, by the local existence Theorem 15.4, there exists $\delta > 0$ and $w \in C^1((b(x) - \delta, b(x) + \delta), U)$ such that

$$\dot{w}(t) = Z(t, w(t)) \text{ and } w(b(x)) = y(b(x)-).$$

Now define $\tilde{y} : (a, b(x) + \delta) \to U$ by

$$\tilde{y}(t) = \begin{cases} y(t) & \text{if } t \in J_x \\ w(t) & \text{if } t \in [b(x), b(x) + \delta) \end{cases}.$$

The reader may now easily show $\tilde{y}$ solves the integral Eq. (15.2) and hence also solves Eq. 15.1 for $t \in (a(x), b(x) + \delta)$.[2] But this violates the maximality of y and hence we must have that $y(b(x)-) \notin U$. The assertions for t near $a(x)$ are proved similarly. ∎

[2] See the argument in Proposition 15.13 for a slightly different method of extending y which avoids the use of the integral equation (15.2).

Example 15.11. Let $X = \mathbb{R}^2$, $J = \mathbb{R}$, $U = \left\{(x,y) \in \mathbb{R}^2 : 0 < r < 1\right\}$ where $r^2 = x^2 + y^2$ and

$$Z(x,y) = \frac{1}{r}(x,y) + \frac{1}{1-r^2}(-y,x).$$

The the unique solution $(x(t), y(t))$ to

$$\frac{d}{dt}(x(t), y(t)) = Z(x(t), y(t)) \text{ with } (x(0), y(0)) = (\frac{1}{2}, 0)$$

is given by

$$(x(t), y(t)) = \left(t + \frac{1}{2}\right)\left(\cos\left(\frac{1}{1/2 - t}\right), \sin\left(\frac{1}{1/2-t}\right)\right)$$

for $t \in J_{(1/2,0)} = (-\infty, 1/2)$. Notice that $\|Z(x(t), y(t))\| \to \infty$ as $t \uparrow 1/2$ and $\operatorname{dist}((x(t), y(t)), U^c) \to 0$ as $t \uparrow 1/2$.

Example 15.12. (Not worked out completely.) Let $X = U = \ell^2$, $\psi \in C^\infty(\mathbb{R}^2)$ be a smooth function such that $\psi = 1$ in a neighborhood of the line segment joining $(1,0)$ to $(0,1)$ and being supported within the $1/10$ – neighborhood of this segment. Choose $a_n \uparrow \infty$ and $b_n \uparrow \infty$ and define

$$Z(x) = \sum_{n=1}^{\infty} a_n \psi(b_n(x_n, x_{n+1}))(e_{n+1} - e_n). \tag{15.22}$$

For any $x \in \ell^2$, only a finite number of terms are non-zero in the above some in a neighborhood of x. Therefor $Z : \ell^2 \to \ell^2$ is a smooth and hence locally Lipshcitz vector field. Let $(y(t), J = (a,b))$ denote the maximal solution to

$$\dot{y}(t) = Z(y(t)) \text{ with } y(0) = e_1.$$

Then if the a_n and b_n are chosen appropriately, then $b < \infty$ and there will exist $t_n \uparrow b$ such that $y(t_n)$ is approximately e_n for all n. So again $y(t_n)$ does not have a limit yet $\sup_{t\in[0,b)} \|y(t)\| < \infty$. The idea is that Z is constructed to blow the particle form e_1 to e_2 to e_3 to e_4 etc. etc. with the time it takes to travel from e_n to e_{n+1} being on order $1/2^n$. The vector field in Eq. (15.22) is a first approximation at such a vector field, it may have to be adjusted a little more to provide an honest example. In this example, we are having problems because $y(t)$ is "going off in dimensions."

Here is another version of Proposition 15.10 which is more useful when $\dim(X) < \infty$.

Proposition 15.13. *Let $Z \in C(J \times U, X)$ be a locally Lipschitz function in x and $y : J_x = (a(x), b(x)) \to U$ be the maximal solution to Eq. (15.1).*

1. *If $b(x) < b$, then for every compact subset $K \subset U$ there exists $T_K < b(x)$ such that $y(t) \notin K$ for all $t \in [T_K, b(x))$.*

2. When $\dim(X) < \infty$, *we may write this condition as: if* $b(x) < b$, *then either*

$$\limsup_{t \uparrow b(x)} \|y(t)\| = \infty \text{ or } \liminf_{t \uparrow b(x)} dist(y(t), U^c) = 0.$$

Proof. 1) Suppose that $b(x) < b$ and, for sake of contradiction, there exists a compact set $K \subset U$ and $t_n \uparrow b(x)$ such that $y(t_n) \in K$ for all n. Since K is compact, by passing to a subsequence if necessary, we may assume $y_\infty := \lim_{n\to\infty} y(t_n)$ exists in $K \subset U$. By the local existence Theorem 15.4, there exists $T_0 > 0$ and $\delta > 0$ such that for each $x' \in B(y_\infty, \delta)$ there exists a unique solution $w(\cdot, x') \in C^1((-T_0, T_0), U)$ solving

$$w(t, x') = Z(t, w(t, x')) \text{ and } w(0, x') = x'.$$

Now choose n sufficiently large so that $t_n \in (b(x) - T_0/2, b(x))$ and $y(t_n) \in B(y_\infty, \delta)$. Define $\tilde{y} : (a(x), b(x) + T_0/2) \to U$ by

$$\tilde{y}(t) = \begin{cases} y(t) & \text{if } t \in J_x \\ w(t - t_n, y(t_n)) & \text{if } t \in (t_n - T_0, b(x) + T_0/2). \end{cases}$$

wherein we have used $(t_n - T_0, b(x) + T_0/2) \subset (t_n - T_0, t_n + T_0)$. By uniqueness of solutions to ODE's $\tilde{y}$ is well defined, $\tilde{y} \in C^1((a(x), b(x) + T_0/2), X)$ and $\tilde{y}$ solves the ODE in Eq. 15.1. But this violates the maximality of y. 2) For each $n \in \mathbb{N}$ let

$$K_n := \{x \in U : \|x\| \le n \text{ and } \operatorname{dist}(x, U^c) \ge 1/n\}.$$

Then $K_n \uparrow U$ and each K_n is a closed bounded set and hence compact if $\dim(X) < \infty$. Therefore if $b(x) < b$, by item 1., there exists $T_n \in [0, b(x))$ such that $y(t) \notin K_n$ for all $t \in [T_n, b(x))$ or equivalently $\|y(t)\| > n$ or $\operatorname{dist}(y(t), U^c) < 1/n$ for all $t \in [T_n, b(x))$. ∎

Remark 15.14. In general it is **not** true that the functions a and b are continuous. For example, let U be the region in $\mathbb{R}^2$ described in polar coordinates by $r > 0$ and $0 < \theta < 3\pi/4$ and $Z(x, y) = (0, -1)$ as in Figure 15.2 below. Then $b(x, y) = y$ for all $x, y > 0$ while $b(x, y) = \infty$ for all $x < 0$ and $y \in \mathbb{R}$ which shows b is discontinuous. On the other hand notice that

$$\{b > t\} = \{x < 0\} \cup \{(x, y) : x \ge 0, y > t\}$$

is an open set for all $t > 0$. An example of a vector field for which $b(x)$ is discontinuous is given in the top left hand corner of Figure 15.2. The map ψ would allow the reader to find an example on $\mathbb{R}^2$ if so desired. Some calculations shows that Z transferred to $\mathbb{R}^2$ by the map ψ is given by the new vector

$$\tilde{Z}(x, y) = -e^{-x} \left(\sin\left(\frac{3\pi}{8} + \frac{3}{4} \tan^{-1}(y) \right), \cos\left(\frac{3\pi}{8} + \frac{3}{4} \tan^{-1}(y) \right) \right).$$

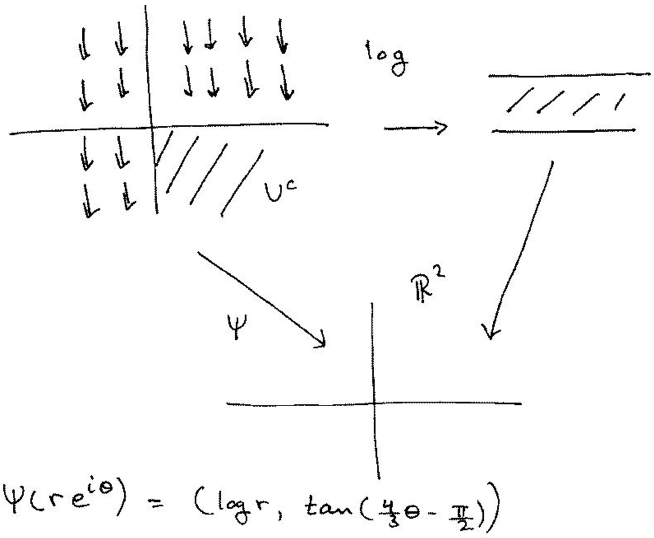


Fig. 15.2. Manufacturing vector fields where $b(x)$ is discontinuous.

Theorem 15.15 (Global Continuity). *Let $Z \in C(J \times U, X)$ be a locally Lipschitz function in x. Then $\mathcal{D}(Z)$ is an open subset of $J \times U$ and the functions $\phi : \mathcal{D}(Z) \to U$ and $\dot{\phi} : \mathcal{D}(Z) \to U$ are continuous. More precisely, for all $x_0 \in U$ and all open intervals J_0 such that $0 \in J_0 \sqsubset\sqsubset J_{x_0}$ there exists $\delta = \delta(x_0, J_0, Z) > 0$ and $C = C(x_0, J_0, Z) < \infty$ such that for all $x \in B(x_0, \delta)$, $J_0 \subset J_x$ and*

$$\|\phi(\cdot, x) - \phi(\cdot, x_0)\|_{BC(J_0,U)} \leq C \|x - x_0\| . \tag{15.23}$$

Proof. Let $|J_0| = b_0 - a_0$, $I = \bar{J}_0$ and $E := y(\bar{J}_0)$ – a compact subset of U and let $\varepsilon > 0$ and $K < \infty$ be given as in Lemma 15.7, i.e. K is the Lipschitz constant for Z on E_ε. Also recall the notation: $\Delta_1(t) = [0, t]$ if $t > 0$ and $\Delta_1(t) = [t, 0]$ if $t < 0$. Suppose that $x \in E_\varepsilon$, then by Corollary 15.3,

$$\|\phi(t, x) - \phi(t, x_0)\| \leq \|x - x_0\| e^{K|t|} \leq \|x - x_0\| e^{K|J_0|} \tag{15.24}$$

for all $t \in J_0 \cap J_x$ such that such that $\phi(\Delta_1(t), x) \subset E_\varepsilon$. Letting $\delta := \varepsilon e^{-K|J_0|}/2$, and assuming $x \in B(x_0, \delta)$, the previous equation implies

$$\|\phi(t, x) - \phi(t, x_0)\| \leq \varepsilon/2 < \varepsilon \;\forall\; t \in J_0 \cap J_x \;\ni\; \phi(\Delta_1(t), x) \subset E_\varepsilon.$$

This estimate further shows that $\phi(t, x)$ remains bounded and strictly away from the boundary of U for all such t. Therefore, it follows from Proposition 15.8 and "continuous induction[3]" that $J_0 \subset J_x$ and Eq. (15.24) is valid for all

[3] See the argument in the proof of Proposition 8.11.

$t \in J_0$. This proves Eq. (15.23) with $C := e^{K|J_0|}$. Suppose that $(t_0, x_0) \in \mathcal{D}(Z)$ and let $0 \in J_0 \sqsubset\sqsubset J_{x_0}$ such that $t_0 \in J_0$ and δ be as above. Then we have just shown $J_0 \times B(x_0, \delta) \subset \mathcal{D}(Z)$ which proves $\mathcal{D}(Z)$ is open. Furthermore, since the evaluation map

$$(t_0, y) \in J_0 \times BC(J_0, U) \xrightarrow{e} y(t_0) \in X$$

is continuous (as the reader should check) it follows that $\phi = e \circ (x \to \phi(\cdot, x)) : J_0 \times B(x_0, \delta) \to U$ is also continuous; being the composition of continuous maps. The continuity of $\dot{\phi}(t_0, x)$ is a consequences of the continuity of ϕ and the differential equation 15.1 Alternatively using Eq. (15.2),

$$\begin{aligned}
\|\phi(t_0, x) - \phi(t, x_0)\| &\le \|\phi(t_0, x) - \phi(t_0, x_0)\| + \|\phi(t_0, x_0) - \phi(t, x_0)\| \\
&\le C \|x - x_0\| + \left| \int_t^{t_0} \|Z(\tau, \phi(\tau, x_0))\| \, d\tau \right| \\
&\le C \|x - x_0\| + M |t_0 - t|
\end{aligned}$$

where C is the constant in Eq. (15.23) and $M = \sup_{\tau \in J_0} \|Z(\tau, \phi(\tau, x_0))\| < \infty$. This clearly shows ϕ is continuous. ■

15.5 Semi-Group Properties of time independent flows

To end this chapter we investigate the semi-group property of the flow associated to the vector-field Z. It will be convenient to introduce the following suggestive notation. For $(t, x) \in \mathcal{D}(Z)$, set $e^{tZ}(x) = \phi(t, x)$. So the path $t \to e^{tZ}(x)$ is the maximal solution to

$$\frac{d}{dt} e^{tZ}(x) = Z(e^{tZ}(x)) \quad \text{with } e^{0Z}(x) = x.$$

This exponential notation will be justified shortly. It is convenient to have the following conventions.

Notation 15.16 *We write $f : X \to X$ to mean a function defined on some open subset $D(f) \subset X$. The open set $D(f)$ will be called the domain of f. Given two functions $f : X \to X$ and $g : X \to X$ with domains $D(f)$ and $D(g)$ respectively, we define the composite function $f \circ g : X \to X$ to be the function with domain*

$$D(f \circ g) = \{x \in X : x \in D(g) \ \text{ and } g(x) \in D(f)\} = g^{-1}(D(f))$$

given by the rule $f \circ g(x) = f(g(x))$ for all $x \in D(f \circ g)$. We now write $f = g$ iff $D(f) = D(g)$ and $f(x) = g(x)$ for all $x \in D(f) = D(g)$. We will also write $f \subset g$ iff $D(f) \subset D(g)$ and $g|_{D(f)} = f$.

Theorem 15.17. *For fixed $t \in \mathbb{R}$ we consider e^{tZ} as a function from X to X with domain $D(e^{tZ}) = \{x \in U : (t,x) \in \mathcal{D}(Z)\}$, where $D(\phi) = \mathcal{D}(Z) \subset \mathbb{R} \times U$, $\mathcal{D}(Z)$ and ϕ are defined in Notation 15.9. Conclusions:*

1. *If $t, s \in \mathbb{R}$ and $t \cdot s \geq 0$, then $e^{tZ} \circ e^{sZ} = e^{(t+s)Z}$.*
2. *If $t \in \mathbb{R}$, then $e^{tZ} \circ e^{-tZ} = Id_{D(e^{-tZ})}$.*
3. *For arbitrary $t, s \in \mathbb{R}$, $e^{tZ} \circ e^{sZ} \subset e^{(t+s)Z}$.*

Proof. Item 1. For simplicity assume that $t, s \geq 0$. The case $t, s \leq 0$ is left to the reader. Suppose that $x \in D(e^{tZ} \circ e^{sZ})$. Then by assumption $x \in D(e^{sZ})$ and $e^{sZ}(x) \in D(e^{tZ})$. Define the path $y(\tau)$ via:

$$y(\tau) = \begin{cases} e^{\tau Z}(x) & \text{if } 0 \leq \tau \leq s \\ e^{(\tau - s)Z}(x) & \text{if } s \leq \tau \leq t + s \end{cases}.$$

It is easy to check that y solves $\dot{y}(\tau) = Z(y(\tau))$ with $y(0) = x$. But since, $e^{\tau Z}(x)$ is the maximal solution we must have that $x \in D(e^{(t+s)Z})$ and $y(t + s) = e^{(t+s)Z}(x)$. That is $e^{(t+s)Z}(x) = e^{tZ} \circ e^{sZ}(x)$. Hence we have shown that $e^{tZ} \circ e^{sZ} \subset e^{(t+s)Z}$. To finish the proof of item 1. it suffices to show that $D(e^{(t+s)Z}) \subset D(e^{tZ} \circ e^{sZ})$. Take $x \in D(e^{(t+s)Z})$, then clearly $x \in D(e^{sZ})$. Set $y(\tau) = e^{(\tau+s)Z}(x)$ defined for $0 \leq \tau \leq t$. Then y solves

$$\dot{y}(\tau) = Z(y(\tau)) \text{ with } y(0) = e^{sZ}(x).$$

But since $\tau \to e^{\tau Z}(e^{sZ}(x))$ is the maximal solution to the above initial valued problem we must have that $y(\tau) = e^{\tau Z}(e^{sZ}(x))$, and in particular at $\tau = t$, $e^{(t+s)Z}(x) = e^{tZ}(e^{sZ}(x))$. This shows that $x \in D(e^{tZ} \circ e^{sZ})$ and in fact $e^{(t+s)Z} \subset e^{tZ} \circ e^{sZ}$.

Item 2. Let $x \in D(e^{-tZ})$ – again assume for simplicity that $t \geq 0$. Set $y(\tau) = e^{(\tau - t)Z}(x)$ defined for $0 \leq \tau \leq t$. Notice that $y(0) = e^{-tZ}(x)$ and $\dot{y}(\tau) = Z(y(\tau))$. This shows that $y(\tau) = e^{\tau Z}(e^{-tZ}(x))$ and in particular that $x \in D(e^{tZ} \circ e^{-tZ})$ and $e^{tZ} \circ e^{-tZ}(x) = x$. This proves item 2.

Item 3. I will only consider the case that $s < 0$ and $t + s \geq 0$, the other cases are handled similarly. Write u for $t + s$, so that $t = -s + u$. We know that $e^{tZ} = e^{uZ} \circ e^{-sZ}$ by item 1. Therefore

$$e^{tZ} \circ e^{sZ} = (e^{uZ} \circ e^{-sZ}) \circ e^{sZ}.$$

Notice in general, one has $(f \circ g) \circ h = f \circ (g \circ h)$ (you prove). Hence, the above displayed equation and item 2. imply that

$$e^{tZ} \circ e^{sZ} = e^{uZ} \circ (e^{-sZ} \circ e^{sZ}) = e^{(t+s)Z} \circ I_{D(e^{sZ})} \subset e^{(t+s)Z}.$$

■

The following result is trivial but conceptually illuminating partial converse to Theorem 15.17.

Proposition 15.18 (Flows and Complete Vector Fields). *Suppose* $U \subset_o X$, $\phi \in C(\mathbb{R} \times U, U)$ *and* $\phi_t(x) = \phi(t, x)$. *Suppose* ϕ *satisfies:*

1. $\phi_0 = I_U$,
2. $\phi_t \circ \phi_s = \phi_{t+s}$ *for all* $t, s \in \mathbb{R}$, *and*
3. $Z(x) := \dot{\phi}(0, x)$ *exists for all* $x \in U$ *and* $Z \in C(U, X)$ *is locally Lipschitz.*

Then $\phi_t = e^{tZ}$.

Proof. Let $x \in U$ and $y(t) := \phi_t(x)$. Then using Item 2.,

$$\dot{y}(t) = \frac{d}{ds}|_0 y(t+s) = \frac{d}{ds}|_0 \phi_{(t+s)}(x) = \frac{d}{ds}|_0 \phi_s \circ \phi_t(x) = Z(y(t)).$$

Since $y(0) = x$ by Item 1. and Z is locally Lipschitz by Item 3., we know by uniqueness of solutions to ODE's (Corollary 15.3) that $\phi_t(x) = y(t) = e^{tZ}(x)$. ■

15.6 Exercises

Exercise 15.1. Find a vector field Z such that $e^{(t+s)Z}$ is not contained in $e^{tZ} \circ e^{sZ}$.

Definition 15.19. *A locally Lipschitz function* $Z : U \subset_o X \to X$ *is said to be a complete vector field if* $\mathcal{D}(Z) = \mathbb{R} \times U$. *That is for any* $x \in U$, $t \to e^{tZ}(x)$ *is defined for all* $t \in \mathbb{R}$.

Exercise 15.2. Suppose that $Z : X \to X$ is a locally Lipschitz function. Assume there is a constant $C > 0$ such that

$$\|Z(x)\| \le C(1 + \|x\|) \text{ for all } x \in X.$$

Then Z is complete. **Hint:** use Gronwall's Lemma 15.2 and Proposition 15.10.

Exercise 15.3. Suppose y is a solution to $\dot{y}(t) = |y(t)|^{1/2}$ with $y(0) = 0$. Show there exists $a, b \in [0, \infty]$ such that

$$y(t) = \begin{cases} \frac{1}{4}(t-b)^2 & \text{if } \quad t \ge b \\ 0 & \text{if } -a < t < b \\ -\frac{1}{4}(t+a)^2 & \text{if } \quad t \le -a. \end{cases}$$

Exercise 15.4. Using the fact that the solutions to Eq. (15.3) are never 0 if $x \neq 0$, show that $y(t) = 0$ is the only solution to Eq. (15.3) with $y(0) = 0$.

Exercise 15.5 (Higher Order ODE). Let X be a Banach space, , $\mathcal{U} \subset_o X^n$ and $f \in C(J \times \mathcal{U}, X)$ be a Locally Lipschitz function in $\mathbf{x} = (x_1, \ldots, x_n)$. Show the n^{th} ordinary differential equation,

$$y^{(n)}(t) = f(t, y(t), \dot{y}(t), \ldots y^{(n-1)}(t)) \text{ with } y^{(k)}(0) = y_0^k \text{ for } k < n \tag{15.25}$$

where $(y_0^0, \ldots, y_0^{n-1})$ is given in $\mathcal{U}$, has a unique solution for small $t \in J$. **Hint**: let $\mathbf{y}(t) = \big(y(t), \dot{y}(t), \ldots y^{(n-1)}(t)\big)$ and rewrite Eq. (15.25) as a first order ODE of the form

$$\dot{\mathbf{y}}(t) = Z(t, \mathbf{y}(t)) \text{ with } \mathbf{y}(0) = (y_0^0, \ldots, y_0^{n-1}).$$

Exercise 15.6. Use the results of Exercises 8.20 and 15.5 to solve

$$\ddot{y}(t) - 2\dot{y}(t) + y(t) = 0 \text{ with } y(0) = a \text{ and } \dot{y}(0) = b.$$

Hint: The 2×2 matrix associated to this system, A, has only one eigenvalue 1 and may be written as $A = I + B$ where $B^2 = 0$.

Exercise 15.7 (Non-Homogeneous ODE). Suppose that $U \subset_o X$ is open and $Z : \mathbb{R} \times U \to X$ is a continuous function. Let $J = (a, b)$ be an interval and $t_0 \in J$. Suppose that $y \in C^1(J, U)$ is a solution to the "non-homogeneous" differential equation:

$$\dot{y}(t) = Z(t, y(t)) \text{ with } y(t_o) = x \in U. \tag{15.26}$$

Define $Y \in C^1(J - t_0, \mathbb{R} \times U)$ by $Y(t) := (t + t_0, y(t + t_0))$. Show that Y solves the "homogeneous" differential equation

$$\dot{Y}(t) = \tilde{Z}(Y(t)) \text{ with } Y(0) = (t_0, y_0), \tag{15.27}$$

where $\tilde{Z}(t, x) := (1, Z(x))$. Conversely, suppose that $Y \in C^1(J - t_0, \mathbb{R} \times U)$ is a solution to Eq. (15.27). Show that $Y(t) = (t + t_0, y(t + t_0))$ for some $y \in C^1(J, U)$ satisfying Eq. (15.26). (In this way the theory of non-homogeneous ode's may be reduced to the theory of homogeneous ode's.)

Exercise 15.8 (Differential Equations with Parameters). Let W be another Banach space, $U \times V \subset_o X \times W$ and $Z \in C(U \times V, X)$ be a locally Lipschitz function on $U \times V$. For each $(x, w) \in U \times V$, let $t \in J_{x,w} \to \phi(t, x, w)$ denote the maximal solution to the ODE

$$\dot{y}(t) = Z(y(t), w) \text{ with } y(0) = x. \tag{15.28}$$

Prove

$$\mathcal{D} := \{(t, x, w) \in \mathbb{R} \times U \times V : t \in J_{x,w}\} \tag{15.29}$$

is open in $\mathbb{R} \times U \times V$ and ϕ and $\dot{\phi}$ are continuous functions on $\mathcal{D}$.

Hint: If $y(t)$ solves the differential equation in (15.28), then $v(t) := (y(t), w)$ solves the differential equation,

$$\dot{v}(t) = \tilde{Z}(v(t)) \text{ with } v(0) = (x, w), \tag{15.30}$$

where $\tilde{Z}(x, w) := (Z(x, w), 0) \in X \times W$ and let $\psi(t, (x, w)) := v(t)$. Now apply the Theorem 15.15 to the differential equation (15.30).

Exercise 15.9 (Abstract Wave Equation). For $A \in L(X)$ and $t \in \mathbb{R}$, let

$$\cos(tA) := \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} t^{2n} A^{2n} \text{ and}$$
$$\frac{\sin(tA)}{A} := \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} t^{2n+1} A^{2n}.$$

Show that the unique solution $y \in C^2(\mathbb{R}, X)$ to

$$\ddot{y}(t) + A^2 y(t) = 0 \text{ with } y(0) = y_0 \text{ and } \dot{y}(0) = \dot{y}_0 \in X \tag{15.31}$$

is given by

$$y(t) = \cos(tA) y_0 + \frac{\sin(tA)}{A} \dot{y}_0.$$

Remark 15.20. Exercise 15.9 can be done by direct verification. Alternatively and more instructively, rewrite Eq. (15.31) as a first order ODE using Exercise 15.5. In doing so you will be lead to compute e^{tB} where $B \in L(X \times X)$ is given by

$$B = \begin{pmatrix} 0 & I \\ -A^2 & 0 \end{pmatrix},$$

where we are writing elements of $X \times X$ as column vectors, $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. You should then show

$$e^{tB} = \begin{pmatrix} \cos(tA) & \frac{\sin(tA)}{A} \\ -A\sin(tA) & \cos(tA) \end{pmatrix}$$

where

$$A\sin(tA) := \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} t^{2n+1} A^{2(n+1)}.$$

Exercise 15.10 (Duhamel's Principle for the Abstract Wave Equation). Continue the notation in Exercise 15.9, but now consider the ODE,

$$\ddot{y}(t) + A^2 y(t) = f(t) \text{ with } y(0) = y_0 \text{ and } \dot{y}(0) = \dot{y}_0 \in X \tag{15.32}$$

where $f \in C(\mathbb{R}, X)$. Show the unique solution to Eq. (15.32) is given by

$$y(t) = \cos(tA) y_0 + \frac{\sin(tA)}{A} \dot{y}_0 + \int_0^t \frac{\sin((t-\tau)A)}{A} f(\tau) d\tau \tag{15.33}$$

Hint: Again this could be proved by direct calculation. However it is more instructive to deduce Eq. (15.33) from Exercise 8.22 and the comments in Remark 15.20.

16

Banach Space Calculus

In this section, X and Y will be Banach space and U will be an open subset of X.

Notation 16.1 (ε, O, and o notation) *Let $0 \in U \subset_o X$, and $f : U \to Y$ be a function. We will write:*

1. *$f(x) = \varepsilon(x)$ if $\lim_{x \to 0} \|f(x)\| = 0$.*
2. *$f(x) = O(x)$ if there are constants $C < \infty$ and $r > 0$ such that $\|f(x)\| \leq C\|x\|$ for all $x \in B(0, r)$. This is equivalent to the condition that $\limsup_{x \to 0} \left(\|x\|^{-1}\|f(x)\|\right) < \infty$, where*
$$\limsup_{x \to 0} \frac{\|f(x)\|}{\|x\|} := \lim_{r \downarrow 0} \sup\{\|f(x)\| : 0 < \|x\| \leq r\}.$$
3. *$f(x) = o(x)$ if $f(x) = \varepsilon(x)O(x)$, i.e. $\lim_{x \to 0} \|f(x)\|/\|x\| = 0$.*

Example 16.2. Here are some examples of properties of these symbols.

1. A function $f : U \subset_o X \to Y$ is continuous at $x_0 \in U$ if $f(x_0 + h) = f(x_0) + \varepsilon(h)$.
2. If $f(x) = \varepsilon(x)$ and $g(x) = \varepsilon(x)$ then $f(x) + g(x) = \varepsilon(x)$.
   Now let $g : Y \to Z$ be another function where Z is another Banach space.
3. If $f(x) = O(x)$ and $g(y) = o(y)$ then $g \circ f(x) = o(x)$.
4. If $f(x) = \varepsilon(x)$ and $g(y) = \varepsilon(y)$ then $g \circ f(x) = \varepsilon(x)$.

16.1 The Differential

Definition 16.3. *A function $f : U \subset_o X \to Y$ is **differentiable** at $x_0 \in U$ if there exists a linear transformation $\Lambda \in L(X, Y)$ such that*

$$f(x_0 + h) - f(x_0) - \Lambda h = o(h). \tag{16.1}$$

*We denote Λ by $f'(x_0)$ or $Df(x_0)$ if it exists. As with continuity, f is **differentiable on** U if f is differentiable at all points in U.*

Remark 16.4. The linear transformation Λ in Definition 16.3 is necessarily unique. Indeed if Λ_1 is another linear transformation such that Eq. (16.1) holds with Λ replaced by Λ_1, then

$$(\Lambda - \Lambda_1)h = o(h),$$

i.e.

$$\limsup_{h\to 0} \frac{\|(\Lambda - \Lambda_1)h\|}{\|h\|} = 0.$$

On the other hand, by definition of the operator norm,

$$\limsup_{h\to 0} \frac{\|(\Lambda - \Lambda_1)h\|}{\|h\|} = \|\Lambda - \Lambda_1\|.$$

The last two equations show that $\Lambda = \Lambda_1$.

Exercise 16.1. Show that a function $f : (a, b) \to X$ is a differentiable at $t \in (a, b)$ in the sense of Definition 8.8 iff it is differentiable in the sense of Definition 16.3. Also show $Df(t)v = v\dot{f}(t)$ for all $v \in \mathbb{R}$.

Example 16.5. If $T \in L(X, Y)$ and $x, h \in X$, then

$$T(x + h) - T(x) - Th = 0$$

which shows $T'(x) = T$ for all $x \in X$.

Example 16.6. Assume that $GL(X, Y)$ is non-empty. Then by Corollary 7.20, $GL(X, Y)$ is an open subset of $L(X, Y)$ and the inverse map $f : GL(X, Y) \to GL(Y, X)$, defined by $f(A) := A^{-1}$, is continuous. We will now show that f is differentiable and

$$f'(A)B = -A^{-1}BA^{-1} \text{ for all } B \in L(X, Y).$$

This is a consequence of the identity,

$$f(A + H) - f(A) = (A + H)^{-1}(A - (A + H))A^{-1} = -(A + H)^{-1}HA^{-1}$$

which may be used to find the estimate,

$$\begin{aligned}
\left\|f(A + H) - f(A) + A^{-1}HA^{-1}\right\| &= \left\|\left[A^{-1} - (A + H)^{-1}\right]HA^{-1}\right\| \\
&\le \left\|A^{-1} - (A + H)^{-1}\right\| \|H\| \left\|A^{-1}\right\| \\
&\le \frac{\|A^{-1}\|^3 \|H\|^2}{1 - \|A^{-1}\| \|H\|} = O\left(\|H\|^2\right)
\end{aligned}$$

wherein we have used the bound in Eq. (7.8) of Corollary 7.20 for the last inequality.

16.2 Product and Chain Rules

The following theorem summarizes some basic properties of the differential.

Theorem 16.7. *The differential D has the following properties:*

1. ***Linearity:*** *D is linear, i.e. $D(f+\lambda g) = Df + \lambda Dg$.*
2. ***Product Rule:*** *If $f : U \subset_o X \to Y$ and $A : U \subset_o X \to L(X,Z)$ are differentiable at x_0 then so is $x \to (Af)(x) := A(x)f(x)$ and*
$$D(Af)(x_0)h = (DA(x_0)h)f(x_0) + A(x_0)Df(x_0)h.$$
3. ***Chain Rule:*** *If $f : U \subset_o X \to V \subset_o Y$ is differentiable at $x_0 \in U$, and $g : V \subset_o Y \to Z$ is differentiable at $y_0 := f(x_0)$, then $g \circ f$ is differentiable at x_0 and $(g \circ f)'(x_0) = g'(y_0)f'(x_0)$.*
4. ***Converse Chain Rule:*** *Suppose that $f : U \subset_o X \to V \subset_o Y$ is **continuous** at $x_0 \in U$, $g : V \subset_o Y \to Z$ is differentiable $y_0 := f(h_o)$, $g'(y_0)$ is invertible, and $g \circ f$ is differentiable at x_0, then f is differentiable at x_0 and*
$$f'(x_0) := [g'(x_0)]^{-1}(g \circ f)'(x_0). \tag{16.2}$$

Proof. Linearity. Let $f, g : U \subset_o X \to Y$ be two functions which are differentiable at $x_0 \in U$ and $\lambda \in \mathbb{R}$, then

$$\begin{aligned}
&(f+\lambda g)(x_0+h) \\
&\qquad = f(x_0) + Df(x_0)h + o(h) + \lambda(g(x_0) + Dg(x_0)h + o(h) \\
&\qquad = (f+\lambda g)(x_0) + (Df(x_0) + \lambda Dg(x_0))h + o(h),
\end{aligned}$$

which implies that $(f+\lambda g)$ is differentiable at x_0 and that

$$D(f+\lambda g)(x_0) = Df(x_0) + \lambda Dg(x_0).$$

Product Rule. The computation,

$$\begin{aligned}
&A(x_0+h)f(x_0+h) \\
&\qquad = (A(x_0) + DA(x_0)h + o(h))(f(x_0) + f'(x_0)h + o(h)) \\
&\qquad = A(x_0)f(x_0) + A(x_0)f'(x_0)h + [DA(x_0)h]f(x_0) + o(h),
\end{aligned}$$

verifies the product rule holds. This may also be considered as a special case of Proposition 16.9. **Chain Rule.** Using $f(x_0+h) - f(x_0) = O(h)$ (see Eq. (16.1)) and $o(O(h)) = o(h)$,

$$\begin{aligned}
&(g \circ f)(x_0+h) \\
&\quad = g(f(x_0)) + g'(f(x_0))(f(x_0+h) - f(x_0)) + o(f(x_0+h) - f(x_0)) \\
&\quad = g(f(x_0)) + g'(f(x_0))(Df(x_0)x_0 + o(h)) + o(f(x_0+h) - f(x_0) \\
&\quad = g(f(x_0)) + g'(f(x_0))Df(x_0)h + o(h).
\end{aligned}$$

Converse Chain Rule. Since g is differentiable at $y_0 = f(x_0)$ and $g'(y_0)$ is invertible,

$$\begin{aligned} & g(f(x_0+h)) - g(f(x_0)) \\ &\quad = g'(f(x_0))(f(x_0+h) - f(x_0)) + o(f(x_0+h) - f(x_0)) \\ &\quad = g'(f(x_0))\left[f(x_0+h) - f(x_0) + o(f(x_0+h) - f(x_0))\right]. \end{aligned}$$

And since $g \circ f$ is differentiable at x_0,

$$(g \circ f)(x_0+h) - g(f(x_0)) = (g \circ f)'(x_0)h + o(h).$$

Comparing these two equations shows that

$$\begin{aligned} & f(x_0+h) - f(x_0) + o(f(x_0+h) - f(x_0)) \\ &\qquad = g'(f(x_0))^{-1}\left[(g \circ f)'(x_0)h + o(h)\right] \end{aligned}$$

which is equivalent to

$$\begin{aligned} & f(x_0+h) - f(x_0) + o(f(x_0+h) - f(x_0)) \\ &\quad = g'(f(x_0))^{-1}\left[(g \circ f)'(x_0)h + o(h)\right] \\ &\quad = g'(f(x_0))^{-1}\{(g \circ f)'(x_0)h + o(h) - o(f(x_0+h) - f(x_0))\} \\ &\quad = g'(f(x_0))^{-1}(g \circ f)'(x_0)h + o(h) + o(f(x_0+h) - f(x_0)). \quad (16.3) \end{aligned}$$

Using the continuity of f, $f(x_0+h) - f(x_0)$ is close to 0 if h is close to zero, and hence

$$\|o(f(x_0+h) - f(x_0))\| \le \frac{1}{2}\|f(x_0+h) - f(x_0)\| \tag{16.4}$$

for all h sufficiently close to 0. (We may replace $\frac{1}{2}$ by any number $\alpha > 0$ above.) Taking the norm of both sides of Eq. (16.3) and making use of Eq. (16.4) shows, for h close to 0, that

$$\begin{aligned} & \|f(x_0+h) - f(x_0)\| \\ &\quad \le \|g'(f(x_0))^{-1}(g \circ f)'(x_0)\|\|h\| + o(\|h\|) + \frac{1}{2}\|f(x_0+h) - f(x_0)\|. \end{aligned}$$

Solving for $\|f(x_0+h) - f(x_0)\|$ in this last equation shows that

$$f(x_0+h) - f(x_0) = O(h). \tag{16.5}$$

(This is an improvement, since the continuity of f only guaranteed that $f(x_0+h) - f(x_0) = \varepsilon(h)$.) Because of Eq. (16.5), we now know that $o(f(x_0+h) - f(x_0)) = o(h)$, which combined with Eq. (16.3) shows that

$$f(x_0+h) - f(x_0) = g'(f(x_0))^{-1}(g \circ f)'(x_0)h + o(h),$$

i.e. f is differentiable at x_0 and $f'(x_0) = g'(f(x_0))^{-1}(g \circ f)'(x_0)$. ∎

Corollary 16.8 (Chain Rule). *Suppose that $\sigma : (a,b) \to U \subset_o X$ is differentiable at $t \in (a,b)$ and $f : U \subset_o X \to Y$ is differentiable at $\sigma(t) \in U$. Then $f \circ \sigma$ is differentiable at t and*

$$d(f \circ \sigma)(t)/dt = f'(\sigma(t))\dot{\sigma}(t).$$

Proposition 16.9 (Product Rule II). *Suppose that $X := X_1 \times \cdots \times X_n$ with each X_i being a Banach space and $T : X_1 \times \cdots \times X_n \to Y$ is a multilinear map, i.e.*

$$x_i \in X_i \to T(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) \in Y$$

is linear when $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ are held fixed. Then the following are equivalent:

1. *T is continuous.*
2. *T is continuous at $0 \in X$.*
3. *There exists a constant $C < \infty$ such that*
$$\|T(x)\|_Y \le C \prod_{i=1}^n \|x_i\|_{X_i} \tag{16.6}$$
for all $x = (x_1, \ldots, x_n) \in X$.
4. *T is differentiable at all $x \in X_1 \times \cdots \times X_n$.*

Moreover if T the differential of T is given by

$$T'(x)\,h = \sum_{i=1}^n T(x_1, \ldots, x_{i-1}, h_i, x_{i+1}, \ldots, x_n) \tag{16.7}$$

where $h = (h_1, \ldots, h_n) \in X$.

Proof. Let us equip X with the norm

$$\|x\|_X := \max\left\{\|x_i\|_{X_i}\right\}.$$

If T is continuous then T is continuous at 0. If T is continuous at 0, using $T(0) = 0$, there exists a $\delta > 0$ such that $\|T(x)\|_Y \le 1$ whenever $\|x\|_X \le \delta$. Now if $x \in X$ is arbitrary, let $x' := \delta\left(\|x_1\|_{X_1}^{-1} x_1, \ldots, \|x_n\|_{X_n}^{-1} x_n\right)$. Then $\|x'\|_X \le \delta$ and hence

$$\left\|\left(\delta^n \prod_{i=1}^n \|x_i\|_{X_i}^{-1}\right) T(x_1, \ldots, x_n)\right\|_Y = \|T(x')\| \le 1$$

from which Eq. (16.6) follows with $C = \delta^{-n}$.

Now suppose that Eq. (16.6) holds. For $x, h \in X$ and $\varepsilon \in \{0,1\}^n$ let $|\varepsilon| = \sum_{i=1}^n \varepsilon_i$ and

$$x^{\varepsilon}(h) := ((1-\varepsilon_1)x_1 + \varepsilon_1 h_1, \dots, (1-\varepsilon_n)x_n + \varepsilon_n h_n) \in X.$$

By the multi-linearity of T,

$$\begin{aligned} T(x+h) = T(x_1+h_1,\dots,x_n+h_n) &= \sum_{\varepsilon\in\{0,1\}^n} T(x^{\varepsilon}(h)) \\ &= T(x) + \sum_{i=1}^n T(x_1,\dots,x_{i-1},h_i,x_{i+1},\dots,x_n) \\ &+ \sum_{\varepsilon\in\{0,1\}^n:|\varepsilon|\ge 2} T(x^{\varepsilon}(h)). \end{aligned} \tag{16.8}$$

From Eq. (16.6),

$$\left\| \sum_{\varepsilon\in\{0,1\}^n:|\varepsilon|\ge 2} T(x^{\varepsilon}(h)) \right\| = O\left(\|h\|^2\right),$$

and so it follows from Eq. (16.8) that $T'(x)$ exists and is given by Eq. (16.7). This completes the proof since it is trivial to check that T being differentiable at $x \in X$ implies continuity of T at $x \in X$. ■

Exercise 16.2. Let $\det : L(\mathbb{R}^n) \to \mathbb{R}$ be the determinant function on $n \times n$ matrices and for $A \in L(\mathbb{R}^n)$ we will let A_i denote the i^{th} – column of A and write $A = (A_1|A_2|\dots|A_n)$.

1. Show $\det'(A)$ exists for all $A \in L(\mathbb{R}^n)$ and
$$\det{}'(A)H = \sum_{i=1}^n \det(A_1|\dots|A_{i-1}|H_i|A_{i+1}|\dots|A_n) \tag{16.9}$$
for all $H \in L(\mathbb{R}^n)$. **Hint**: recall that $\det(A)$ is a multilinear function of its columns.
2. Use Eq. (16.9) along with basic properties of the determinant to show $\det'(I)H = \operatorname{tr}(H)$.
3. Suppose now that $A \in GL(\mathbb{R}^n)$, show
$$\det{}'(A)H = \det(A)\operatorname{tr}(A^{-1}H).$$
Hint: Notice that $\det(A+H) = \det(A)\det\left(I + A^{-1}H\right)$.
4. If $A \in L(\mathbb{R}^n)$, show $\det\left(e^A\right) = e^{\operatorname{tr}(A)}$. **Hint:** use the previous item and Corollary 16.8 to show
$$\frac{d}{dt}\det\left(e^{tA}\right) = \det\left(e^{tA}\right)\operatorname{tr}(A).$$

Definition 16.10. *Let X and Y be Banach spaces and let $\mathcal{L}^1(X,Y) := L(X,Y)$ and for $k \geq 2$ let $\mathcal{L}^k(X,Y)$ be defined inductively by $\mathcal{L}^{k+1}(X,Y) = L(X,\mathcal{L}^k(X,Y))$. For example $\mathcal{L}^2(X,Y) = L(X,L(X,Y))$ and $\mathcal{L}^3(X,Y) = L\left(X,L(X,L(X,Y))\right)$.*

Suppose $f : U \subset_o X \to Y$ is a function. If f is differentiable on U, then it makes sense to ask if $f' = Df : U \to L(X,Y) = \mathcal{L}^1(X,Y)$ is differentiable. If Df is differentiable on U then $f'' = D^2 f := DDf : U \to \mathcal{L}^2(X,Y)$. Similarly we define $f^{(n)} = D^n f : U \to \mathcal{L}^n(X,Y)$ inductively.

Definition 16.11. *Given $k \in \mathbb{N}$, let $C^k(U,Y)$ denote those functions $f : U \to Y$ such that $f^{(j)} := D^j f : U \to \mathcal{L}^j(X,Y)$ exists and is continuous for $j = 1,2,\ldots,k$.*

Example 16.12. Let us continue on with Example 16.6 but now let $X = Y$ to simplify the notation. So $f : GL(X) \to GL(X)$ is the map $f(A) = A^{-1}$ and

$$f'(A) = -L_{A^{-1}} R_{A^{-1}}, \text{ i.e. } f' = -L_f R_f.$$

where $L_A B = AB$ and $R_A B = AB$ for all $A, B \in L(X)$. As the reader may easily check, the maps

$$A \in L(X) \to L_A, R_A \in L(L(X))$$

are linear and bounded. So by the chain and the product rule we find $f''(A)$ exists for all $A \in L(X)$ and

$$f''(A)B = -L_{f'(A)B} R_f - L_f R_{f'(A)B}.$$

More explicitly

$$[f''(A)B]\,C = A^{-1}BA^{-1}CA^{-1} + A^{-1}CA^{-1}BA^{-1}. \tag{16.10}$$

Working inductively one shows $f : GL(X) \to GL(X)$ defined by $f(A) := A^{-1}$ is C^∞.

16.3 Partial Derivatives

Definition 16.13 (Partial or Directional Derivative). *Let $f : U \subset_o X \to Y$ be a function, $x_0 \in U$, and $v \in X$. We say that f is differentiable at x_0 in the direction v iff $\frac{d}{dt}|_0(f(x_0 + tv)) =: (\partial_v f)(x_0)$ exists. We call $(\partial_v f)(x_0)$ the directional or partial derivative of f at x_0 in the direction v.*

Notice that if f is differentiable at x_0, then $\partial_v f(x_0)$ exists and is equal to $f'(x_0)v$, see Corollary 16.8.

Proposition 16.14. *Let $f : U \subset_o X \to Y$ be a continuous function and $D \subset X$ be a dense subspace of X. Assume $\partial_v f(x)$ exists for all $x \in U$ and $v \in D$, and there exists a continuous function $A : U \to L(X,Y)$ such that $\partial_v f(x) = A(x)v$ for all $v \in D$ and $x \in U \cap D$. Then $f \in C^1(U,Y)$ and $Df = A$.*

Proof. Let $x_0 \in U$, $\varepsilon > 0$ such that $B(x_0, 2\varepsilon) \subset U$ and $M := \sup\{\|A(x)\| : x \in B(x_0, 2\varepsilon)\} < \infty$[1]. For $x \in B(x_0,\varepsilon) \cap D$ and $v \in D \cap B(0,\varepsilon)$, by the fundamental theorem of calculus,

$$\begin{aligned} f(x+v) - f(x) &= \int_0^1 \frac{df(x+tv)}{dt}\, dt \\ &= \int_0^1 (\partial_v f)(x+tv)\, dt = \int_0^1 A(x+tv)\, v\, dt. \end{aligned} \tag{16.11}$$

For general $x \in B(x_0,\varepsilon)$ and $v \in B(0,\varepsilon)$, choose $x_n \in B(x_0,\varepsilon) \cap D$ and $v_n \in D \cap B(0,\varepsilon)$ such that $x_n \to x$ and $v_n \to v$. Then

$$f(x_n + v_n) - f(x_n) = \int_0^1 A(x_n + tv_n)\, v_n\, dt \tag{16.12}$$

holds for all n. The left side of this last equation tends to $f(x+v) - f(x)$ by the continuity of f. For the right side of Eq. (16.12) we have

$$\begin{aligned} &\|\int_0^1 A(x+tv)\, v\, dt - \int_0^1 A(x_n + tv_n)\, v_n\, dt\| \\ &\quad \le \int_0^1 \|A(x+tv) - A(x_n + tv_n)\|\|v\|\, dt + M\|v - v_n\|. \end{aligned}$$

It now follows by the continuity of A, the fact that $\|A(x+tv) - A(x_n+tv_n)\| \le M$, and the dominated convergence theorem that right side of Eq. (16.12) converges to $\int_0^1 A(x+tv)\, v\, dt$. Hence Eq. (16.11) is valid for all $x \in B(x_0,\varepsilon)$ and $v \in B(0,\varepsilon)$. We also see that

$$f(x+v) - f(x) - A(x)v = \varepsilon(v)v, \tag{16.13}$$

where $\varepsilon(v) := \int_0^1 [A(x+tv) - A(x)]\, dt$. Now

[1] It should be noted well, unlike in finite dimensions closed and bounded sets need not be compact, so it is not sufficient to choose ϵ sufficiently small so that $\overline{B(x_0, 2\epsilon)} \subset U$. Here is a counter example. Let $X \equiv H$ be a Hilbert space, $\{e_n\}_{n=1}^\infty$ be an orthonormal set. Define $f(x) \equiv \sum_{n=1}^\infty n\phi(\|x - e_n\|)$, where ϕ is any continuous function on $\mathbb{R}$ such that $\phi(0) = 1$ and ϕ is supported in $(-1,1)$. Notice that $\|e_n - e_m\|^2 = 2$ for all $m \neq n$, so that $\|e_n - e_m\| = \sqrt{2}$. Using this fact it is rather easy to check that for any $x_0 \in H$, there is an $\epsilon > 0$ such that for all $x \in B(x_0,\epsilon)$, only one term in the sum defining f is non-zero. Hence, f is continuous. However, $f(e_n) = n \to \infty$ as $n \to \infty$.

$$\begin{aligned}\|\varepsilon(v)\| &\leq \int_0^1 \|A(x+tv) - A(x)\| \, dt \\ &\leq \max_{t\in[0,1]} \|A(x+tv) - A(x)\| \to 0 \text{ as } v \to 0,\end{aligned}$$

by the continuity of A. Thus, we have shown that f is differentiable and that $Df(x) = A(x)$. ■

Corollary 16.15. *Suppose now that $X = \mathbb{R}^d$, $f : U \subset_o X \to Y$ be a continuous function such that $\partial_i f(x) := \partial_{e_i} f(x)$ exists and is continuous on U for $i = 1, 2, \ldots, d$, where $\{e_i\}_{i=1}^d$ is the standard basis for $\mathbb{R}^d$. Then $f \in C^1(U, Y)$ and $Df(x) e_i = \partial_i f(x)$ for all i.*

Proof. For $x \in U$, let $A(x) : \mathbb{R}^d \to Y$ be the unique linear map such that $A(x) e_i = \partial_i f(x)$ for $i = 1, 2, \ldots, d$. Then $A : U \to L(\mathbb{R}^d, Y)$ is a continuous map. Now let $v \in \mathbb{R}^d$ and $v^{(i)} := (v_1, v_2, \ldots, v_i, 0, \ldots, 0)$ for $i = 1, 2, \ldots, d$ and $v^{(0)} := 0$. Then for $t \in \mathbb{R}$ near 0, using the fundamental theorem of calculus and the definition of $\partial_i f(x)$,

$$\begin{aligned} f(x+tv) - f(x) &= \sum_{i=1}^d \left[f\left(x + tv^{(i)}\right) - f\left(x + tv^{(i-1)}\right)\right] \\ &= \sum_{i=1}^d \int_0^1 \frac{d}{ds} f\left(x + tv^{(i-1)} + stv_i e_i\right) ds \\ &= \sum_{i=1}^d tv_i \int_0^1 \partial_i f\left(x + tv^{(i-1)} + stv_i e_i\right) ds \\ &= \sum_{i=1}^d tv_i \int_0^1 A\left(x + tv^{(i-1)} + stv_i e_i\right) e_i ds. \end{aligned}$$

Using the continuity of A, it now follows that

$$\begin{aligned} \lim_{t\to 0} \frac{f(x+tv) - f(x)}{t} &= \sum_{i=1}^d v_i \lim_{t\to 0} \int_0^1 A\left(x + tv^{(i-1)} + stv_i e_i\right) e_i ds \\ &= \sum_{i=1}^d v_i \int_0^1 A(x) e_i ds = A(x) v \end{aligned}$$

which shows $\partial_v f(x)$ exists and $\partial_v f(x) = A(x) v$. The result now follows from an application of Proposition 16.14. ■

16.4 Higher Order Derivatives

It is somewhat inconvenient to work with the Banach spaces $\mathcal{L}^k(X, Y)$ in Definition 16.10. For this reason we will introduce an isomorphic Banach space, $M_k(X, Y)$.

Definition 16.16. *For $k \in \{1, 2, 3, \ldots\}$, let $M_k(X, Y)$ denote the set of functions $f : X^k \to Y$ such that*

1. *For $i \in \{1, 2, \ldots, k\}$, $v \in X \to f\langle v_1, v_2, \ldots, v_{i-1}, v, v_{i+1}, \ldots, v_k\rangle \in Y$ is linear* [2] *for all $\{v_i\}_{i=1}^n \subset X$.*
2. *The norm $\|f\|_{M_k(X,Y)}$ should be finite, where*

$$\|f\|_{M_k(X,Y)} := \sup\{\frac{\|f\langle v_1, v_2, \ldots, v_k\rangle\|_Y}{\|v_1\|\|v_2\| \cdots \|v_k\|} : \{v_i\}_{i=1}^k \subset X \setminus \{0\}\}.$$

Lemma 16.17. *There are linear operators $j_k : \mathcal{L}^k(X, Y) \to M_k(X, Y)$ defined inductively as follows: $j_1 = Id_{L(X,Y)}$ (notice that $M_1(X, Y) = \mathcal{L}^1(X, Y) = L(X, Y)$) and*

$$(j_{k+1}A)\langle v_0, v_1, \ldots, v_k\rangle = (j_k(Av_0))\langle v_1, v_2, \ldots, v_k\rangle \quad \forall v_i \in X.$$

(Notice that $Av_0 \in \mathcal{L}^k(X, Y)$.) Moreover, the maps j_k are isometric isomorphisms.

Proof. To get a feeling for what j_k is let us write out j_2 and j_3 explicitly. If $A \in \mathcal{L}^2(X, Y) = L(X, L(X, Y))$, then $(j_2A)\langle v_1, v_2\rangle = (Av_1)v_2$ and if $A \in \mathcal{L}^3(X, Y) = L(X, L(X, L(X, Y)))$, $(j_3A)\langle v_1, v_2, v_3\rangle = ((Av_1)v_2)v_3$ for all $v_i \in X$. It is easily checked that j_k is linear for all k. We will now show by induction that j_k is an isometry and in particular that j_k is injective. Clearly this is true if $k = 1$ since j_1 is the identity map. For $A \in \mathcal{L}^{k+1}(X, Y)$,

$$\begin{aligned}
&\|j_{k+1}A\|_{M_{k+1}(X,Y)} \\
&\quad := \sup\{\frac{\|(j_k(Av_0))\langle v_1, v_2, \ldots, v_k\rangle\|_Y}{\|v_0\|\|v_1\|\|v_2\| \cdots \|v_k\|} : \{v_i\}_{i=0}^k \subset X \setminus \{0\}\} \\
&\quad = \sup\{\frac{\|(j_k(Av_0))\|_{M_k(X,Y)}}{\|v_0\|} : v_0 \in X \setminus \{0\}\} \\
&\quad = \sup\{\frac{\|Av_0\|_{\mathcal{L}^k(X,Y)}}{\|v_0\|} : v_0 \in X \setminus \{0\}\} \\
&\quad = \|A\|_{L(X,\mathcal{L}^k(X,Y))} := \|A\|_{\mathcal{L}^{k+1}(X,Y)},
\end{aligned}$$

wherein the second to last inequality we have used the induction hypothesis. This shows that j_{k+1} is an isometry provided j_k is an isometry. To finish the proof it suffices to show that j_k is surjective for all k. Again this is true for $k = 1$. Suppose that j_k is invertible for some $k \geq 1$. Given $f \in M_{k+1}(X, Y)$ we must produce $A \in \mathcal{L}^{k+1}(X, Y) = L(X, \mathcal{L}^k(X, Y))$ such that $j_{k+1}A = f$. If such an equation is to hold, then for $v_0 \in X$, we would have $j_k(Av_0) = f\langle v_0, \cdots\rangle$. That is $Av_0 = j_k^{-1}(f\langle v_0, \cdots\rangle)$. It is easily checked that A so defined is linear, bounded, and $j_{k+1}A = f$. ■

From now on we will identify $\mathcal{L}^k$ with M_k without further mention. In particular, we will view $D^k f$ as function on U with values in $M_k(X, Y)$.

---

[2] I will routinely write $f\langle v_1, v_2, \ldots, v_k\rangle$ rather than $f(v_1, v_2, \ldots, v_k)$ when the function f depends on each of variables linearly, i.e. f is a multi-linear function.

Theorem 16.18 (Differentiability). *Suppose $k \in \{1, 2, \ldots\}$ and D is a dense subspace of X, $f : U \subset_o X \to Y$ is a function such that $(\partial_{v_1} \partial_{v_2} \cdots \partial_{v_l} f)(x)$ exists for all $x \in D \cap U$, $\{v_i\}_{i=1}^l \subset D$, and $l = 1, 2, \ldots k$. Further assume there exists continuous functions $A_l : U \subset_o X \to M_l(X, Y)$ such that such that $(\partial_{v_1} \partial_{v_2} \cdots \partial_{v_l} f)(x) = A_l(x)\langle v_1, v_2, \ldots, v_l\rangle$ for all $x \in D \cap U$, $\{v_i\}_{i=1}^l \subset D$, and $l = 1, 2, \ldots k$. Then $D^l f(x)$ exists and is equal to $A_l(x)$ for all $x \in U$ and $l = 1, 2, \ldots, k$.*

Proof. We will prove the theorem by induction on k. We have already proved the theorem when $k = 1$, see Proposition 16.14. Now suppose that $k > 1$ and that the statement of the theorem holds when k is replaced by $k-1$. Hence we know that $D^l f(x) = A_l(x)$ for all $x \in U$ and $l = 1, 2, \ldots, k-1$. We are also given that

$$(\partial_{v_1} \partial_{v_2} \cdots \partial_{v_k} f)(x) = A_k(x)\langle v_1, v_2, \ldots, v_k\rangle \quad \forall x \in U \cap D, \{v_i\} \subset D. \tag{16.14}$$

Now we may write $(\partial_{v_2} \cdots \partial_{v_k} f)(x)$ as $(D^{k-1} f)(x)\langle v_2, v_3, \ldots, v_k\rangle$ so that Eq. (16.14) may be written as

$$\begin{aligned} &\partial_{v_1}(D^{k-1} f)(x)\langle v_2, v_3, \ldots, v_k\rangle) \\ &\qquad = A_k(x)\langle v_1, v_2, \ldots, v_k\rangle \quad \forall x \in U \cap D, \{v_i\} \subset D. \end{aligned} \tag{16.15}$$

So by the fundamental theorem of calculus, we have that

$$\begin{aligned} &((D^{k-1} f)(x + v_1) - (D^{k-1} f)(x))\langle v_2, v_3, \ldots, v_k\rangle \\ &\qquad = \int_0^1 A_k(x + tv_1)\langle v_1, v_2, \ldots, v_k\rangle \, dt \end{aligned} \tag{16.16}$$

for all $x \in U \cap D$ and $\{v_i\} \subset D$ with v_1 sufficiently small. By the same argument given in the proof of Proposition 16.14, Eq. (16.16) remains valid for all $x \in U$ and $\{v_i\} \subset X$ with v_1 sufficiently small. We may write this last equation alternatively as,

$$(D^{k-1} f)(x + v_1) - (D^{k-1} f)(x) = \int_0^1 A_k(x + tv_1)\langle v_1, \cdots\rangle \, dt. \tag{16.17}$$

Hence

$$\begin{aligned} &(D^{k-1} f)(x + v_1) - (D^{k-1} f)(x) - A_k(x)\langle v_1, \cdots\rangle \\ &\qquad = \int_0^1 [A_k(x + tv_1) - A_k(x)]\langle v_1, \cdots\rangle \, dt \end{aligned}$$

from which we get the estimate,

$$\|(D^{k-1} f)(x + v_1) - (D^{k-1} f)(x) - A_k(x)\langle v_1, \cdots\rangle\| \le \varepsilon(v_1)\|v_1\| \tag{16.18}$$

where $\varepsilon(v_1) := \int_0^1 \|A_k(x + tv_1) - A_k(x)\| \, dt$. Notice by the continuity of A_k that $\varepsilon(v_1) \to 0$ as $v_1 \to 0$. Thus it follow from Eq. (16.18) that $D^{k-1} f$ is differentiable and that $(D^k f)(x) = A_k(x)$. ∎

Example 16.19. Let $f : GL(X, Y) \to GL(Y, X)$ be defined by $f(A) := A^{-1}$. We assume that $GL(X, Y)$ is not empty. Then f is infinitely differentiable and

$$\begin{aligned}(D^k f)(A)\langle V_1, V_2, \dots, V_k\rangle \\ &= (-1)^k \sum_\sigma \{B^{-1}V_{\sigma(1)}B^{-1}V_{\sigma(2)}B^{-1}\cdots B^{-1}V_{\sigma(k)}B^{-1}\}, \quad (16.19)\end{aligned}$$

where sum is over all permutations of σ of $\{1, 2, \dots, k\}$.

Let me check Eq. (16.19) in the case that $k = 2$. Notice that we have already shown that $(\partial_{V_1} f)(B) = Df(B)V_1 = -B^{-1}V_1B^{-1}$. Using the product rule we find that

$$(\partial_{V_2}\partial_{V_1} f)(B) = B^{-1}V_2B^{-1}V_1B^{-1} + B^{-1}V_1B^{-1}V_2B^{-1} =: A_2(B)\langle V_1, V_2\rangle.$$

Notice that $\|A_2(B)\langle V_1, V_2\rangle\| \le 2\|B^{-1}\|^3\|V_1\| \cdot \|V_2\|$, so that $\|A_2(B)\| \le 2\|B^{-1}\|^3 < \infty$. Hence $A_2 : GL(X, Y) \to M_2(L(X, Y), L(Y, X))$. Also

$$\begin{aligned}\|(A_2(B) - A_2(C))\langle V_1, V_2\rangle\| &\le 2\|B^{-1}V_2B^{-1}V_1B^{-1} - C^{-1}V_2C^{-1}V_1C^{-1}\| \\ &\le 2\|B^{-1}V_2B^{-1}V_1B^{-1} - B^{-1}V_2B^{-1}V_1C^{-1}\| \\ &\quad + 2\|B^{-1}V_2B^{-1}V_1C^{-1} - B^{-1}V_2C^{-1}V_1C^{-1}\| \\ &\quad\quad + 2\|B^{-1}V_2C^{-1}V_1C^{-1} - C^{-1}V_2C^{-1}V_1C^{-1}\| \\ &\le 2\|B^{-1}\|^2\|V_2\|\|V_1\|\|B^{-1} - C^{-1}\| \\ &\quad + 2\|B^{-1}\|\|C^{-1}\|\|V_2\|\|V_1\|\|B^{-1} - C^{-1}\| \\ &\quad\quad + 2\|C^{-1}\|^2\|V_2\|\|V_1\|\|B^{-1} - C^{-1}\|.\end{aligned}$$

This shows that

$$\|A_2(B) - A_2(C)\| \le 2\|B^{-1} - C^{-1}\|\{\|B^{-1}\|^2 + \|B^{-1}\|\|C^{-1}\| + \|C^{-1}\|^2\}.$$

Since $B \to B^{-1}$ is differentiable and hence continuous, it follows that $A_2(B)$ is also continuous in B. Hence by Theorem 16.18 $D^2 f(A)$ exists and is given as in Eq. (16.19)

Example 16.20. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a C^∞– function and $F(x) := \int_0^1 f(x(t))\,dt$ for $x \in X := C([0, 1], \mathbb{R})$ equipped with the norm $\|x\| := \max_{t\in[0,1]} |x(t)|$. Then $F : X \to \mathbb{R}$ is also infinitely differentiable and

$$(D^k F)(x)\langle v_1, v_2, \dots, v_k\rangle = \int_0^1 f^{(k)}(x(t))v_1(t)\cdots v_k(t)\,dt, \quad (16.20)$$

for all $x \in X$ and $\{v_i\} \subset X$.

To verify this example, notice that

$$(\partial_v F)(x) := \frac{d}{ds}|_0 F(x+sv) = \frac{d}{ds}|_0 \int_0^1 f(x(t)+sv(t))\,dt$$
$$= \int_0^1 \frac{d}{ds}|_0 f(x(t)+sv(t))\,dt = \int_0^1 f'(x(t))v(t)\,dt.$$

Similar computations show that

$$(\partial_{v_1}\partial_{v_2}\cdots\partial_{v_k} f)(x) = \int_0^1 f^{(k)}(x(t))v_1(t)\cdots v_k(t)\,dt =: A_k(x)\langle v_1, v_2, \dots, v_k\rangle.$$

Now for $x, y \in X$,

$$|A_k(x)\langle v_1, v_2, \dots, v_k\rangle - A_k(y)\langle v_1, v_2, \dots, v_k\rangle|$$
$$\le \int_0^1 |f^{(k)}(x(t)) - f^{(k)}(y(t))| \cdot |v_1(t)\cdots v_k(t)\,|dt$$
$$\le \prod_{i=1}^k \|v_i\| \int_0^1 |f^{(k)}(x(t)) - f^{(k)}(y(t))|dt,$$

which shows that

$$\|A_k(x) - A_k(y)\| \le \int_0^1 |f^{(k)}(x(t)) - f^{(k)}(y(t))|dt.$$

This last expression is easily seen to go to zero as $y \to x$ in X. Hence A_k is continuous. Thus we may apply Theorem 16.18 to conclude that Eq. (16.20) is valid.

16.5 Inverse and Implicit Function Theorems

In this section, let X be a Banach space, $R > 0$, $U = B = B(0,R) \subset X$ and $\varepsilon : U \to X$ be a continuous function such that $\varepsilon(0) = 0$. Our immediate goal is to give a sufficient condition on ε so that $F(x) := x + \varepsilon(x)$ is a homeomorphism from U to $F(U)$ with $F(U)$ being an open subset of X. Let's start by looking at the one dimensional case first. So for the moment assume that $X = \mathbb{R}$, $U = (-1,1)$, and $\varepsilon : U \to \mathbb{R}$ is C^1. Then F will be injective iff F is either strictly increasing or decreasing. Since we are thinking that F is a "small" perturbation of the identity function we will assume that F is strictly increasing, i.e. $F' = 1 + \varepsilon' > 0$. This positivity condition is not so easily interpreted for operators on a Banach space. However the condition that $|\varepsilon'| \le \alpha < 1$ is easily interpreted in the Banach space setting and it implies $1 + \varepsilon' > 0$.

Lemma 16.21. *Suppose that $U = B = B(0,R)$ $(R > 0)$ is a ball in X and $\varepsilon : B \to X$ is a C^1 function such that $\|D\varepsilon\| \le \alpha < \infty$ on U. Then*

$$\|\varepsilon(x) - \varepsilon(y)\| \le \alpha\|x - y\| \text{ for all } x, y \in U. \tag{16.21}$$

Proof. By the fundamental theorem of calculus and the chain rule:

$$\begin{aligned}\varepsilon(y)-\varepsilon(x) &= \int_0^1 \frac{d}{dt}\varepsilon(x+t(y-x))dt \\ &= \int_0^1 [D\varepsilon(x+t(y-x))](y-x)dt.\end{aligned}$$

Therefore, by the triangle inequality and the assumption that $\|D\varepsilon(x)\| \le \alpha$ on B,

$$\|\varepsilon(y)-\varepsilon(x)\| \le \int_0^1 \|D\varepsilon(x+t(y-x))\|dt \cdot \|(y-x)\| \le \alpha\|(y-x)\|.$$

■

Remark 16.22. It is easily checked that if $\varepsilon : U = B(0,R) \to X$ is C^1 and satisfies (16.21) then $\|D\varepsilon\| \le \alpha$ on U.

Using the above remark and the analogy to the one dimensional example, one is lead to the following proposition.

Proposition 16.23. *Suppose $\alpha \in (0,1)$, $R > 0$, $U = B(0,R) \subset_o X$ and $\varepsilon : U \to X$ is a continuous function such that $\varepsilon(0) = 0$ and*

$$\|\varepsilon(x)-\varepsilon(y)\| \le \alpha\|x-y\| \quad \forall\ x,y \in U. \tag{16.22}$$

Then $F : U \to X$ defined by $F(x) := x + \varepsilon(x)$ for $x \in U$ satisfies:

1. *F is an injective map and $G = F^{-1} : V := F(U) \to U$ is continuous.*
2. *If $x_0 \in U$, $z_0 = F(x_0)$ and $r > 0$ such the $B(x_0,r) \subset U$, then*

$$B(z_0,(1-\alpha)r) \subset F(B(x_0,r)) \subset B(z_0,(1+\alpha)r). \tag{16.23}$$

In particular, for all $r \le R$,

$$B(0,(1-\alpha)\,r) \subset F(B(0,r)) \subset B(0,(1+\alpha)\,r), \tag{16.24}$$

see Figure 16.1 below.

3. *$V := F(U)$ is open subset of X and $F : U \to V$ is a homeomorphism.*

Proof.

1. Using the definition of F and the estimate in Eq. (16.22),

$$\begin{aligned}\|x-y\| &= \|(F(x)-F(y)) - (\varepsilon(x)-\varepsilon(y))\| \\ &\le \|F(x)-F(y)\| + \|\varepsilon(x)-\varepsilon(y)\| \\ &\le \|F(x)-F(y)\| + \alpha\|(x-y)\|\end{aligned}$$

for all $x, y \in U$. This implies

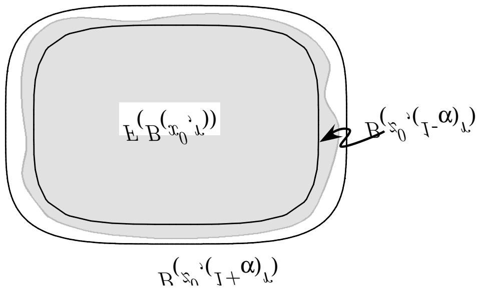


Fig. 16.1. Nesting of $F\left(B(x_0,r)\right)$ between $B(z_0,(1-\alpha)r)$ and $B(z_0,(1+\alpha)r)$.

$$\|x-y\| \le (1-\alpha)^{-1}\|F(x)-F(y)\| \tag{16.25}$$

which shows F is injective on U and hence shows the inverse function $G = F^{-1} : V := F(U) \to U$ is well defined. Moreover, replacing x, y in Eq. (16.25) by $G(x)$ and $G(y)$ respectively with $x, y \in V$ shows

$$\|G(x) - G(y)\| \le (1-\alpha)^{-1}\|x-y\| \text{ for all } x, y \in V. \tag{16.26}$$

Hence G is Lipschitz on V and hence continuous.

2. Let $x_0 \in U$, $r > 0$ and $z_0 = F(x_0) = x_0 + \varepsilon(x_0)$ be as in item 2. The second inclusion in Eq. (16.23) follows from the simple computation:

$$\begin{aligned}\|F(x_0+h) - z_0\| &= \|h + \varepsilon(x_0+h) - \varepsilon(x_0)\| \\ &\le \|h\| + \|\varepsilon(x_0+h) - \varepsilon(x_0)\| \\ &\le (1+\alpha)\|h\| < (1+\alpha)\, r\end{aligned}$$

for all $h \in B(0,r)$. To prove the first inclusion in Eq. (16.23) we must find, for every $z \in B(z_0,(1-\alpha)r)$, an $h \in B(0,r)$ such that $z = F(x_0+h)$ or equivalently an $h \in B(0,r)$ solving

$$z - z_0 = F(x_0+h) - F(x_0) = h + \varepsilon(x_0+h) - \varepsilon(x_0).$$

Let $k := z - z_0$ and for $h \in B(0,r)$, let $\delta(h) := \varepsilon(x_0+h) - \varepsilon(x_0)$. With this notation it suffices to show for each $k \in B(z_0,(1-\alpha)r)$ there exists $h \in B(0,r)$ such that $k = h + \delta(h)$. Notice that $\delta(0) = 0$ and

$$\|\delta(h_1) - \delta(h_2)\| = \|\varepsilon(x_0+h_1) - \varepsilon(x_0+h_2)\| \le \alpha\|h_1 - h_2\| \tag{16.27}$$

for all $h_1, h_2 \in B(0,r)$. We are now going to solve the equation $k = h + \delta(h)$ for h by the method of successive approximations starting with $h_0 = 0$ and then defining h_n inductively by

$$h_{n+1} = k - \delta(h_n). \tag{16.28}$$

A simple induction argument using Eq. (16.27) shows that

$$\|h_{n+1}-h_n\| \le \alpha^n \|k\| \text{ for all } n \in \mathbb{N}_0$$

and in particular that

$$\|h_N\| = \left\| \sum_{n=0}^{N-1} (h_{n+1}-h_n) \right\| \le \sum_{n=0}^{N-1} \|h_{n+1}-h_n\|$$
$$\le \sum_{n=0}^{N-1} \alpha^n \|k\| = \frac{1-\alpha^N}{1-\alpha} \|k\| . \tag{16.29}$$

Since $\|k\| < (1-\alpha) r$, this implies that $\|h_N\| < r$ for all N showing the approximation procedure is well defined. Let

$$h := \lim_{N\to\infty} h_n = \sum_{n=0}^{\infty} (h_{n+1}-h_n) \in X$$

which exists since the sum in the previous equation is absolutely convergent. Passing to the limit in Eqs. (16.29) and (16.28) shows that $\|h\| \le (1-\alpha)^{-1} \|k\| < r$ and $h = k - \delta(h)$, i.e. $h \in B(0,r)$ solves $k = h + \delta(h)$ as desired.

3. Given $x_0 \in U$, the first inclusion in Eq. (16.23) shows that $z_0 = F(x_0)$ is in the interior of $F(U)$. Since $z_0 \in F(U)$ was arbitrary, it follows that $V = F(U)$ is open. The continuity of the inverse function has already been proved in item 1.

■

For the remainder of this section let X and Y be two Banach spaces, $U \subset_o X$, $k \ge 1$, and $f \in C^k(U,Y)$.

Lemma 16.24. *Suppose $x_0 \in U$, $R > 0$ is such that $B^X(x_0,R) \subset U$ and $T : B^X(x_0,R) \to Y$ is a C^1 – function such that $T'(x_0)$ is invertible. Let*

$$\alpha(R) := \sup_{x\in B^X(x_0,R)} \left\|T'(x_0)^{-1}T'(x) - I\right\|_{L(X)} \tag{16.30}$$

and $\varepsilon \in C^1\left(B^X(0,R),X\right)$ be defined by

$$\varepsilon(h) = T'(x_0)^{-1}\left[T(x_0+h) - T(x_0)\right] - h \tag{16.31}$$

so that

$$T(x_0+h) = T(x_0) + T'(x_0)\left(h + \varepsilon(h)\right). \tag{16.32}$$

Then $\varepsilon(h) = o(h)$ as $h \to 0$ and

$$\|\varepsilon(h') - \varepsilon(h)\| \le \alpha(R) \|h'-h\| \text{ for all } h,h' \in B^X(0,R). \tag{16.33}$$

If $\alpha(R) < 1$ (which may be achieved by shrinking R if necessary), then $T'(x)$ is invertible for all $x \in B^X(x_0,R)$ and

$$\sup_{x\in B^X(x_0,R)} \left\|T'(x)^{-1}\right\|_{L(Y,X)} \le \frac{1}{1-\alpha(R)} \left\|T'(x_0)^{-1}\right\|_{L(Y,X)}. \tag{16.34}$$

Proof. By definition of $T'(x_0)$ and using $T'(x_0)^{-1}$ exists,

$$T(x_0+h)-T(x_0)=T'(x_0)h+o(h)$$

from which it follows that $\varepsilon(h)=o(h)$. In fact by the fundamental theorem of calculus,

$$\varepsilon(h)=\int_0^1 \left(T'(x_0)^{-1}T'(x_0+th)-I\right)hdt$$

but we will not use this here. Let $h,h'\in B^X(0,R)$ and apply the fundamental theorem of calculus to $t\to T(x_0+t(h'-h))$ to conclude

$$\begin{aligned}\varepsilon(h')-\varepsilon(h)&=T'(x_0)^{-1}\left[T(x_0+h')-T(x_0+h)\right]-(h'-h)\\&=\left[\int_0^1 \left(T'(x_0)^{-1}T'(x_0+t(h'-h))-I\right)dt\right](h'-h).\end{aligned}$$

Taking norms of this equation gives

$$\begin{aligned}\|\varepsilon(h')-\varepsilon(h)\|&\le\left[\int_0^1 \left\|T'(x_0)^{-1}T'(x_0+t(h'-h))-I\right\|dt\right]\|h'-h\|\\&\le\alpha(R)\|h'-h\|\end{aligned}$$

It only remains to prove Eq. (16.34), so suppose now that $\alpha(R)<1$. Then by Proposition 7.19, $T'(x_0)^{-1}T'(x)=I-\left(I-T'(x_0)^{-1}T'(x)\right)$ is invertible and

$$\left\|\left[T'(x_0)^{-1}T'(x)\right]^{-1}\right\|\le\frac{1}{1-\alpha(R)}\text{ for all }x\in B^X(x_0,R).$$

Since $T'(x)=T'(x_0)\left[T'(x_0)^{-1}T'(x)\right]$ this implies $T'(x)$ is invertible and

$$\left\|T'(x)^{-1}\right\|=\left\|\left[T'(x_0)^{-1}T'(x)\right]^{-1}T'(x_0)^{-1}\right\|\le\frac{1}{1-\alpha(R)}\left\|T'(x_0)^{-1}\right\|$$

for all $x\in B^X(x_0,R)$. ∎

Theorem 16.25 (Inverse Function Theorem). *Suppose $U\subset_o X$, $k\ge 1$ and $T\in C^k(U,Y)$ such that $T'(x)$ is invertible for all $x\in U$. Further assume $x_0\in U$ and $R>0$ such that $B^X(x_0,R)\subset U$.*

1. For all $r\le R$,

$$T(B^X(x_0,r))\subset T(x_0)+T'(x_0)B^X(0,(1+\alpha(r))r)\,. \tag{16.35}$$

2. If we further assume that

$$\alpha(R):=\sup_{x\in B^X(x_0,R)}\left\|T'(x_0)^{-1}T'(x)-I\right\|<1,$$

which may always be achieved by taking R sufficiently small, then

$$T(x_0)+T'(x_0)B^X(0,(1-\alpha(r))r)\subset T(B^X(x_0,r)) \tag{16.36}$$

for all $r\le R$, see Figure 16.2.

3. *$T : U \to Y$ is an open mapping, in particular $V := T(U) \subset_o Y$.*
4. *Again if R is sufficiently small so that $\alpha(R) < 1$, then $T|_{B^X(x_0,R)} : B^X(x_0,R) \to T(B^X(x_0,R))$ is invertible and $T|^{-1}_{B^X(x_0,R)} : T\left(B^X(x_0,R)\right) \to B^X(x_0,R)$ is a C^k – map.*
5. *If T is injective, then $T^{-1} : V \to U$ is also a C^k – map and*

$$\left(T^{-1}\right)'(y) = \left[T'(T^{-1}(y))\right]^{-1} \text{ for all } y \in V.$$

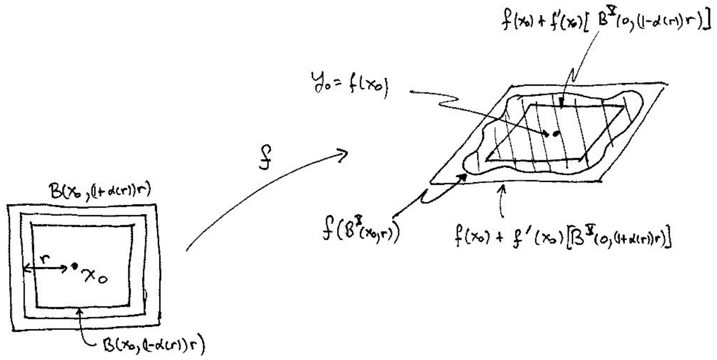


Fig. 16.2. The nesting of $T(B^X(x_0,r))$ between $T(x_0)+T'(x_0)\,B^X(0,(1-\alpha(r))r)$ and$T(x_0)+T'(x_0)\,B^X(0,(1+\alpha(r))r)$.

Proof. Let $\varepsilon \in C^1\left(B^X(0,R), X\right)$ be as defined in Eq. (16.31).

1. Using Eqs. (16.32) and (16.24),
$$\begin{aligned} T\left(B^X(x_0,r)\right) &= T(x_0) + T'(x_0)(I+\varepsilon)\left(B^X(0,r)\right) \\ &\subset T(x_0) + T'(x_0)\,B^X(0,(1+\alpha(r))\,r) \end{aligned} \tag{16.37}$$
which proves Eq. (16.35).
2. Now assume $\alpha(R) < 1$, then by Eqs. (16.37) and (16.24),
$$\begin{aligned} &T(x_0) + T'(x_0)\,B^X(0,(1-\alpha(r))\,r) \\ &\qquad \subset T(x_0) + T'(x_0)(I+\varepsilon)\left(B^X(0,r)\right) = T\left(B^X(x_0,r)\right) \end{aligned}$$
which proves Eq. (16.36).
3. Notice that $h \in X \to T(x_0) + T'(x_0)\,h \in Y$ is a homeomorphism. The fact that T is an open map follows easily from Eq. (16.36) which shows that $T(x_0)$ is interior of $T(W)$ for any $W \subset_o X$ with $x_0 \in W$.
4. The fact that $T|_{B^X(x_0,R)} : B^X(x_0,R) \to T(B^X(x_0,R))$ is invertible with a continuous inverse follows from Eq. (16.32) and Proposition 16.23. It

now follows from the converse to the chain rule, Theorem 16.7, that $g := T|_{B^X(x_0,R)}^{-1} : T\left(B^X(x_0,R)\right) \to B^X(x_0,R)$ is differentiable and

$$g'(y) = [T'(g(y))]^{-1} \text{ for all } y \in T\left(B^X(x_0,R)\right).$$

This equation shows g is C^1. Now suppose that $k \geq 2$. Since $T' \in C^{k-1}(B, L(X))$ and $i(A) := A^{-1}$ is a smooth map by Example 16.19, $g' = i \circ T' \circ g$ is C^1, i.e. g is C^2. If $k \geq 2$, we may use the same argument to now show g is C^3. Continuing this way inductively, we learn g is C^k.

5. Since differentiability and smoothness is local, the assertion in item 5. follows directly from what has already been proved.

■

Theorem 16.26 (Implicit Function Theorem). *Suppose that X, Y, and W are three Banach spaces, $k \geq 1$, $A \subset X \times Y$ is an open set, (x_0, y_0) is a point in A, and $f : A \to W$ is a C^k – map such $f(x_0, y_0) = 0$. Assume that $D_2 f(x_0, y_0) := D(f(x_0, \cdot))(y_0) : Y \to W$ is a bounded invertible linear transformation. Then there is an open neighborhood U_0 of x_0 in X such that for all connected open neighborhoods U of x_0 contained in U_0, there is a unique continuous function $u : U \to Y$ such that $u(x_0) = y_o$, $(x, u(x)) \in A$ and $f(x, u(x)) = 0$ for all $x \in U$. Moreover u is necessarily C^k and*

$$Du(x) = -D_2 f(x, u(x))^{-1} D_1 f(x, u(x)) \text{ for all } x \in U. \tag{16.38}$$

Proof. By replacing f by $(x, y) \to D_2 f(x_0, y_0)^{-1} f(x, y)$ if necessary, we may assume with out loss of generality that $W = Y$ and $D_2 f(x_0, y_0) = I_Y$. Define $F : A \to X \times Y$ by $F(x, y) := (x, f(x, y))$ for all $(x, y) \in A$. Notice that

$$DF(x, y) = \begin{bmatrix} I & D_1 f(x, y) \\ 0 & D_2 f(x, y) \end{bmatrix}$$

which is invertible iff $D_2 f(x, y)$ is invertible and if $D_2 f(x, y)$ is invertible then

$$DF(x, y)^{-1} = \begin{bmatrix} I & -D_1 f(x, y) D_2 f(x, y)^{-1} \\ 0 & D_2 f(x, y)^{-1} \end{bmatrix}.$$

Since $D_2 f(x_0, y_0) = I$ is invertible, the inverse function theorem guarantees that there exists a neighborhood U_0 of x_0 and V_0 of y_0 such that $U_0 \times V_0 \subset A$, $F(U_0 \times V_0)$ is open in $X \times Y$, $F|_{(U_0 \times V_0)}$ has a C^k–inverse which we call F^{-1}. Let $\pi_2(x, y) := y$ for all $(x, y) \in X \times Y$ and define C^k – function u_0 on U_0 by $u_0(x) := \pi_2 \circ F^{-1}(x, 0)$. Since $F^{-1}(x, 0) = (\tilde{x}, u_0(x))$ iff

$$(x, 0) = F(\tilde{x}, u_0(x)) = (\tilde{x}, f(\tilde{x}, u_0(x))),$$

it follows that $x = \tilde{x}$ and $f(x, u_0(x)) = 0$. Thus

$$(x, u_0(x)) = F^{-1}(x, 0) \in U_0 \times V_0 \subset A$$

and $f(x,u_0(x)) = 0$ for all $x \in U_0$. Moreover, u_0 is C^k being the composition of the C^k– functions, $x \to (x,0)$, F^{-1}, and π_2. So if $U \subset U_0$ is a connected set containing x_0, we may define $u := u_0|_U$ to show the existence of the functions u as described in the statement of the theorem. The only statement left to prove is the uniqueness of such a function u. Suppose that $u_1 : U \to Y$ is another continuous function such that $u_1(x_0) = y_0$, and $(x, u_1(x)) \in A$ and $f(x, u_1(x)) = 0$ for all $x \in U$. Let

$$O := \{x \in U | u(x) = u_1(x)\} = \{x \in U | u_0(x) = u_1(x)\}.$$

Clearly O is a (relatively) closed subset of U which is not empty since $x_0 \in O$. Because U is connected, if we show that O is also an open set we will have shown that $O = U$ or equivalently that $u_1 = u_0$ on U. So suppose that $x \in O$, i.e. $u_0(x) = u_1(x)$. For $\tilde{x}$ near $x \in U$,

$$0 = 0 - 0 = f(\tilde{x}, u_0(\tilde{x})) - f(\tilde{x}, u_1(\tilde{x})) = R(\tilde{x})(u_1(\tilde{x}) - u_0(\tilde{x})) \tag{16.39}$$

where

$$R(\tilde{x}) := \int_0^1 D_2 f((\tilde{x}, u_0(\tilde{x}) + t(u_1(\tilde{x}) - u_0(\tilde{x}))) dt. \tag{16.40}$$

From Eq. (16.40) and the continuity of u_0 and u_1, $\lim_{\tilde{x} \to x} R(\tilde{x}) = D_2 f(x, u_0(x))$ which is invertible.[3] Thus $R(\tilde{x})$ is invertible for all $\tilde{x}$ sufficiently close to x which combined with Eq. (16.39) implies that $u_1(\tilde{x}) = u_0(\tilde{x})$ for all $\tilde{x}$ sufficiently close to x. Since $x \in O$ was arbitrary, we have shown that O is open. ■

16.6 Smooth Dependence of ODE's on Initial Conditions*

In this subsection, let X be a Banach space, $U \subset_o X$ and J be an open interval with $0 \in J$.

Lemma 16.27. *If $Z \in C(J \times U, X)$ such that $D_x Z(t,x)$ exists for all $(t,x) \in J \times U$ and $D_x Z(t,x) \in C(J \times U, X)$ then Z is locally Lipschitz in x, see Definition 15.6.*

Proof. Suppose $I \sqsubset\sqsubset J$ and $x \in U$. By the continuity of DZ, for every $t \in I$ there an open neighborhood N_t of $t \in I$ and $\varepsilon_t > 0$ such that $B(x, \varepsilon_t) \subset U$ and

$$\sup \{\|D_x Z(t', x')\| : (t', x') \in N_t \times B(x, \varepsilon_t)\} < \infty.$$

By the compactness of I, there exists a finite subset $\Lambda \subset I$ such that $I \subset \cup_{t \in I} N_t$. Let $\varepsilon(x, I) := \min \{\varepsilon_t : t \in \Lambda\}$ and

[3] Notice that $DF(x, u_0(x))$ is invertible for all $x \in U_0$ since $F|_{U_0 \times V_0}$ has a C^1 inverse. Therefore $D_2 f(x, u_0(x))$ is also invertible for all $x \in U_0$.

$$K(x, I) := \sup \{ \|DZ(t, x')\| (t, x') \in I \times B(x, \varepsilon(x, I))\} < \infty.$$

Then by the fundamental theorem of calculus and the triangle inequality,

$$\begin{aligned}\|Z(t, x_1) - Z(t, x_0)\| &\leq \left(\int_0^1 \|D_x Z(t, x_0 + s(x_1 - x_0)\| \, ds \right) \|x_1 - x_0\| \\ &\leq K(x, I) \|x_1 - x_0\|\end{aligned}$$

for all $x_0, x_1 \in B(x, \varepsilon(x, I))$ and $t \in I$. ■

Theorem 16.28 (Smooth Dependence of ODE's on Initial Conditions). *Let X be a Banach space, $U \subset_o X$, $Z \in C(\mathbb{R} \times U, X)$ such that $D_x Z \in C(\mathbb{R} \times U, X)$ and $\phi : \mathcal{D}(Z) \subset \mathbb{R} \times X \to X$ denote the maximal solution operator to the ordinary differential equation*

$$\dot{y}(t) = Z(t, y(t)) \text{ with } y(0) = x \in U, \tag{16.41}$$

see Notation 15.9 and Theorem 15.15. Then $\phi \in C^1(\mathcal{D}(Z), U)$, $\partial_t D_x \phi(t, x)$ exists and is continuous for $(t, x) \in \mathcal{D}(Z)$ and $D_x \phi(t, x)$ satisfies the linear differential equation,

$$\frac{d}{dt} D_x \phi(t, x) = [(D_x Z)(t, \phi(t, x))] D_x \phi(t, x) \text{ with } D_x \phi(0, x) = I_X \tag{16.42}$$

for $t \in J_x$.

Proof. Let $x_0 \in U$ and J be an open interval such that $0 \in J \subset \bar{J} \sqsubset\sqsubset J_{x_0}$, $y_0 := y(\cdot, x_0)|_J$ and

$$\mathcal{O}_\varepsilon := \{ y \in BC(J, U) : \|y - y_0\|_\infty < \varepsilon \} \subset_o BC(J, X).$$

By Lemma 16.27, Z is locally Lipschitz and therefore Theorem 15.15 is applicable. By Eq. (15.23) of Theorem 15.15, there exists $\varepsilon > 0$ and $\delta > 0$ such that $G : B(x_0, \delta) \to \mathcal{O}_\varepsilon$ defined by $G(x) := \phi(\cdot, x)|_J$ is continuous. By Lemma 16.29 below, for $\varepsilon > 0$ sufficiently small the function $F : \mathcal{O}_\varepsilon \to BC(J, X)$ defined by

$$F(y) := y - \int_0^{\cdot} Z(t, y(t)) dt. \tag{16.43}$$

is C^1 and

$$DF(y)v = v - \int_0^{\cdot} D_y Z(t, y(t)) v(t) dt. \tag{16.44}$$

By the existence and uniqueness Theorem 8.21 for linear ordinary differential equations, $DF(y)$ is invertible for any $y \in BC(J, U)$. By the definition of ϕ, $F(G(x)) = h(x)$ for all $x \in B(x_0, \delta)$ where $h : X \to BC(J, X)$ is defined by $h(x)(t) = x$ for all $t \in J$, i.e. $h(x)$ is the constant path at x. Since h is a bounded linear map, h is smooth and $Dh(x) = h$ for all $x \in X$.

We may now apply the converse to the chain rule in Theorem 16.7 to conclude $G \in C^1(B(x_0,\delta),\mathcal{O})$ and $DG(x) = [DF(G(x))]^{-1}Dh(x)$ or equivalently, $DF(G(x))DG(x) = h$ which in turn is equivalent to

$$D_x\phi(t,x) - \int_0^t [DZ(\phi(\tau,x)]D_x\phi(\tau,x)\,d\tau = I_X.$$

As usual this equation implies $D_x\phi(t,x)$ is differentiable in t, $D_x\phi(t,x)$ is continuous in (t,x) and $D_x\phi(t,x)$ satisfies Eq. (16.42). ■

Lemma 16.29. *Continuing the notation used in the proof of Theorem 16.28 and further let*

$$f(y) := \int_0^{\cdot} Z(\tau, y(\tau))\,d\tau \text{ for } y \in \mathcal{O}_\varepsilon.$$

Then $f \in C^1(\mathcal{O}_\varepsilon, Y)$ and for all $y \in \mathcal{O}_\varepsilon$,

$$f'(y)h = \int_0^{\cdot} D_xZ(\tau,y(\tau))h(\tau)\,d\tau =: \Lambda_y h.$$

Proof. Let $h \in Y$ be sufficiently small and $\tau \in J$, then by fundamental theorem of calculus,

$$\begin{aligned} Z(\tau, &y(\tau)+h(\tau)) - Z(\tau,y(\tau)) \\ &= \int_0^1 [D_xZ(\tau,y(\tau)+rh(\tau)) - D_xZ(\tau,y(\tau))]dr \end{aligned}$$

and therefore,

$$\begin{aligned} f(y&+h) - f(y) - \Lambda_y h(t) \\ &= \int_0^t [Z(\tau,y(\tau)+h(\tau)) - Z(\tau,y(\tau)) - D_xZ(\tau,y(\tau))h(\tau)\,]\,d\tau \\ &= \int_0^t d\tau \int_0^1 dr[D_xZ(\tau,y(\tau)+rh(\tau)) - D_xZ(\tau,y(\tau))]h(\tau). \end{aligned}$$

Therefore,

$$\|(f(y+h) - f(y) - \Lambda_y h)\|_\infty \le \|h\|_\infty \delta(h) \tag{16.45}$$

where

$$\delta(h) := \int_J d\tau \int_0^1 dr\,\|D_xZ(\tau,y(\tau)+rh(\tau)) - D_xZ(\tau,y(\tau))\|\,.$$

With the aide of Lemmas 16.27 and Lemma 15.7,

$$(r,\tau,h) \in [0,1]\times J\times Y \to \|D_xZ(\tau,y(\tau)+rh(\tau))\|$$

is bounded for small h provided $\varepsilon > 0$ is sufficiently small. Thus it follows from the dominated convergence theorem that $\delta(h) \to 0$ as $h \to 0$ and hence Eq. (16.45) implies $f'(y)$ exists and is given by Λ_y. Similarly,

$$
\begin{aligned}
&||f'(y+h) - f'(y)||_{op} \\
&\qquad \leq \int_J \|D_x Z(\tau, y(\tau) + h(\tau)) - D_x Z(\tau, y(\tau))\| \, d\tau \to 0 \text{ as } h \to 0
\end{aligned}
$$

showing f' is continuous. ∎

Remark 16.30. If $Z \in C^k(U, X)$, then an inductive argument shows that $\phi \in C^k(\mathcal{D}(Z), X)$. For example if $Z \in C^2(U, X)$ then $(y(t), u(t)) := (\phi(t,x), D_x\phi(t,x))$ solves the ODE,

$$
\frac{d}{dt}(y(t), u(t)) = \tilde{Z}\left((y(t), u(t))\right) \text{ with } (y(0), u(0)) = (x, Id_X)
$$

where $\tilde{Z}$ is the C^1 – vector field defined by

$$
\tilde{Z}\,(x, u) = (Z(x), D_x Z(x) u)\,.
$$

Therefore Theorem 16.28 may be applied to this equation to deduce: $D_x^2\phi(t,x)$ and $D_x^2\dot{\phi}(t,x)$ exist and are continuous. We may now differentiate Eq. (16.42) to find $D_x^2\phi(t,x)$ satisfies the ODE,

$$
\begin{aligned}
\frac{d}{dt} D_x^2 \phi(t,x) &= [\left(\partial_{D_x\phi(t,x)} D_x Z\right)(t, \phi(t,x))] D_x\phi(t,x) \\
&\quad + [(D_x Z)\,(t, \phi(t,x))] D_x^2 \phi(t,x)
\end{aligned}
$$

with $D_x^2\phi(0,x) = 0$.

16.7 Existence of Periodic Solutions

A detailed discussion of the inverse function theorem on Banach and Frechét spaces may be found in Richard Hamilton's, "The Inverse Function Theorem of Nash and Moser." The applications in this section are taken from this paper. In what follows we say $f \in C^k_{2\pi}(\mathbb{R}, (c,d))$ if $f \in C^k_{2\pi}(\mathbb{R}, (c,d))$ and f is 2π – periodic, i.e. $f\,(x + 2\pi) = f\,(x)$ for all $x \in \mathbb{R}$.

Theorem 16.31 (Taken from Hamilton, p. 110.). *Let $p : U := (a,b) \to V := (c,d)$ be a smooth function with $p' > 0$ on (a,b). For every $g \in C^\infty_{2\pi}(\mathbb{R}, (c,d))$ there exists a unique function $y \in C^\infty_{2\pi}(\mathbb{R}, (a,b))$ such that*

$$
\dot{y}(t) + p(y(t)) = g(t).
$$

Proof. Let $\tilde{V} := C^0_{2\pi}(\mathbb{R}, (c,d)) \subset_o C^0_{2\pi}(\mathbb{R}, \mathbb{R})$ and $\tilde{U} \subset_o C^1_{2\pi}(\mathbb{R}, (a,b))$ be given by

$$
\tilde{U} := \left\{ y \in C^1_{2\pi}(\mathbb{R}, \mathbb{R}) : a < y(t) < b \;\&\; c < \dot{y}(t) + p(y(t)) < d \;\forall\; t \right\}.
$$

The proof will be completed by showing $P : \tilde{U} \to \tilde{V}$ defined by

$$P(y)(t) = \dot{y}(t) + p(y(t)) \text{ for } y \in \tilde{U} \text{ and } t \in \mathbb{R}$$

is bijective. Note that if $P(y)$ is smooth then so is y.

Step 1. The differential of P is given by $P'(y)h = \dot{h} + p'(y)h$, see Exercise 16.8. We will now show that the linear mapping $P'(y)$ is invertible. Indeed let $f = p'(y) > 0$, then the general solution to the Eq. $\dot{h} + fh = k$ is given by

$$h(t) = e^{-\int_0^t f(\tau)d\tau} h_0 + \int_0^t e^{-\int_\tau^t f(s)ds} k(\tau)d\tau$$

where h_0 is a constant. We wish to choose h_0 so that $h(2\pi) = h_0$, i.e. so that

$$h_0 \left(1 - e^{-c(f)}\right) = \int_0^{2\pi} e^{-\int_\tau^t f(s)ds} k(\tau)d\tau$$

where

$$c(f) = \int_0^{2\pi} f(\tau)d\tau = \int_0^{2\pi} p'(y(\tau))d\tau > 0.$$

The unique solution $h \in C^1_{2\pi}(\mathbb{R}, \mathbb{R})$ to $P'(y)h = k$ is given by

$$\begin{aligned} h(t) &= \left(1 - e^{-c(f)}\right)^{-1} e^{-\int_0^t f(\tau)d\tau} \int_0^{2\pi} e^{-\int_\tau^t f(s)ds} k(\tau)d\tau + \int_0^t e^{-\int_\tau^t f(s)ds} k(\tau)d\tau \\ &= \left(1 - e^{-c(f)}\right)^{-1} e^{-\int_0^t f(s)ds} \int_0^{2\pi} e^{-\int_\tau^t f(s)ds} k(\tau)d\tau + \int_0^t e^{-\int_\tau^t f(s)ds} k(\tau)d\tau. \end{aligned}$$

Therefore $P'(y)$ is invertible for all y. Hence by the inverse function Theorem 16.25, $P : \tilde{U} \to \tilde{V}$ is an open mapping which is locally invertible.

Step 2. Let us now prove $P : \tilde{U} \to \tilde{V}$ is injective. For this suppose $y_1, y_2 \in \tilde{U}$ such that $P(y_1) = g = P(y_2)$ and let $z = y_2 - y_1$. Since

$$\dot{z}(t) + p(y_2(t)) - p(y_1(t)) = g(t) - g(t) = 0,$$

if $t_m \in \mathbb{R}$ is point where $z(t_m)$ takes on its maximum, then $\dot{z}(t_m) = 0$ and hence

$$p(y_2(t_m)) - p(y_1(t_m)) = 0.$$

Since p is increasing this implies $y_2(t_m) = y_1(t_m)$ and hence $z(t_m) = 0$. This shows $z(t) \leq 0$ for all t and a similar argument using a minimizer of z shows $z(t) \geq 0$ for all t. So we conclude $y_1 = y_2$.

Step 3. Let $W := P(\tilde{U})$, we wish to show $W = \tilde{V}$. By step 1., we know W is an open subset of $\tilde{V}$ and since $\tilde{V}$ is connected, to finish the proof it suffices to show W is relatively closed in $\tilde{V}$. So suppose $y_j \in \tilde{U}$ such that $g_j := P(y_j) \to g \in \tilde{V}$. We must now show $g \in W$, i.e. $g = P(y)$ for some $y \in W$. If t_m is a maximizer of y_j, then $\dot{y}_j(t_m) = 0$ and hence $g_j(t_m) = p(y_j(t_m)) < d$ and therefore $y_j(t_m) < b$ because p is increasing. A similar argument works for the minimizers then allows us to conclude $\operatorname{Ran} p \circ y_j) \subset \operatorname{Ran} g_j) \sqsubset\sqsubset (c, d)$

for all j. Since g_j is converging uniformly to g, there exists $c < \gamma < \delta < d$ such that $\mathrm{Ran}(p \circ y_j) \subset \mathrm{Ran}(g_j) \subset [\gamma, \delta]$ for all j. Again since $p' > 0$,

$$\mathrm{Ran}(y_j) \subset p^{-1}([\gamma, \delta]) = [\alpha, \beta] \sqsubset\sqsubset (a, b) \text{ for all } j.$$

In particular $\sup\{|\dot{y}_j(t)| : t \in \mathbb{R} \text{ and } j\} < \infty$ since

$$\dot{y}_j(t) = g_j(t) - p(y_j(t)) \subset [\gamma, \delta] - [\gamma, \delta] \tag{16.46}$$

which is a compact subset of $\mathbb{R}$. The Ascoli-Arzela Theorem 11.29 now allows us to assume, by passing to a subsequence if necessary, that y_j is converging uniformly to $y \in C^0_{2\pi}(\mathbb{R}, [\alpha, \beta])$. It now follows that

$$\dot{y}_j(t) = g_j(t) - p(y_j(t)) \to g - p(y)$$

uniformly in t. Hence we concluded that $y \in C^1_{2\pi}(\mathbb{R}, \mathbb{R}) \cap C^0_{2\pi}(\mathbb{R}, [\alpha, \beta])$, $\dot{y}_j \to y$ and $P(y) = g$. This has proved that $g \in W$ and hence that W is relatively closed in $\tilde{V}$. ■

16.8 Contraction Mapping Principle

Some of the arguments uses in this chapter and in Chapter 15 may be abstracted to a general principle of finding fixed points on a complete metric space. This is the content of this chapter.

Theorem 16.32. *Suppose that (X, ρ) is a complete metric space and $S : X \to X$ is a contraction, i.e. there exists $\alpha \in (0, 1)$ such that $\rho(S(x), S(y)) \leq \alpha\rho(x, y)$ for all $x, y \in X$. Then S has a unique fixed point in X, i.e. there exists a unique point $x \in X$ such that $S(x) = x$.*

Proof. For uniqueness suppose that x and x' are two fixed points of S, then

$$\rho(x, x') = \rho(S(x), S(x')) \leq \alpha\rho(x, x').$$

Therefore $(1 - \alpha)\rho(x, x') \leq 0$ which implies that $\rho(x, x') = 0$ since $1 - \alpha > 0$. Thus $x = x'$. For existence, let $x_0 \in X$ be any point in X and define $x_n \in X$ inductively by $x_{n+1} = S(x_n)$ for $n \geq 0$. We will show that $x := \lim_{n\to\infty} x_n$ exists in X and because S is continuous this will imply,

$$x = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} S(x_n) = S(\lim_{n\to\infty} x_n) = S(x),$$

showing x is a fixed point of S. So to finish the proof, because X is complete, it suffices to show $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence in X. An easy inductive computation shows, for $n \geq 0$, that

$$\rho(x_{n+1}, x_n) = \rho(S(x_n), S(x_{n-1})) \leq \alpha\rho(x_n, x_{n-1}) \leq \cdots \leq \alpha^n\rho(x_1, x_0).$$

Another inductive argument using the triangle inequality shows, for $m > n$, that,

$$\rho(x_m, x_n) \leq \rho(x_m, x_{m-1}) + \rho(x_{m-1}, x_n) \leq \cdots \leq \sum_{k=n}^{m-1} \rho(x_{k+1}, x_k).$$

Combining the last two inequalities gives (using again that $\alpha \in (0,1)$),

$$\rho(x_m, x_n) \leq \sum_{k=n}^{m-1} \alpha^k \rho(x_1, x_0) \leq \rho(x_1, x_0)\alpha^n \sum_{l=0}^{\infty} \alpha^l = \rho(x_1, x_0)\frac{\alpha^n}{1-\alpha}.$$

This last equation shows that $\rho(x_m, x_n) \to 0$ as $m, n \to \infty$, i.e. $\{x_n\}_{n=0}^{\infty}$ is a Cauchy sequence. ∎

Corollary 16.33 (Contraction Mapping Principle II). *Suppose that (X, ρ) is a complete metric space and $S : X \to X$ is a continuous map such that $S^{(n)}$ is a contraction for some $n \in \mathbb{N}$. Here*

$$S^{(n)} := \overbrace{S \circ S \circ \ldots \circ S}^{n \text{ times}}$$

and we are assuming there exists $\alpha \in (0,1)$ such that $\rho(S^{(n)}(x), S^{(n)}(y)) \leq \alpha\rho(x, y)$ for all $x, y \in X$. Then S has a unique fixed point in X.

Proof. Let $T := S^{(n)}$, then $T : X \to X$ is a contraction and hence T has a unique fixed point $x \in X$. Since any fixed point of S is also a fixed point of T, we see if S has a fixed point then it must be x. Now

$$T(S(x)) = S^{(n)}(S(x)) = S(S^{(n)}(x)) = S(T(x)) = S(x),$$

which shows that $S(x)$ is also a fixed point of T. Since T has only one fixed point, we must have that $S(x) = x$. So we have shown that x is a fixed point of S and this fixed point is unique. ∎

Lemma 16.34. *Suppose that (X, ρ) is a complete metric space, $n \in \mathbb{N}$, Z is a topological space, and $\alpha \in (0,1)$. Suppose for each $z \in Z$ there is a map $S_z : X \to X$ with the following properties:*

Contraction property $\rho(S_z^{(n)}(x), S_z^{(n)}(y)) \leq \alpha\rho(x, y)$ for all $x, y \in X$ and $z \in Z$.

Continuity in z For each $x \in X$ the map $z \in Z \to S_z(x) \in X$ is continuous.

By Corollary 16.33 above, for each $z \in Z$ there is a unique fixed point $G(z) \in X$ of S_z.

Conclusion: *The map $G : Z \to X$ is continuous.*

Proof. Let $T_z := S_z^{(n)}$. If $z, w \in Z$, then

$$\begin{aligned}\rho(G(z), G(w)) &= \rho(T_z(G(z)), T_w(G(w))) \\ &\le \rho(T_z(G(z)), T_w(G(z))) + \rho(T_w(G(z)), T_w(G(w))) \\ &\le \rho(T_z(G(z)), T_w(G(z))) + \alpha\rho(G(z), G(w)).\end{aligned}$$

Solving this inequality for $\rho(G(z), G(w))$ gives

$$\rho(G(z), G(w)) \le \frac{1}{1-\alpha}\rho(T_z(G(z)), T_w(G(z))).$$

Since $w \to T_w(G(z))$ is continuous it follows from the above equation that $G(w) \to G(z)$ as $w \to z$, i.e. G is continuous. ∎

16.9 Exercises

Exercise 16.3. Suppose that $A : \mathbb{R} \to L(X)$ is a continuous function and $V : \mathbb{R} \to L(X)$ is the unique solution to the linear differential equation

$$\dot{V}(t) = A(t)V(t) \text{ with } V(0) = I. \tag{16.47}$$

Assuming that $V(t)$ is invertible for all $t \in \mathbb{R}$, show that $V^{-1}(t) := [V(t)]^{-1}$ must solve the differential equation

$$\frac{d}{dt}V^{-1}(t) = -V^{-1}(t)A(t) \text{ with } V^{-1}(0) = I. \tag{16.48}$$

See Exercise 8.13 as well.

Exercise 16.4 (Differential Equations with Parameters). Let W be another Banach space, $U \times V \subset_o X \times W$ and $Z \in C^1(U \times V, X)$. For each $(x, w) \in U \times V$, let $t \in J_{x,w} \to \phi(t, x, w)$ denote the maximal solution to the ODE

$$\dot{y}(t) = Z(y(t), w) \text{ with } y(0) = x \tag{16.49}$$

and

$$\mathcal{D} := \{(t, x, w) \in \mathbb{R} \times U \times V : t \in J_{x,w}\}$$

as in Exercise 15.8.

1. Prove that ϕ is C^1 and that $D_w\phi(t, x, w)$ solves the differential equation:

$$\frac{d}{dt}D_w\phi(t, x, w) = (D_xZ)(\phi(t, x, w), w)D_w\phi(t, x, w) + (D_wZ)(\phi(t, x, w), w)$$

with $D_w\phi(0, x, w) = 0 \in L(W, X)$. **Hint:** See the hint for Exercise 15.8 with the reference to Theorem 15.15 being replace by Theorem 16.28.

2. Also show with the aid of Duhamel's principle (Exercise 8.23) and Theorem 16.28 that

$$D_w\phi(t,x,w) = D_x\phi(t,x,w)\int_0^t D_x\phi(\tau,x,w)^{-1}(D_wZ)(\phi(\tau,x,w),w)d\tau$$

Exercise 16.5. (Differential of e^A) Let $f : L(X) \to GL(X)$ be the exponential function $f(A) = e^A$. Prove that f is differentiable and that

$$Df(A)B = \int_0^1 e^{(1-t)A}Be^{tA}\,dt. \tag{16.50}$$

Hint: Let $B \in L(X)$ and define $w(t,s) = e^{t(A+sB)}$ for all $t,s \in \mathbb{R}$. Notice that

$$dw(t,s)/dt = (A+sB)w(t,s) \text{ with } w(0,s) = I \in L(X). \tag{16.51}$$

Use Exercise 16.4 to conclude that w is C^1 and that $w'(t,0) := dw(t,s)/ds|_{s=0}$ satisfies the differential equation,

$$\frac{d}{dt}w'(t,0) = Aw'(t,0) + Be^{tA} \text{ with } w(0,0) = 0 \in L(X). \tag{16.52}$$

Solve this equation by Duhamel's principle (Exercise 8.23) and then apply Proposition 16.14 to conclude that f is differentiable with differential given by Eq. (16.50).

Exercise 16.6 (Local ODE Existence). Let S_x be defined as in Eq. (15.15) from the proof of Theorem 15.4. Verify that S_x satisfies the hypothesis of Corollary 16.33. In particular we could have used Corollary 16.33 to prove Theorem 15.4.

Exercise 16.7 (Local ODE Existence Again). Let $J = (-1,1)$, $Z \in C^1(X,X)$, $Y := BC(J,X)$ and for $y \in Y$ and $s \in J$ let $y_s \in Y$ be defined by $y_s(t) := y(st)$. Use the following outline to prove the ODE

$$\dot{y}(t) = Z(y(t)) \text{ with } y(0) = x \tag{16.53}$$

has a unique solution for small t and this solution is C^1 in x.

1. If y solves Eq. (16.53) then y_s solves

$$\dot{y}_s(t) = sZ(y_s(t)) \text{ with } y_s(0) = x$$

or equivalently

$$y_s(t) = x + s\int_0^t Z(y_s(\tau))d\tau. \tag{16.54}$$

Notice that when $s = 0$, the unique solution to this equation is $y_0(t) = x$.

2. Let $F : J \times Y \to J \times Y$ be defined by
$$F(s, y) := (s, y(t) - s \int_0^t Z(y(\tau))d\tau).$$
Show the differential of F is given by
$$F'(s, y)(a, v) = \left(a, t \to v(t) - s \int_0^t Z'(y(\tau))v(\tau)d\tau - a \int_0^{\cdot} Z(y(\tau))d\tau\right).$$
3. Verify $F'(0, y) : \mathbb{R} \times Y \to \mathbb{R} \times Y$ is invertible for all $y \in Y$ and notice that $F(0, y) = (0, y)$.
4. For $x \in X$, let $C_x \in Y$ be the constant path at x, i.e. $C_x(t) = x$ for all $t \in J$. Use the inverse function Theorem 16.25 to conclude there exists $\varepsilon > 0$ and a C^1 map $\phi : (-\varepsilon, \varepsilon) \times B(x_0, \varepsilon) \to Y$ such that
$$F(s, \phi(s, x)) = (s, C_x) \text{ for all } (s, x) \in (-\varepsilon, \varepsilon) \times B(x_0, \varepsilon).$$
5. Show, for $s \le \varepsilon$ that $y_s(t) := \phi(s, x)(t)$ satisfies Eq. (16.54). Now define $y(t, x) = \phi(\varepsilon/2, x)(2t/\varepsilon)$ and show $y(t, x)$ solve Eq. (16.53) for $|t| < \varepsilon/2$ and $x \in B(x_0, \varepsilon)$.

Exercise 16.8. Show P defined in Theorem 16.31 is continuously differentiable and $P'(y)h = \dot{h} + p'(y)h$.

Exercise 16.9. Embedded sub-manifold problems.

Exercise 16.10. Lagrange Multiplier problems.

16.9.1 Alternate construction of g. To be made into an exercise.

Suppose $U \subset_o X$ and $f : U \to Y$ is a C^2 – function. Then we are looking for a function $g(y)$ such that $f(g(y)) = y$. Fix an $x_0 \in U$ and $y_0 = f(x_0) \in Y$. Suppose such a g exists and let $x(t) = g(y_0 + th)$ for some $h \in Y$. Then differentiating $f(x(t)) = y_0 + th$ implies
$$\frac{d}{dt} f(x(t)) = f'(x(t))\dot{x}(t) = h$$
or equivalently that
$$\dot{x}(t) = [f'(x(t))]^{-1} h = Z(h, x(t)) \text{ with } x(0) = x_0 \tag{16.55}$$
where $Z(h, x) = [f'(x(t))]^{-1} h$. Conversely if x solves Eq. (16.55) we have $\frac{d}{dt} f(x(t)) = h$ and hence that
$$f(x(1)) = y_0 + h.$$
Thus if we define
$$g(y_0 + h) := e^{Z(h,\cdot)}(x_0),$$
then $f(g(y_0 + h)) = y_0 + h$ for all h sufficiently small. This shows f is an open mapping.

Part V

Lebesgue Integration Theory

17

Introduction: What are measures and why "measurable" sets

Definition 17.1 (Preliminary). *A* ***measure*** μ *"on" a set* X *is a function* $\mu : 2^X \to [0, \infty]$ *such that*

1. $\mu(\emptyset) = 0$
2. *If* $\{A_i\}_{i=1}^N$ *is a finite* ($N < \infty$) *or countable* ($N = \infty$) *collection of subsets of* X *which are pair-wise disjoint (i.e.* $A_i \cap A_j = \emptyset$ *if* $i \neq j$*) then*

$$\mu(\cup_{i=1}^N A_i) = \sum_{i=1}^N \mu(A_i).$$

Example 17.2. Suppose that X is any set and $x \in X$ is a point. For $A \subset X$, let

$$\delta_x(A) = \begin{cases} 1 \text{ if } & x \in A \\ 0 \text{ if } & x \notin A. \end{cases}$$

Then $\mu = \delta_x$ is a measure on X called the Dirac delta measure at x.

Example 17.3. Suppose that μ is a measure on X and $\lambda > 0$, then $\lambda \cdot \mu$ is also a measure on X. Moreover, if $\{\mu_\alpha\}_{\alpha \in J}$ are all measures on X, then $\mu = \sum_{\alpha \in J} \mu_\alpha$, i.e.

$$\mu(A) = \sum_{\alpha \in J} \mu_\alpha(A) \text{ for all } A \subset X$$

is a measure on X. (See Section 2 for the meaning of this sum.) To prove this we must show that μ is countably additive. Suppose that $\{A_i\}_{i=1}^\infty$ is a collection of pair-wise disjoint subsets of X, then

$$\begin{aligned} \mu(\cup_{i=1}^\infty A_i) &= \sum_{i=1}^\infty \mu(A_i) = \sum_{i=1}^\infty \sum_{\alpha \in J} \mu_\alpha(A_i) \\ &= \sum_{\alpha \in J} \sum_{i=1}^\infty \mu_\alpha(A_i) = \sum_{\alpha \in J} \mu_\alpha(\cup_{i=1}^\infty A_i) \\ &= \mu(\cup_{i=1}^\infty A_i) \end{aligned}$$

wherein the third equality we used Theorem 4.22 and in the fourth we used that fact that μ_α is a measure.

Example 17.4. Suppose that X is a set $\lambda : X \to [0,\infty]$ is a function. Then

$$\mu := \sum_{x\in X} \lambda(x)\delta_x$$

is a measure, explicitly

$$\mu(A) = \sum_{x\in A} \lambda(x)$$

for all $A \subset X$.

17.1 The problem with Lebesgue "measure"

So far all of the examples of measures given above are "counting" type measures, i.e. a weighted count of the number of points in a set. We certainly are going to want other types of measures too. In particular, it will be of great interest to have a measure on $\mathbb{R}$ (called Lebesgue measure) which measures the "length" of a subset of $\mathbb{R}$. Unfortunately as the next theorem shows, there is no such reasonable measure of length if we insist on measuring all subsets of $\mathbb{R}$.

Theorem 17.5. *There is no measure* $\mu : 2^{\mathbb{R}} \to [0,\infty]$ *such that*

1. $\mu([a,b)) = (b-a)$ *for all* $a < b$ *and*
2. *is translation invariant, i.e.* $\mu(A+x) = \mu(A)$ *for all* $x \in \mathbb{R}$ *and* $A \in 2^{\mathbb{R}}$, *where*

$$A + x := \{y + x : y \in A\} \subset \mathbb{R}.$$

In fact the theorem is still true even if (1) is replaced by the weaker condition that $0 < \mu((0,1]) < \infty$.

The counting measure $\mu(A) = \#(A)$ is translation invariant. However $\mu((0,1]) = \infty$ in this case and so μ does not satisfy condition 1.

Proof. First proof. Let us identify $[0,1)$ with the unit circle $S^1 := \{z \in \mathbb{C} : |z| = 1\}$ by the map

$$\phi(t) = e^{i2\pi t} = (\cos 2\pi t + i \sin 2\pi t) \in S^1$$

for $t \in [0,1)$. Using this identification we may use μ to define a function ν on 2^{S^1} by $\nu(\phi(A)) = \mu(A)$ for all $A \subset [0,1)$. This new function is a measure on S^1 with the property that $0 < \nu((0,1]) < \infty$. For $z \in S^1$ and $N \subset S^1$ let

$$zN := \{zn \in S^1 : n \in N\}, \tag{17.1}$$

that is to say $e^{i\theta}N$ is N rotated counter clockwise by angle θ. We now claim that ν is invariant under these rotations, i.e.

$$\nu(zN) = \nu(N) \tag{17.2}$$

for all $z \in S^1$ and $N \subset S^1$. To verify this, write $N = \phi(A)$ and $z = \phi(t)$ for some $t \in [0,1)$ and $A \subset [0,1)$. Then

$$\phi(t)\phi(A) = \phi(t + A \operatorname{mod} 1)$$

where for $A \subset [0,1)$ and $\alpha \in [0,1)$,

$$\begin{aligned} t + A \operatorname{mod} 1 &:= \{a + t \operatorname{mod} 1 \in [0,1) : a \in N\} \\ &= (a + A \cap \{a < 1 - t\}) \cup ((t-1) + A \cap \{a \geq 1 - t\})\,. \end{aligned}$$

Thus

$$\begin{aligned} \nu(\phi(t)\phi(A)) &= \mu(t + A \operatorname{mod} 1) \\ &= \mu\left((a + A \cap \{a < 1-t\}) \cup ((t-1) + A \cap \{a \geq 1-t\})\right) \\ &= \mu\left((a + A \cap \{a < 1-t\})\right) + \mu\left(((t-1) + A \cap \{a \geq 1-t\})\right) \\ &= \mu\left(A \cap \{a < 1-t\}\right) + \mu\left(A \cap \{a \geq 1-t\}\right) \\ &= \mu\left((A \cap \{a < 1-t\}) \cup (A \cap \{a \geq 1-t\})\right) \\ &= \mu(A) = \nu(\phi(A)). \end{aligned}$$

Therefore it suffices to prove that no finite non-trivial measure ν on S^1 such that Eq. (17.2) holds. To do this we will "construct" a non-measurable set $N = \phi(A)$ for some $A \subset [0,1)$. Let

$$R := \{z = e^{i2\pi t} : t \in \mathbb{Q}\} = \{z = e^{i2\pi t} : t \in [0,1) \cap \mathbb{Q}\}$$

– a countable subgroup of S^1. As above R acts on S^1 by rotations and divides S^1 up into equivalence classes, where $z, w \in S^1$ are equivalent if $z = rw$ for some $r \in R$. Choose (using the axiom of choice) one representative point n from each of these equivalence classes and let $N \subset S^1$ be the set of these representative points. Then every point $z \in S^1$ may be uniquely written as $z = nr$ with $n \in N$ and $r \in R$. That is to say

$$S^1 = \coprod_{r \in R} (rN) \tag{17.3}$$

where $\coprod_\alpha A_\alpha$ is used to denote the union of pair-wise disjoint sets $\{A_\alpha\}$. By Eqs. (17.2) and (17.3),

$$\nu(S^1) = \sum_{r \in R} \nu(rN) = \sum_{r \in R} \nu(N).$$

The right member from this equation is either 0 or ∞, 0 if $\nu(N)=0$ and ∞ if $\nu(N)>0$. In either case it is not equal $\nu(S^1)\in(0,1)$. Thus we have reached the desired contradiction. ■

Proof. Second proof of Theorem 17.5. For $N\subset[0,1)$ and $\alpha\in[0,1)$, let

$$\begin{aligned} N^\alpha &= N+\alpha \bmod 1\\ &= \{a+\alpha \bmod 1\in[0,1) : a\in N\}\\ &= (\alpha+N\cap\{a<1-\alpha\})\cup((\alpha-1)+N\cap\{a\geq 1-\alpha\}). \end{aligned}$$

Then

$$\begin{aligned} \mu(N^\alpha) &= \mu(\alpha+N\cap\{a<1-\alpha\})+\mu((\alpha-1)+N\cap\{a\geq 1-\alpha\})\\ &= \mu(N\cap\{a<1-\alpha\})+\mu(N\cap\{a\geq 1-\alpha\})\\ &= \mu(N\cap\{a<1-\alpha\}\cup(N\cap\{a\geq 1-\alpha\}))\\ &= \mu(N). \end{aligned} \tag{17.4}$$

We will now construct a bad set N which coupled with Eq. (17.4) will lead to a contradiction. Set

$$Q_x := \{x+r\in\mathbb{R} : r\in\mathbb{Q}\} = x+\mathbb{Q}.$$

Notice that $Q_x\cap Q_y\neq\emptyset$ implies that $Q_x=Q_y$. Let $\mathcal{O}=\{Q_x : x\in\mathbb{R}\}$ – the orbit space of the $\mathbb{Q}$ action. For all $A\in\mathcal{O}$ choose $f(A)\in[0,1/3)\cap A$[1] and define $N=f(\mathcal{O})$. Then observe:

1. $f(A)=f(B)$ implies that $A\cap B\neq\emptyset$ which implies that $A=B$ so that f is injective.
2. $\mathcal{O}=\{Q_n : n\in N\}$.

Let R be the countable set,

$$R := \mathbb{Q}\cap[0,1).$$

We now claim that

$$N^r\cap N^s=\emptyset \text{ if } r\neq s \text{ and} \tag{17.5}$$

$$[0,1)=\cup_{r\in R}N^r. \tag{17.6}$$

Indeed, if $x\in N^r\cap N^s\neq\emptyset$ then $x=r+n \bmod 1$ and $x=s+n' \bmod 1$, then $n-n'\in\mathbb{Q}$, i.e. $Q_n=Q_{n'}$. That is to say, $n=f(Q_n)=f(Q_{n'})=n'$ and hence that $s=r \bmod 1$, but $s,r\in[0,1)$ implies that $s=r$. Furthermore, if $x\in[0,1)$ and $n:=f(Q_x)$, then $x-n=r\in\mathbb{Q}$ and $x\in N^{r \bmod 1}$. Now that we have constructed N, we are ready for the contradiction. By Equations (17.4–17.6) we find

[1] We have used the Axiom of choice here, i.e. $\prod_{A\in\mathcal{F}}(A\cap[0,1/3])\neq\emptyset$

$$1 = \mu([0,1)) = \sum_{r \in R} \mu(N^r) = \sum_{r \in R} \mu(N)$$
$$= \begin{cases} \infty \text{ if } \mu(N) > 0 \\ 0 \text{ if } \mu(N) = 0 \end{cases}.$$

which is certainly inconsistent. Incidentally we have just produced an example of so called "non – measurable" set. ■

Because of Theorem 17.5, it is necessary to modify Definition 17.1. Theorem 17.5 points out that we will have to give up the idea of trying to measure all subsets of $\mathbb{R}$ but only measure some sub-collections of "measurable" sets. This leads us to the notion of σ – algebra discussed in the next chapter. Our revised notion of a measure will appear in Definition 19.1 of Chapter 19 below.

18

Measurability

18.1 Algebras and σ – Algebras

Definition 18.1. *A collection of subsets $\mathcal{A}$ of a set X is an **algebra** if*

1. $\emptyset, X \in \mathcal{A}$
2. *$A \in \mathcal{A}$ implies that $A^c \in \mathcal{A}$*
3. *$\mathcal{A}$ is closed under finite unions, i.e. if $A_1, \ldots, A_n \in \mathcal{A}$ then $A_1 \cup \cdots \cup A_n \in \mathcal{A}$.*
   In view of conditions 1. and 2., 3. is equivalent to

3′. *$\mathcal{A}$ is closed under finite intersections.*

Definition 18.2. *A collection of subsets $\mathcal{M}$ of X is a σ – **algebra** (or sometimes called a σ – field) if $\mathcal{M}$ is an algebra which also closed under countable unions, i.e. if $\{A_i\}_{i=1}^{\infty} \subset \mathcal{M}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{M}$. (Notice that since $\mathcal{M}$ is also closed under taking complements, $\mathcal{M}$ is also closed under taking countable intersections.) A pair $(X, \mathcal{M})$, where X is a set and $\mathcal{M}$ is a σ – algebra on X, is called a **measurable space.***

The reader should compare these definitions with that of a topology in Definition 10.1. Recall that the elements of a topology are called open sets. Analogously, elements of and algebra $\mathcal{A}$ or a σ – algebra $\mathcal{M}$ will be called **measurable** sets.

Example 18.3. Here are some examples of algebras.

1. $\mathcal{M} = 2^X$, then $\mathcal{M}$ is a topology, an algebra and a σ – algebra.
2. Let $X = \{1, 2, 3\}$, then $\tau = \{\emptyset, X, \{2, 3\}\}$ is a topology on X which is not an algebra.
3. $\tau = \mathcal{A} = \{\{1\}, \{2, 3\}, \emptyset, X\}$ is a topology, an algebra, and a σ – algebra on X. The sets X, $\{1\}$, $\{2, 3\}$, $\emptyset$ are open and closed. The sets $\{1, 2\}$ and $\{1, 3\}$ are neither open nor **closed** and are not measurable.

The reader should compare this example with Example 10.3.

Proposition 18.4. *Let $\mathcal{E}$ be any collection of subsets of X. Then there exists a unique smallest algebra $\mathcal{A}(\mathcal{E})$ and σ – algebra $\sigma(\mathcal{E})$ which contains $\mathcal{E}$.*

Proof. The proof is the same as the analogous Proposition 10.6 for topologies, i.e.

$$\mathcal{A}(\mathcal{E}) := \bigcap\{\mathcal{A} : \mathcal{A} \text{ is an algebra such that } \mathcal{E} \subset \mathcal{A}\}$$

and

$$\sigma(\mathcal{E}) := \bigcap\{\mathcal{M} : \mathcal{M} \text{ is a } \sigma - \text{algebra such that } \mathcal{E} \subset \mathcal{M}\}.$$

■

Example 18.5. Suppose $X = \{1, 2, 3\}$ and $\mathcal{E} = \{\emptyset, X, \{1, 2\}, \{1, 3\}\}$, see Figure 18.1.

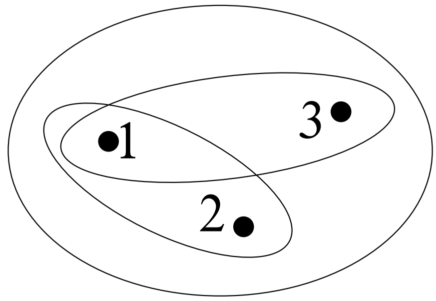


Fig. 18.1. A collection of subsets.

Then

$$\tau(\mathcal{E}) = \{\emptyset, X, \{1\}, \{1, 2\}, \{1, 3\}\}$$
$$\mathcal{A}(\mathcal{E}) = \sigma(\mathcal{E}) = 2^X.$$

The next proposition is the analogue to Proposition 10.7 for topologies and enables us to give and explicit descriptions of $\mathcal{A}(\mathcal{E})$. On the other hand it should be noted that $\sigma(\mathcal{E})$ typically does not admit a simple concrete description.

Proposition 18.6. *Let X be a set and $\mathcal{E} \subset 2^X$. Let $\mathcal{E}^c := \{A^c : A \in \mathcal{E}\}$ and $\mathcal{E}_c := \mathcal{E} \cup \{X, \emptyset\} \cup \mathcal{E}^c$ Then*

$$\mathcal{A}(\mathcal{E}) := \{\textit{finite unions of finite intersections of elements from } \mathcal{E}_c\}. \quad (18.1)$$

Proof. Let $\mathcal{A}$ denote the right member of Eq. (18.1). From the definition of an algebra, it is clear that $\mathcal{E} \subset \mathcal{A} \subset \mathcal{A}(\mathcal{E})$. Hence to finish that proof it suffices to show $\mathcal{A}$ is an algebra. The proof of these assertions are routine except for possibly showing that $\mathcal{A}$ is closed under complementation. To check $\mathcal{A}$ is closed under complementation, let $Z \in \mathcal{A}$ be expressed as

$$Z = \bigcup_{i=1}^{N} \bigcap_{j=1}^{K} A_{ij}$$

where $A_{ij} \in \mathcal{E}_c$. Therefore, writing $B_{ij} = A_{ij}^c \in \mathcal{E}_c$, we find that

$$Z^c = \bigcap_{i=1}^{N} \bigcup_{j=1}^{K} B_{ij} = \bigcup_{j_1,\dots,j_N=1}^{K} (B_{1j_1} \cap B_{2j_2} \cap \dots \cap B_{Nj_N}) \in \mathcal{A}$$

wherein we have used the fact that $B_{1j_1} \cap B_{2j_2} \cap \dots \cap B_{Nj_N}$ is a finite intersection of sets from $\mathcal{E}_c$. ∎

Remark 18.7. One might think that in general $\sigma(\mathcal{E})$ may be described as the countable unions of countable intersections of sets in $\mathcal{E}^c$. However this is in general **false,** since if

$$Z = \bigcup_{i=1}^{\infty} \bigcap_{j=1}^{\infty} A_{ij}$$

with $A_{ij} \in \mathcal{E}_c$, then

$$Z^c = \bigcup_{j_1=1, j_2=1, \dots j_N=1, \dots}^{\infty} \left(\bigcap_{\ell=1}^{\infty} A_{\ell, j_\ell}^c \right)$$

which is now an **uncountable** union. Thus the above description is not correct. In general it is complicated to explicitly describe $\sigma(\mathcal{E})$, see Proposition 1.23 on page 39 of Folland for details. Also see Proposition 18.13 below.

Exercise 18.1. Let τ be a topology on a set X and $\mathcal{A} = \mathcal{A}(\tau)$ be the algebra generated by τ. Show $\mathcal{A}$ is the collection of subsets of X which may be written as finite union of sets of the form $F \cap V$ where F is closed and V is open.

The following notion will be useful in the sequel and plays an analogous role for algebras as a base (Definition 10.8) does for a topology.

Definition 18.8. *A set $\mathcal{E} \subset 2^X$ is said to be an **elementary family or elementary class** provided that*

- $\emptyset \in \mathcal{E}$
- *$\mathcal{E}$ is closed under finite intersections*
- *if $E \in \mathcal{E}$, then E^c is a finite disjoint union of sets from $\mathcal{E}$. (In particular $X = \emptyset^c$ is a finite disjoint union of elements from $\mathcal{E}$.)*

Example 18.9. Let $X = \mathbb{R}$, then

$$\begin{aligned}\mathcal{E} &:= \left\{(a,b] \cap \mathbb{R} : a, b \in \bar{\mathbb{R}}\right\} \\ &= \{(a,b] : a \in [-\infty, \infty) \text{ and } a < b < \infty\} \cup \{\emptyset, \mathbb{R}\}\end{aligned}$$

is an elementary family.

Exercise 18.2. Let $\mathcal{A} \subset 2^X$ and $\mathcal{B} \subset 2^Y$ be elementary families. Show the collection

$$\mathcal{E} = \mathcal{A} \times \mathcal{B} = \{A \times B : A \in \mathcal{A} \text{ and } B \in \mathcal{B}\}$$

is also an elementary family.

Proposition 18.10. *Suppose $\mathcal{E} \subset 2^X$ is an elementary family, then $\mathcal{A} = \mathcal{A}(\mathcal{E})$ consists of sets which may be written as finite disjoint unions of sets from $\mathcal{E}$.*

Proof. This could be proved making use of Proposition 18.6. However it is easier to give a direct proof. Let $\mathcal{A}$ denote the collection of sets which may be written as finite disjoint unions of sets from $\mathcal{E}$. Clearly $\mathcal{E} \subset \mathcal{A} \subset \mathcal{A}(\mathcal{E})$ so it suffices to show $\mathcal{A}$ is an algebra since $\mathcal{A}(\mathcal{E})$ is the smallest algebra containing $\mathcal{E}$. By the properties of $\mathcal{E}$, we know that $\emptyset, X \in \mathcal{A}$. Now suppose that $A_i = \coprod_{F \in \Lambda_i} F \in \mathcal{A}$ where, for $i = 1, 2, \dots, n$, Λ_i is a finite collection of disjoint sets from $\mathcal{E}$. Then

$$\bigcap_{i=1}^n A_i = \bigcap_{i=1}^n \left(\coprod_{F \in \Lambda_i} F\right) = \bigcup_{(F_1, ,\dots, F_n) \in \Lambda_1 \times \dots \times \Lambda_n} (F_1 \cap F_2 \cap \dots \cap F_n)$$

and this is a disjoint (you check) union of elements from $\mathcal{E}$. Therefore $\mathcal{A}$ is closed under finite intersections. Similarly, if $A = \coprod_{F \in \Lambda} F$ with Λ being a finite collection of disjoint sets from $\mathcal{E}$, then $A^c = \bigcap_{F \in \Lambda} F^c$. Since by assumption $F^c \in \mathcal{A}$ for $F \in \Lambda \subset \mathcal{E}$ and $\mathcal{A}$ is closed under finite intersections, it follows that $A^c \in \mathcal{A}$. ∎

Definition 18.11. *Let X be a set. We say that a family of sets $\mathcal{F} \subset 2^X$ is a **partition** of X if distinct members of $\mathcal{F}$ are disjoint and if X is the union of the sets in $\mathcal{F}$.*

Example 18.12. Let X be a set and $\mathcal{E} = \{A_1, \dots, A_n\}$ where $A_1, \dots, A_n$ is a partition of X. In this case

$$\mathcal{A}(\mathcal{E}) = \sigma(\mathcal{E}) = \tau(\mathcal{E}) = \{\cup_{i \in \Lambda} A_i : \Lambda \subset \{1, 2, \dots, n\}\}$$

where $\cup_{i \in \Lambda} A_i := \emptyset$ when $\Lambda = \emptyset$. Notice that

$$\#\left(\mathcal{A}(\mathcal{E})\right) = \#(2^{\{1,2,\dots,n\}}) = 2^n.$$

Proposition 18.13. *Suppose that $\mathcal{M} \subset 2^X$ is a σ – algebra and $\mathcal{M}$ is at most a countable set. Then there exists a unique **finite** partition $\mathcal{F}$ of X such that $\mathcal{F} \subset \mathcal{M}$ and every element $B \in \mathcal{M}$ is of the form*

$$B = \cup \{A \in \mathcal{F} : A \subset B\}. \tag{18.2}$$

In particular $\mathcal{M}$ is actually a finite set and $\#(\mathcal{M}) = 2^n$ for some $n \in \mathbb{N}$.

Proof. For each $x \in X$ let

$$A_x = \cap \{A \in \mathcal{M} : x \in A\} \in \mathcal{M},$$

wherein we have used $\mathcal{M}$ is a countable σ – algebra to insure $A_x \in \mathcal{M}$. Hence A_x is the smallest set in $\mathcal{M}$ which contains x. Let $C = A_x \cap A_y$. If $x \notin C$ then $A_x \setminus C \subset A_x$ is an element of $\mathcal{M}$ which contains x and since A_x is the smallest member of $\mathcal{M}$ containing x, we must have that $C = \emptyset$. Similarly if $y \notin C$ then $C = \emptyset$. Therefore if $C \neq \emptyset$, then $x, y \in A_x \cap A_y \in \mathcal{M}$ and $A_x \cap A_y \subset A_x$ and $A_x \cap A_y \subset A_y$ from which it follows that $A_x = A_x \cap A_y = A_y$. This shows that $\mathcal{F} = \{A_x : x \in X\} \subset \mathcal{M}$ is a (necessarily countable) partition of X for which Eq. (18.2) holds for all $B \in \mathcal{M}$. Enumerate the elements of $\mathcal{F}$ as $\mathcal{F} = \{P_n\}_{n=1}^N$ where $N \in \mathbb{N}$ or $N = \infty$. If $N = \infty$, then the correspondence

$$a \in \{0,1\}^{\mathbb{N}} \to A_a = \cup\{P_n : a_n = 1\} \in \mathcal{M}$$

is bijective and therefore, by Lemma 2.6, $\mathcal{M}$ is uncountable. Thus any countable σ – algebra is necessarily finite. This finishes the proof modulo the uniqueness assertion which is left as an exercise to the reader. ■

Example 18.14. Let $X = \mathbb{R}$ and

$$\mathcal{E} = \{(a, \infty) : a \in \mathbb{R}\} \cup \{\mathbb{R}, \emptyset\} = \{(a, \infty) \cap \mathbb{R} : a \in \bar{\mathbb{R}}\} \subset 2^{\mathbb{R}}.$$

Notice that $\mathcal{E}_f = \mathcal{E}$ and that $\mathcal{E}$ is closed under unions, which shows that $\tau(\mathcal{E}) = \mathcal{E}$, i.e. $\mathcal{E}$ is already a topology. Since $(a, \infty)^c = (-\infty, a]$ we find that $\mathcal{E}_c = \{(a, \infty), (-\infty, a], -\infty \leq a < \infty\} \cup \{\mathbb{R}, \emptyset\}$. Noting that

$$(a, \infty) \cap (-\infty, b] = (a, b]$$

it follows that $\mathcal{A}(\mathcal{E}) = \mathcal{A}(\tilde{\mathcal{E}})$ where

$$\tilde{\mathcal{E}} := \left\{(a, b] \cap \mathbb{R} : a, b \in \bar{\mathbb{R}}\right\}.$$

Since $\tilde{\mathcal{E}}$ is an elementary family of subsets of $\mathbb{R}$, Proposition 18.10 implies $\mathcal{A}(\mathcal{E})$ may be described as being those sets which are finite disjoint unions of sets from $\tilde{\mathcal{E}}$. The σ – algebra, $\sigma(\mathcal{E})$, generated by $\mathcal{E}$ is **very complicated**. Here are some sets in $\sigma(\mathcal{E})$ – most of which are not in $\mathcal{A}(\mathcal{E})$.

(a) $(a, b) = \bigcup_{n=1}^{\infty} (a, b - \frac{1}{n}] \in \sigma(\mathcal{E})$.

(b) All of the standard open subsets of $\mathbb{R}$ are in $\sigma(\mathcal{E})$.
(c) $\{x\} = \bigcap_n \left(x - \frac{1}{n}, x\right] \in \sigma(\mathcal{E})$
(d) $[a,b] = \{a\} \cup (a,b] \in \sigma(\mathcal{E})$
(e) Any countable subset of $\mathbb{R}$ is in $\sigma(\mathcal{E})$.

Remark 18.15. In the above example, one may replace $\mathcal{E}$ by $\mathcal{E} = \{(a,\infty) : a \in \mathbb{Q}\} \cup \{\mathbb{R}, \emptyset\}$, in which case $\mathcal{A}(\mathcal{E})$ may be described as being those sets which are finite disjoint unions of sets from the following list

$$\{(a,\infty), (-\infty, a], (a,b] : a,b \in \mathbb{Q}\} \cup \{\emptyset, \mathbb{R}\}.$$

This shows that $\mathcal{A}(\mathcal{E})$ is a countable set – a useful fact which will be needed later.

Notation 18.16 *For a general topological space (X,τ), the* ***Borel σ – algebra*** *is the σ – algebra $\mathcal{B}_X := \sigma(\tau)$ on X. In particular if $X = \mathbb{R}^n$, $\mathcal{B}_{\mathbb{R}^n}$ will be used to denote the Borel σ – algebra on $\mathbb{R}^n$ when $\mathbb{R}^n$ is equipped with its standard Euclidean topology.*

Exercise 18.3. Verify the σ – algebra, $\mathcal{B}_{\mathbb{R}}$, is generated by any of the following collection of sets:

$$1.\ \{(a,\infty) : a \in \mathbb{R}\},\ 2.\ \{(a,\infty) : a \in \mathbb{Q}\}\ \text{or}\ 3.\ \{[a,\infty) : a \in \mathbb{Q}\}.$$

Proposition 18.17. *If τ is a second countable topology on X and $\mathcal{E}$ is a countable collection of subsets of X such that $\tau = \tau(\mathcal{E})$, then $\mathcal{B}_X := \sigma(\tau) = \sigma(\mathcal{E})$, i.e. $\sigma(\tau(\mathcal{E})) = \sigma(\mathcal{E})$.*

Proof. Let $\mathcal{E}_f$ denote the collection of subsets of X which are finite intersection of elements from $\mathcal{E}$ along with X and $\emptyset$. Notice that $\mathcal{E}_f$ is still countable (you prove). A set Z is in $\tau(\mathcal{E})$ iff Z is an arbitrary union of sets from $\mathcal{E}_f$. Therefore $Z = \bigcup_{A \in \mathcal{F}} A$ for some subset $\mathcal{F} \subset \mathcal{E}_f$ which is necessarily countable. Since $\mathcal{E}_f \subset \sigma(\mathcal{E})$ and $\sigma(\mathcal{E})$ is closed under countable unions it follows that $Z \in \sigma(\mathcal{E})$ and hence that $\tau(\mathcal{E}) \subset \sigma(\mathcal{E})$. Lastly, since $\mathcal{E} \subset \tau(\mathcal{E}) \subset \sigma(\mathcal{E})$, $\sigma(\mathcal{E}) \subset \sigma(\tau(\mathcal{E})) \subset \sigma(\mathcal{E})$. ■

18.2 Measurable Functions

Our notion of a "measurable" function will be analogous to that for a continuous function. For motivational purposes, suppose $(X, \mathcal{M}, \mu)$ is a measure space and $f : X \to \mathbb{R}_+$. Roughly speaking, in the next Chapter we are going to define $\int_X f d\mu$ as a certain limit of sums of the form,

$$\sum_{0<a_1<a_2<a_3<\dots}^{\infty} a_i \mu(f^{-1}(a_i, a_{i+1}]).$$

For this to make sense we will need to require $f^{-1}((a,b]) \in \mathcal{M}$ for all $a < b$. Because of Lemma 18.22 below, this last condition is equivalent to the condition $f^{-1}(\mathcal{B}_{\mathbb{R}}) \subset \mathcal{M}$.

Definition 18.18. *Let $(X, \mathcal{M})$ and $(Y, \mathcal{F})$ be measurable spaces. A function $f : X \to Y$ is* ***measurable*** *if $f^{-1}(\mathcal{F}) \subset \mathcal{M}$. We will also say that f is $\mathcal{M}/\mathcal{F}$ – measurable or $(\mathcal{M}, \mathcal{F})$ – measurable.*

Example 18.19 (Characteristic Functions). Let $(X, \mathcal{M})$ be a measurable space and $A \subset X$. We define the **characteristic function** $1_A : X \to \mathbb{R}$ by

$$1_A(x) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ if } x \notin A. \end{cases}$$

If $A \in \mathcal{M}$, then 1_A is $(\mathcal{M}, 2^{\mathbb{R}})$ – measurable because $1_A^{-1}(W)$ is either $\emptyset$, X, A or A^c for any $W \subset \mathbb{R}$. Conversely, if $\mathcal{F}$ is any σ – algebra on $\mathbb{R}$ containing a set $W \subset \mathbb{R}$ such that $1 \in W$ and $0 \in W^c$, then $A \in \mathcal{M}$ if 1_A is $(\mathcal{M}, \mathcal{F})$ – measurable. This is because $A = 1_A^{-1}(W) \in \mathcal{M}$.

Exercise 18.4. Suppose $f : X \to Y$ is a function, $\mathcal{F} \subset 2^Y$ and $\mathcal{M} \subset 2^X$. Show $f^{-1}\mathcal{F}$ and $f_*\mathcal{M}$ (see Notation 2.7) are algebras (σ – algebras) provided $\mathcal{F}$ and $\mathcal{M}$ are algebras (σ – algebras).

Remark 18.20. Let $f : X \to Y$ be a function. Given a σ – algebra $\mathcal{F} \subset 2^Y$, the σ – algebra $\mathcal{M} := f^{-1}(\mathcal{F})$ is the smallest σ – algebra on X such that f is $(\mathcal{M}, \mathcal{F})$ - measurable . Similarly, if $\mathcal{M}$ is a σ - algebra on X then $\mathcal{F} = f_*\mathcal{M}$ is the largest σ – algebra on Y such that f is $(\mathcal{M}, \mathcal{F})$ - measurable .

Recall from Definition 2.8 that for $\mathcal{E} \subset 2^X$ and $A \subset X$ that

$$\mathcal{E}_A = i_A^{-1}(\mathcal{E}) = \{A \cap E : E \in \mathcal{E}\}$$

where $i_A : A \to X$ is the inclusion map. Because of Exercise 10.3, when $\mathcal{E} = \mathcal{M}$ is an algebra (σ – algebra), $\mathcal{M}_A$ is an algebra (σ – algebra) on A and we call $\mathcal{M}_A$ the relative or induced algebra (σ – algebra) on A.

The next two Lemmas are direct analogues of their topological counter parts in Lemmas 10.13 and 10.14. For completeness, the proofs will be given even though they are same as those for Lemmas 10.13 and 10.14.

Lemma 18.21. *Suppose that $(X, \mathcal{M})$, $(Y, \mathcal{F})$ and $(Z, \mathcal{G})$ are measurable spaces. If $f : (X, \mathcal{M}) \to (Y, \mathcal{F})$ and $g : (Y, \mathcal{F}) \to (Z, \mathcal{G})$ are measurable functions then $g \circ f : (X, \mathcal{M}) \to (Z, \mathcal{G})$ is measurable as well.*

Proof. By assumption $g^{-1}(\mathcal{G}) \subset \mathcal{F}$ and $f^{-1}(\mathcal{F}) \subset \mathcal{M}$ so that

$$(g \circ f)^{-1}(\mathcal{G}) = f^{-1}\left(g^{-1}(\mathcal{G})\right) \subset f^{-1}(\mathcal{F}) \subset \mathcal{M}.$$

■

Lemma 18.22. *Suppose that $f : X \to Y$ is a function and $\mathcal{E} \subset 2^Y$ and $A \subset Y$ then*

$$\sigma\left(f^{-1}(\mathcal{E})\right) = f^{-1}(\sigma(\mathcal{E})) \text{ and} \tag{18.3}$$

$$(\sigma(\mathcal{E}))_A = \sigma(\mathcal{E}_A\ ). \tag{18.4}$$

(Similar assertion hold with $\sigma(\cdot)$ being replaced by $\mathcal{A}(\cdot)$.) Moreover, if $\mathcal{F} = \sigma(\mathcal{E})$ and $\mathcal{M}$ is a σ – algebra on X, then f is $(\mathcal{M}, \mathcal{F})$ – measurable iff $f^{-1}(\mathcal{E}) \subset \mathcal{M}$.

Proof. By Exercise 18.4, $f^{-1}(\sigma(\mathcal{E}))$ is a σ – algebra and since $\mathcal{E} \subset \mathcal{F}$, $f^{-1}(\mathcal{E}) \subset f^{-1}(\sigma(\mathcal{E}))$. It now follows that $\sigma\ (f^{-1}(\mathcal{E})) \subset f^{-1}(\sigma\ (\mathcal{E}))$. For the reverse inclusion, notice that

$$f_*\sigma\left(f^{-1}(\mathcal{E})\right) = \left\{B \subset Y : f^{-1}(B) \in \sigma\left(f^{-1}(\mathcal{E})\right)\right\}$$

is a σ – algebra which contains $\mathcal{E}$ and thus $\sigma(\mathcal{E}) \subset f_*\sigma\ \left(f^{-1}(\mathcal{E})\right)$. Hence if $B \in \sigma(\mathcal{E})$ we know that $f^{-1}(B) \in \sigma\left(f^{-1}(\mathcal{E})\right)$, i.e. $f^{-1}(\sigma(\mathcal{E})) \subset \sigma\left(f^{-1}(\mathcal{E})\right)$ and Eq. (18.3) has been proved. Applying Eq. (18.3) with $X = A$ and $f = i_A$ being the inclusion map implies

$$(\sigma(\mathcal{E}))_A = i_A^{-1}(\sigma(\mathcal{E})) = \sigma(i_A^{-1}(\mathcal{E})) = \sigma(\mathcal{E}_A).$$

Lastly if $f^{-1}\mathcal{E} \subset \mathcal{M}$, then $f^{-1}\sigma(\mathcal{E}) = \sigma\left(f^{-1}\mathcal{E}\right) \subset \mathcal{M}$ which shows f is $(\mathcal{M}, \mathcal{F})$ – measurable. ■

Corollary 18.23. *Suppose that $(X, \mathcal{M})$ is a measurable space. Then the following conditions on a function $f : X \to \mathbb{R}$ are equivalent:*

1. *f is $(\mathcal{M}, \mathcal{B}_\mathbb{R})$ – measurable,*
2. *$f^{-1}((a, \infty)) \in \mathcal{M}$ for all $a \in \mathbb{R}$,*
3. *$f^{-1}((a, \infty)) \in \mathcal{M}$ for all $a \in \mathbb{Q}$,*
4. *$f^{-1}((-\infty, a]) \in \mathcal{M}$ for all $a \in \mathbb{R}$.*

Proof. An exercise in using Lemma 18.22 and is the content of Exercise 18.8. ■

Here is yet another way to generate σ – algebras. (Compare with the analogous topological Definition 10.20.)

Definition 18.24 (σ – Algebras Generated by Functions). *Let X be a set and suppose there is a collection of measurable spaces $\{(Y_\alpha, \mathcal{F}_\alpha) : \alpha \in A\}$ and functions $f_\alpha : X \to Y_\alpha$ for all $\alpha \in A$. Let $\sigma(f_\alpha : \alpha \in A)$ denote the smallest σ – algebra on X such that each f_α is measurable, i.e.*

$$\sigma(f_\alpha : \alpha \in A) = \sigma(\cup_\alpha f_\alpha^{-1}(\mathcal{F}_\alpha)).$$

Proposition 18.25. *Assuming the notation in Definition 18.24 and additionally let $(Z, \mathcal{M})$ be a measurable space and $g : Z \to X$ be a function. Then g is $(\mathcal{M}, \sigma(f_\alpha : \alpha \in A))$ – measurable iff $f_\alpha \circ g$ is $(\mathcal{M}, \mathcal{F}_\alpha)$–measurable for all $\alpha \in A$.*

Proof. This proof is essentially the same as the proof of the topological analogue in Proposition 10.21. ($\Rightarrow$) If g is $(\mathcal{M}, \sigma(f_\alpha : \alpha \in A))$ – measurable, then the composition $f_\alpha \circ g$ is $(\mathcal{M}, \mathcal{F}_\alpha)$ – measurable by Lemma 18.21. ($\Leftarrow$) Let

$$\mathcal{G} = \sigma(f_\alpha : \alpha \in A) = \sigma\left(\cup_{\alpha \in A} f_\alpha^{-1}(\mathcal{F}_\alpha)\right).$$

If $f_\alpha \circ g$ is $(\mathcal{M}, \mathcal{F}_\alpha)$ – measurable for all α, then

$$g^{-1} f_\alpha^{-1}(\mathcal{F}_\alpha) \subset \mathcal{M} \,\forall\, \alpha \in A$$

and therefore

$$g^{-1}\left(\cup_{\alpha \in A} f_\alpha^{-1}(\mathcal{F}_\alpha)\right) = \cup_{\alpha \in A} g^{-1} f_\alpha^{-1}(\mathcal{F}_\alpha) \subset \mathcal{M}.$$

Hence

$$g^{-1}(\mathcal{G}) = g^{-1}\left(\sigma\left(\cup_{\alpha \in A} f_\alpha^{-1}(\mathcal{F}_\alpha)\right)\right) = \sigma(g^{-1}\left(\cup_{\alpha \in A} f_\alpha^{-1}(\mathcal{F}_\alpha)\right) \subset \mathcal{M}$$

which shows that g is $(\mathcal{M}, \mathcal{G})$ – measurable. ∎

Definition 18.26. *A function $f : X \to Y$ between two topological spaces is **Borel measurable** if $f^{-1}(\mathcal{B}_Y) \subset \mathcal{B}_X$.*

Proposition 18.27. *Let X and Y be two topological spaces and $f : X \to Y$ be a continuous function. Then f is Borel measurable.*

Proof. Using Lemma 18.22 and $\mathcal{B}_Y = \sigma(\tau_Y)$,

$$f^{-1}(\mathcal{B}_Y) = f^{-1}(\sigma(\tau_Y)) = \sigma(f^{-1}(\tau_Y)) \subset \sigma(\tau_X) = \mathcal{B}_X.$$

∎

Definition 18.28. *Given measurable spaces $(X, \mathcal{M})$ and $(Y, \mathcal{F})$ and a subset $A \subset X$. We say a function $f : A \to Y$ is measurable iff f is $\mathcal{M}_A/\mathcal{F}$ – measurable.*

Proposition 18.29 (Localizing Measurability). *Let $(X, \mathcal{M})$ and $(Y, \mathcal{F})$ be measurable spaces and $f : X \to Y$ be a function.*

1. *If f is measurable and $A \subset X$ then $f|_A : A \to Y$ is measurable.*
2. *Suppose there exist $A_n \in \mathcal{M}$ such that $X = \cup_{n=1}^\infty A_n$ and $f|A_n$ is $\mathcal{M}_{A_n}$ measurable for all n, then f is $\mathcal{M}$ – measurable.*

Proof. As the reader will notice, the proof given below is essentially identical to the proof of Proposition 10.19 which is the topological analogue of this proposition. 1. If $f : X \to Y$ is measurable, $f^{-1}(B) \in \mathcal{M}$ for all $B \in \mathcal{F}$ and therefore

$$f|_A^{-1}(B) = A \cap f^{-1}(B) \in \mathcal{M}_A \text{ for all } B \in \mathcal{F}.$$

2. If $B \in \mathcal{F}$, then

$$f^{-1}(B) = \cup_{n=1}^{\infty} \left(f^{-1}(B) \cap A_n \right) = \cup_{n=1}^{\infty} f|_{A_n}^{-1}(B).$$

Since each $A_n \in \mathcal{M}$, $\mathcal{M}_{A_n} \subset \mathcal{M}$ and so the previous displayed equation shows $f^{-1}(B) \in \mathcal{M}$. ∎

Proposition 18.30. *If $(X, \mathcal{M})$ is a measurable space, then*

$$f = (f_1, f_2, \dots, f_n) : X \to \mathbb{R}^n$$

is $(\mathcal{M}, \mathcal{B}_{\mathbb{R}^n})$ – measurable iff $f_i : X \to \mathbb{R}$ is $(\mathcal{M}, \mathcal{B}_{\mathbb{R}})$ – measurable for each i. In particular, a function $f : X \to \mathbb{C}$ is $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ – measurable iff $\operatorname{Re} f$ and $\operatorname{Im} f$ are $(\mathcal{M}, \mathcal{B}_{\mathbb{R}})$ – measurable.

Proof. This is formally a consequence of Corollary 18.65 and Proposition 18.60 below. Nevertheless it is instructive to give a direct proof now. Let $\tau = \tau_{\mathbb{R}^n}$ denote the usual topology on $\mathbb{R}^n$ and $\pi_i : \mathbb{R}^n \to \mathbb{R}$ be projection onto the i^{th} – factor. Since π_i is continuous, π_i is $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable and therefore if $f : X \to \mathbb{R}^n$ is measurable then so is $f_i = \pi_i \circ f$. Now suppose $f_i : X \to \mathbb{R}$ is measurable for all $i = 1, 2, \dots, n$. Let

$$\mathcal{E} := \{(a, b) : a, b \in \mathbb{Q}^n \ni a < b\},$$

where, for $a, b \in \mathbb{R}^n$, we write $a < b$ iff $a_i < b_i$ for $i = 1, 2, \dots, n$ and let

$$(a, b) = (a_1, b_1) \times \dots \times (a_n, b_n).$$

Since $\mathcal{E} \subset \tau$ and every element $V \in \tau$ may be written as a (necessarily) countable union of elements from $\mathcal{E}$, we have $\sigma(\mathcal{E}) \subset \mathcal{B}_{\mathbb{R}^n} = \sigma(\tau) \subset \sigma(\mathcal{E})$, i.e. $\sigma(\mathcal{E}) = \mathcal{B}_{\mathbb{R}^n}$. (This part of the proof is essentially a direct proof of Corollary 18.65 below.) Because

$$f^{-1}((a, b)) = f_1^{-1}((a_1, b_1)) \cap f_2^{-1}((a_2, b_2)) \cap \dots \cap f_n^{-1}((a_n, b_n)) \in \mathcal{M}$$

for all $a, b \in \mathbb{Q}$ with $a < b$, it follows that $f^{-1}\mathcal{E} \subset \mathcal{M}$ and therefore

$$f^{-1}\mathcal{B}_{\mathbb{R}^n} = f^{-1}\sigma(\mathcal{E}) = \sigma\left(f^{-1}\mathcal{E}\right) \subset \mathcal{M}.$$

∎

Corollary 18.31. *Let $(X, \mathcal{M})$ be a measurable space and $f, g : X \to \mathbb{C}$ be $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ – measurable functions. Then $f \pm g$ and $f \cdot g$ are also $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ – measurable.*

Proof. Define $F : X \to \mathbb{C} \times \mathbb{C}$, $A_{\pm} : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ and $M : \mathbb{C} \times \mathbb{C} \longrightarrow \mathbb{C}$ by $F(x) = (f(x), g(x))$, $A_{\pm}(w, z) = w \pm z$ and $M(w, z) = wz$. Then $A_{\pm}$ and M are continuous and hence $(\mathcal{B}_{\mathbb{C}^2}, \mathcal{B}_{\mathbb{C}})$ – measurable. Also F is $(\mathcal{M}, \mathcal{B}_{\mathbb{C}} \otimes \mathcal{B}_{\mathbb{C}}) = (\mathcal{M}, \mathcal{B}_{\mathbb{C}^2})$ – measurable since $\pi_1 \circ F = f$ and $\pi_2 \circ F = g$ are $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ – measurable. Therefore $A_{\pm} \circ F = f \pm g$ and $M \circ F = f \cdot g$, being the composition of measurable functions, are also measurable. ∎

Lemma 18.32. *Let $\alpha \in \mathbb{C}$, $(X, \mathcal{M})$ be a measurable space and $f : X \to \mathbb{C}$ be a $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ – measurable function. Then*

$$F(x) := \begin{cases} \frac{1}{f(x)} & \text{if} \quad f(x) \neq 0 \\ \alpha & \text{if} \quad f(x) = 0 \end{cases}$$

is measurable.

Proof. Define $i : \mathbb{C} \to \mathbb{C}$ by

$$i(z) = \begin{cases} \frac{1}{z} \text{ if } & z \neq 0 \\ 0 \text{ if } & z = 0. \end{cases}$$

For any open set $V \subset \mathbb{C}$ we have

$$i^{-1}(V) = i^{-1}(V \setminus \{0\}) \cup i^{-1}(V \cap \{0\})$$

Because i is continuous except at $z = 0$, $i^{-1}(V \setminus \{0\})$ is an open set and hence in $\mathcal{B}_{\mathbb{C}}$. Moreover, $i^{-1}(V \cap \{0\}) \in \mathcal{B}_{\mathbb{C}}$ since $i^{-1}(V \cap \{0\})$ is either the empty set or the one point set $\{0\}$. Therefore $i^{-1}(\tau_{\mathbb{C}}) \subset \mathcal{B}_{\mathbb{C}}$ and hence $i^{-1}(\mathcal{B}_{\mathbb{C}}) = i^{-1}(\sigma(\tau_{\mathbb{C}})) = \sigma(i^{-1}(\tau_{\mathbb{C}})) \subset \mathcal{B}_{\mathbb{C}}$ which shows that i is Borel measurable. Since $F = i \circ f$ is the composition of measurable functions, F is also measurable. ■

We will often deal with functions $f : X \to \bar{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. When talking about measurability in this context we will refer to the σ – algebra on $\bar{\mathbb{R}}$ defined by

$$\mathcal{B}_{\bar{\mathbb{R}}} := \sigma\left(\{[a, \infty] : a \in \mathbb{R}\}\right). \tag{18.5}$$

Proposition 18.33 (The Structure of $\mathcal{B}_{\bar{\mathbb{R}}}$). *Let $\mathcal{B}_{\mathbb{R}}$ and $\mathcal{B}_{\bar{\mathbb{R}}}$ be as above, then*

$$\mathcal{B}_{\bar{\mathbb{R}}} = \{A \subset \bar{\mathbb{R}} : A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}\}. \tag{18.6}$$

In particular $\{\infty\}, \{-\infty\} \in \mathcal{B}_{\bar{\mathbb{R}}}$ and $\mathcal{B}_{\mathbb{R}} \subset \mathcal{B}_{\bar{\mathbb{R}}}$.

Proof. Let us first observe that

$$\begin{aligned} \{-\infty\} &= \cap_{n=1}^{\infty}[-\infty, -n) = \cap_{n=1}^{\infty}[-n, \infty]^c \in \mathcal{B}_{\bar{\mathbb{R}}}, \\ \{\infty\} &= \cap_{n=1}^{\infty}[n, \infty] \in \mathcal{B}_{\bar{\mathbb{R}}} \text{ and } \mathbb{R} = \bar{\mathbb{R}} \setminus \{\pm\infty\} \in \mathcal{B}_{\bar{\mathbb{R}}}. \end{aligned}$$

Letting $i : \mathbb{R} \to \bar{\mathbb{R}}$ be the inclusion map,

$$\begin{aligned} i^{-1}\left(\mathcal{B}_{\bar{\mathbb{R}}}\right) &= \sigma\left(i^{-1}\left(\left\{[a, \infty] : a \in \bar{\mathbb{R}}\right\}\right)\right) = \sigma\left(\left\{i^{-1}\left([a, \infty]\right) : a \in \bar{\mathbb{R}}\right\}\right) \\ &= \sigma\left(\left\{[a, \infty] \cap \mathbb{R} : a \in \bar{\mathbb{R}}\right\}\right) = \sigma\left(\{[a, \infty) : a \in \mathbb{R}\}\right) = \mathcal{B}_{\mathbb{R}}. \end{aligned}$$

Thus we have shown

$$\mathcal{B}_{\mathbb{R}} = i^{-1}\left(\mathcal{B}_{\bar{\mathbb{R}}}\right) = \{A \cap \mathbb{R} : A \in \mathcal{B}_{\bar{\mathbb{R}}}\}.$$

This implies:

1. $A \in \mathcal{B}_{\bar{\mathbb{R}}} \implies A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}$ and
2. if $A \subset \bar{\mathbb{R}}$ is such that $A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}$ there exists $B \in \mathcal{B}_{\bar{\mathbb{R}}}$ such that $A \cap \mathbb{R} = B \cap \mathbb{R}$. Because $A \Delta B \subset \{\pm\infty\}$ and $\{\infty\}, \{-\infty\} \in \mathcal{B}_{\bar{\mathbb{R}}}$ we may conclude that $A \in \mathcal{B}_{\bar{\mathbb{R}}}$ as well.

This proves Eq. (18.6). ■

The proofs of the next two corollaries are left to the reader, see Exercises 18.5 and 18.6.

Corollary 18.34. *Let $(X, \mathcal{M})$ be a measurable space and $f : X \to \bar{\mathbb{R}}$ be a function. Then the following are equivalent*

1. *f is $(\mathcal{M}, \mathcal{B}_{\bar{\mathbb{R}}})$ - measurable,*
2. *$f^{-1}((a, \infty]) \in \mathcal{M}$ for all $a \in \mathbb{R}$,*
3. *$f^{-1}((-\infty, a]) \in \mathcal{M}$ for all $a \in \mathbb{R}$,*
4. *$f^{-1}(\{-\infty\}) \in \mathcal{M}$, $f^{-1}(\{\infty\}) \in \mathcal{M}$ and $f^0 : X \to \mathbb{R}$ defined by*

$$f^0(x) := 1_{\mathbb{R}}(f(x)) = \begin{cases} f(x) & \text{if } \quad f(x) \in \mathbb{R} \\ 0 & \text{if } f(x) \in \{\pm\infty\} \end{cases}$$

is measurable.

Corollary 18.35. *Let $(X, \mathcal{M})$ be a measurable space, $f, g : X \to \bar{\mathbb{R}}$ be functions and define $f \cdot g : X \to \bar{\mathbb{R}}$ and $(f+g) : X \to \bar{\mathbb{R}}$ using the conventions, $0 \cdot \infty = 0$ and $(f+g)(x) = 0$ if $f(x) = \infty$ and $g(x) = -\infty$ or $f(x) = -\infty$ and $g(x) = \infty$. Then $f \cdot g$ and $f+g$ are measurable functions on X if both f and g are measurable.*

Exercise 18.5. Prove Corollary 18.34 noting that the equivalence of items 1. – 3. is a direct analogue of Corollary 18.23. Use Proposition 18.33 to handle item 4.

Exercise 18.6. Prove Corollary 18.35.

Proposition 18.36 (Closure under sups, infs and limits). *Suppose that $(X, \mathcal{M})$ is a measurable space and $f_j : (X, \mathcal{M}) \to \overline{\mathbb{R}}$ for $j \in \mathbb{N}$ is a sequence of $\mathcal{M}/\mathcal{B}_{\overline{\mathbb{R}}}$ – measurable functions. Then*

$$\sup\nolimits_j f_j, \quad \inf\nolimits_j f_j, \quad \limsup_{j\to\infty} f_j \text{ and } \liminf_{j\to\infty} f_j$$

are all $\mathcal{M}/\mathcal{B}_{\overline{\mathbb{R}}}$ – measurable functions. (Note that this result is in generally false when $(X, \mathcal{M})$ is a topological space and measurable is replaced by continuous in the statement.)

Proof. Define $g_+(x) := \sup_j f_j(x)$, then

$$\begin{aligned} \{x : g_+(x) \le a\} &= \{x : f_j(x) \le a \ \forall j\} \\ &= \cap_j \{x : f_j(x) \le a\} \in \mathcal{M} \end{aligned}$$

so that g_+ is measurable. Similarly if $g_-(x) = \inf_j f_j(x)$ then

$$\{x : g_-(x) \geq a\} = \cap_j \{x : f_j(x) \geq a\} \in \mathcal{M}.$$

Since

$$\limsup_{j\to\infty} f_j = \inf_n \sup \{f_j : j \geq n\} \text{ and}$$
$$\liminf_{j\to\infty} f_j = \sup_n \inf \{f_j : j \geq n\}$$

we are done by what we have already proved. ■

Definition 18.37. *Given a function* $f : X \to \bar{\mathbb{R}}$ *let* $f_+(x) := \max\{f(x), 0\}$ *and* $f_-(x) := \max(-f(x), 0) = -\min(f(x), 0)$. *Notice that* $f = f_+ - f_-$.

Corollary 18.38. *Suppose* $(X, \mathcal{M})$ *is a measurable space and* $f : X \to \bar{\mathbb{R}}$ *is a function. Then* f *is measurable iff* $f_\pm$ *are measurable.*

Proof. If f is measurable, then Proposition 18.36 implies $f_\pm$ are measurable. Conversely if $f_\pm$ are measurable then so is $f = f_+ - f_-$. ■

18.2.1 More general pointwise limits

Lemma 18.39. *Suppose that* $(X, \mathcal{M})$ *is a measurable space,* (Y, d) *is a metric space and* $f_j : X \to Y$ *is* $(\mathcal{M}, \mathcal{B}_Y)$ – *measurable for all* j. *Also assume that for each* $x \in X$, $f(x) = \lim_{n\to\infty} f_n(x)$ *exists. Then* $f : X \to Y$ *is also* $(\mathcal{M}, \mathcal{B}_Y)$ – *measurable.*

Proof. Let $V \in \tau_d$ and $W_m := \{y \in Y : d_{V^c}(y) > 1/m\}$ for $m = 1, 2, \ldots$. Then $W_m \in \tau_d$,

$$W_m \subset \bar{W}_m \subset \{y \in Y : d_{V^c}(y) \geq 1/m\} \subset V$$

for all m and $W_m \uparrow V$ as $m \to \infty$. The proof will be completed by verifying the identity,

$$f^{-1}(V) = \cup_{m=1}^\infty \cup_{N=1}^\infty \cap_{n\geq N} f_n^{-1}(W_m) \in \mathcal{M}.$$

If $x \in f^{-1}(V)$ then $f(x) \in V$ and hence $f(x) \in W_m$ for some m. Since $f_n(x) \to f(x)$, $f_n(x) \in W_m$ for almost all n. That is $x \in \cup_{m=1}^\infty \cup_{N=1}^\infty \cap_{n\geq N} f_n^{-1}(W_m)$. Conversely when $x \in \cup_{m=1}^\infty \cup_{N=1}^\infty \cap_{n\geq N} f_n^{-1}(W_m)$ there exists an m such that $f_n(x) \in W_m \subset \bar{W}_m$ for almost all n. Since $f_n(x) \to f(x) \in \bar{W}_m \subset V$, it follows that $x \in f^{-1}(V)$. ■

Remark 18.40. In the previous Lemma 18.39 it is possible to let (Y, τ) be any topological space which has the "regularity" property that if $V \in \tau$ there exists $W_m \in \tau$ such that $W_m \subset \bar{W}_m \subset V$ and $V = \cup_{m=1}^\infty W_m$. Moreover, some extra condition is necessary on the topology τ in order for Lemma 18.39 to be correct. For example if $Y = \{1, 2, 3\}$ and $\tau = \{Y, \emptyset, \{1, 2\}, \{2, 3\}, \{2\}\}$ as in Example 10.36 and $X = \{a, b\}$ with the trivial σ – algebra. Let $f_j(a) = f_j(b) = 2$ for all j, then f_j is constant and hence measurable. Let $f(a) = 1$ and $f(b) = 2$, then $f_j \to f$ as $j \to \infty$ with f being non-measurable. Notice that the Borel σ – algebra on Y is 2^Y.

18.3 σ – Function Algebras

In this subsection, we are going to relate σ – algebras of subsets of a set X to certain algebras of functions on X. We will begin this endeavor after proving the simple but very useful approximation Theorem 18.42 below.

Definition 18.41. *Let $(X, \mathcal{M})$ be a measurable space. A function $\phi : X \to \mathbb{F}$ ($\mathbb{F}$ denotes either $\mathbb{R}$, $\mathbb{C}$ or $[0,\infty] \subset \bar{\mathbb{R}}$) is a* ***simple function*** *if ϕ is $\mathcal{M}$ – $\mathcal{B}_{\mathbb{F}}$ measurable and $\phi(X)$ contains only finitely many elements.*

Any such simple functions can be written as

$$\phi = \sum_{i=1}^{n} \lambda_i 1_{A_i} \text{ with } A_i \in \mathcal{M} \text{ and } \lambda_i \in \mathbb{F}. \tag{18.7}$$

Indeed, take $\lambda_1, \lambda_2, \ldots, \lambda_n$ to be an enumeration of the range of ϕ and $A_i = \phi^{-1}(\{\lambda_i\})$. Note that this argument shows that any simple function may be written intrinsically as

$$\phi = \sum_{y \in \mathbb{F}} y 1_{\phi^{-1}(\{y\})}. \tag{18.8}$$

The next theorem shows that simple functions are "pointwise dense" in the space of measurable functions.

Theorem 18.42 (Approximation Theorem). *Let $f : X \to [0,\infty]$ be measurable and define, see Figure 18.2,*

$$\begin{aligned}\phi_n(x) &:= \sum_{k=0}^{2^{2n}-1} \frac{k}{2^n} 1_{f^{-1}\left(\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right)}(x) + 2^n 1_{f^{-1}((2^n,\infty])}(x) \\ &= \sum_{k=0}^{2^{2n}-1} \frac{k}{2^n} 1_{\left\{\frac{k}{2^n} < f \le \frac{k+1}{2^n}\right\}}(x) + 2^n 1_{\{f > 2^n\}}(x)\end{aligned}$$

then $\phi_n \le f$ for all n, $\phi_n(x) \uparrow f(x)$ for all $x \in X$ and $\phi_n \uparrow f$ uniformly on the sets $X_M := \{x \in X : f(x) \le M\}$ with $M < \infty$. Moreover, if $f : X \to \mathbb{C}$ is a measurable function, then there exists simple functions ϕ_n such that $\lim_{n\to\infty} \phi_n(x) = f(x)$ for all x and $|\phi_n| \uparrow |f|$ as $n \to \infty$.

Proof. Since

$$\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right] = \left(\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}}\right] \cup \left(\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}\right],$$

if $x \in f^{-1}\left(\left(\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}}\right]\right)$ then $\phi_n(x) = \phi_{n+1}(x) = \frac{2k}{2^{n+1}}$ and if $x \in f^{-1}\left(\left(\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}\right]\right)$ then $\phi_n(x) = \frac{2k}{2^{n+1}} < \frac{2k+1}{2^{n+1}} = \phi_{n+1}(x)$. Similarly

$$(2^n, \infty] = (2^n, 2^{n+1}] \cup (2^{n+1}, \infty],$$

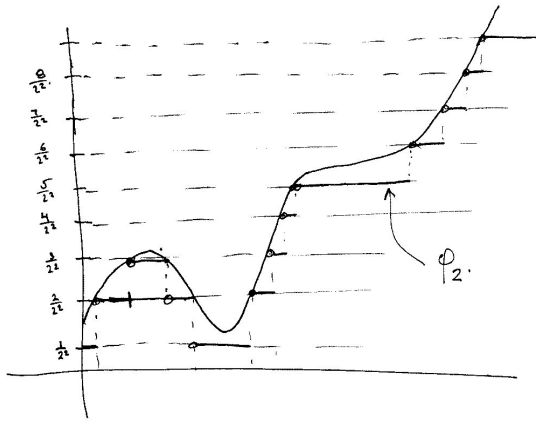


Fig. 18.2. Constructing simple functions approximating a function, $f : X \to [0, \infty]$.

and so for $x \in f^{-1}((2^{n+1}, \infty])$, $\phi_n(x) = 2^n < 2^{n+1} = \phi_{n+1}(x)$ and for $x \in f^{-1}((2^n, 2^{n+1}])$, $\phi_{n+1}(x) \geq 2^n = \phi_n(x)$. Therefore $\phi_n \leq \phi_{n+1}$ for all n. It is clear by construction that $\phi_n(x) \leq f(x)$ for all x and that $0 \leq f(x) - \phi_n(x) \leq 2^{-n}$ if $x \in X_{2^n}$. Hence we have shown that $\phi_n(x) \uparrow f(x)$ for all $x \in X$ and $\phi_n \uparrow f$ uniformly on bounded sets. For the second assertion, first assume that $f : X \to \mathbb{R}$ is a measurable function and choose $\phi_n^{\pm}$ to be simple functions such that $\phi_n^{\pm} \uparrow f_{\pm}$ as $n \to \infty$ and define $\phi_n = \phi_n^+ - \phi_n^-$. Then

$$|\phi_n| = \phi_n^+ + \phi_n^- \leq \phi_{n+1}^+ + \phi_{n+1}^- = |\phi_{n+1}|$$

and clearly $|\phi_n| = \phi_n^+ + \phi_n^- \uparrow f_+ + f_- = |f|$ and $\phi_n = \phi_n^+ - \phi_n^- \to f_+ - f_- = f$ as $n \to \infty$. Now suppose that $f : X \to \mathbb{C}$ is measurable. We may now choose simple function u_n and v_n such that $|u_n| \uparrow |\mathrm{Re}\, f|$, $|v_n| \uparrow |\mathrm{Im}\, f|$, $u_n \to \mathrm{Re}\, f$ and $v_n \to \mathrm{Im}\, f$ as $n \to \infty$. Let $\phi_n = u_n + iv_n$, then

$$|\phi_n|^2 = u_n^2 + v_n^2 \uparrow |\mathrm{Re}\, f|^2 + |\mathrm{Im}\, f|^2 = |f|^2$$

and $\phi_n = u_n + iv_n \to \mathrm{Re}\, f + i \,\mathrm{Im}\, f = f$ as $n \to \infty$. ■

For the rest of this section let X be a given set.

Definition 18.43 (Bounded Convergence). *We say that a sequence of functions f_n from X to $\mathbb{R}$ or $\mathbb{C}$ **converges boundedly** to a function f if $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in X$ and*

$$\sup\{|f_n(x)| : x \in X \text{ and } n = 1, 2, \ldots\} < \infty.$$

Definition 18.44. *A **function algebra** $\mathcal{H}$ on X is a linear subspace of $\ell^\infty(X, \mathbb{R})$ which contains 1 and is closed under pointwise multiplication, i.e. $\mathcal{H}$ is a subalgebra of $\ell^\infty(X, \mathbb{R})$ which contains 1. If $\mathcal{H}$ is further closed under bounded convergence then $\mathcal{H}$ is said to be a σ – **function algebra.***

Example 18.45. Suppose $\mathcal{M}$ is a σ – algebra on X, then

$$\ell^{\infty}(\mathcal{M},\mathbb{R}) := \{f \in \ell^{\infty}(X,\mathbb{R}) : f \text{ is } \mathcal{M}/\mathcal{B}_{\mathbb{R}} - \text{measurable}\} \tag{18.9}$$

is a σ – function algebra. The next theorem will show that these are the only example of σ – function algebras. (See Exercise 18.7 below for examples of function algebras on X.)

Notation 18.46 *If $\mathcal{H} \subset \ell^{\infty}(X,\mathbb{R})$ be a function algebra, let*

$$\mathcal{M}(\mathcal{H}) := \{A \subset X : 1_A \in \mathcal{H}\}. \tag{18.10}$$

Theorem 18.47. *Let $\mathcal{H}$ be a σ – function algebra on a set X. Then*

1. *$\mathcal{M}(\mathcal{H})$ is a σ – algebra on X.*
2. *$\mathcal{H} = \ell^{\infty}(\mathcal{M}(\mathcal{H}),\mathbb{R})$.*
3. *The map*
$$\mathcal{M} \in \{\sigma - \textit{algebras on } X\} \to \ell^{\infty}(\mathcal{M},\mathbb{R}) \in \{\sigma - \textit{function algebras on } X\} \tag{18.11}$$
is bijective with inverse given by $\mathcal{H} \to \mathcal{M}(\mathcal{H})$.

Proof. Let $\mathcal{M} := \mathcal{M}(\mathcal{H})$.

1. Since $0, 1 \in \mathcal{H}$, $\emptyset, X \in \mathcal{M}$. If $A \in \mathcal{M}$ then, since $\mathcal{H}$ is a linear subspace of $\ell^{\infty}(X,\mathbb{R})$, $1_{A^c} = 1 - 1_A \in \mathcal{H}$ which shows $A^c \in \mathcal{M}$. If $\{A_n\}_{n=1}^{\infty} \subset \mathcal{M}$, then since $\mathcal{H}$ is an algebra,
$$1_{\cap_{n=1}^{N} A_n} = \prod_{n=1}^{N} 1_{A_n} =: f_N \in \mathcal{H}$$
for all $N \in \mathbb{N}$. Because $\mathcal{H}$ is closed under bounded convergence it follows that
$$1_{\cap_{n=1}^{\infty} A_n} = \lim_{N\to\infty} f_N \in \mathcal{H}$$
and this implies $\cap_{n=1}^{\infty} A_n \in \mathcal{M}$. Hence we have shown $\mathcal{M}$ is a σ – algebra.
2. Since $\mathcal{H}$ is an algebra, $p(f) \in \mathcal{H}$ for any $f \in \mathcal{H}$ and any polynomial p on $\mathbb{R}$. The Weierstrass approximation Theorem 8.34, asserts that polynomials on $\mathbb{R}$ are uniformly dense in the space of continuos functions on any compact subinterval of $\mathbb{R}$. Hence if $f \in \mathcal{H}$ and $\phi \in C(\mathbb{R})$, there exists polynomials p_n on $\mathbb{R}$ such that $p_n \circ f(x)$ converges to $\phi \circ f(x)$ uniformly (and hence boundedly) in $x \in X$ as $n \to \infty$. Therefore $\phi \circ f \in \mathcal{H}$ for all $f \in \mathcal{H}$ and $\phi \in C(\mathbb{R})$ and in particular $|f| \in \mathcal{H}$ and $f_{\pm} := \frac{|f| \pm f}{2} \in \mathcal{H}$ if $f \in \mathcal{H}$. Fix an $\alpha \in \mathbb{R}$ and for $n \in \mathbb{N}$ let $\phi_n(t) := (t-\alpha)_+^{1/n}$, where $(t-\alpha)_+ := \max\{t-\alpha, 0\}$. Then $\phi_n \in C(\mathbb{R})$ and $\phi_n(t) \to 1_{t>\alpha}$ as $n \to \infty$ and the convergence is bounded when t is restricted to any compact subset of $\mathbb{R}$. Hence if $f \in \mathcal{H}$ it follows that $1_{f>\alpha} = \lim_{n\to\infty} \phi_n(f) \in \mathcal{H}$ for all $\alpha \in \mathbb{R}$,

i.e. $\{f > \alpha\} \in \mathcal{M}$ for all $\alpha \in \mathbb{R}$. Therefore if $f \in \mathcal{H}$ then $f \in \ell^\infty(\mathcal{M}, \mathbb{R})$ and we have shown $\mathcal{H} \subset \ell^\infty(\mathcal{M}, \mathbb{R})$. Conversely if $f \in \ell^\infty(\mathcal{M}, \mathbb{R})$, then for any $\alpha < \beta$, $\{\alpha < f \leq \beta\} \in \mathcal{M} = \mathcal{M}(\mathcal{H})$ and so by assumption $1_{\{\alpha < f \leq \beta\}} \in \mathcal{H}$. Combining this remark with the approximation Theorem 18.42 and the fact that $\mathcal{H}$ is closed under bounded convergence shows that $f \in \mathcal{H}$. Hence we have shown $\ell^\infty(\mathcal{M}, \mathbb{R}) \subset \mathcal{H}$ which combined with $\mathcal{H} \subset \ell^\infty(\mathcal{M}, \mathbb{R})$ already proved shows $\mathcal{H} = \ell^\infty(\mathcal{M}(\mathcal{H}), \mathbb{R})$.

3. Items 1. and 2. shows the map in Eq. (18.11) is surjective. To see the map is injective suppose $\mathcal{M}$ and $\mathcal{F}$ are two σ – algebras on X such that $\ell^\infty(\mathcal{M}, \mathbb{R}) = \ell^\infty(\mathcal{F}, \mathbb{R})$, then

$$\begin{aligned}\mathcal{M} &= \{A \subset X : 1_A \in \ell^\infty(\mathcal{M}, \mathbb{R})\} \\ &= \{A \subset X : 1_A \in \ell^\infty(\mathcal{F}, \mathbb{R})\} = \mathcal{F}.\end{aligned}$$

■

Notation 18.48 *Suppose M is a subset of $\ell^\infty(X, \mathbb{R})$.*

1. *Let $\mathcal{H}(M)$ denote the smallest subspace of $\ell^\infty(X, \mathbb{R})$ which contains M and the constant functions and is closed under bounded convergence.*
2. *Let $\mathcal{H}_\sigma(M)$ denote the smallest σ – function algebra containing M.*

Theorem 18.49. *Suppose M is a subset of $\ell^\infty(X, \mathbb{R})$, then $\mathcal{H}_\sigma(M) = \ell^\infty(\sigma(M), \mathbb{R})$ or in other words the following diagram commutes:*

$$\begin{array}{ccccc} & M & \longrightarrow & \sigma(M) & \\ M & \{\textit{Multiplicative Subsets}\} & \longrightarrow & \{\sigma - \textit{algebras}\} & \mathcal{M} \\ \downarrow & \downarrow & & \downarrow & \downarrow \\ \mathcal{H}_\sigma(M) & \{\sigma\text{–}\textit{ function algebras}\} & = & \{\sigma\text{–}\textit{ function algebras}\} & \ell^\infty(\mathcal{M}, \mathbb{R}). \end{array}$$

Proof. Since $\ell^\infty(\sigma(M), \mathbb{R})$ is σ – function algebra which contains M it follows that

$$\mathcal{H}_\sigma(M) \subset \ell^\infty(\sigma(M), \mathbb{R}).$$

For the opposite inclusion, let

$$\mathcal{M} = \mathcal{M}(\mathcal{H}_\sigma(M)) := \{A \subset X : 1_A \in \mathcal{H}_\sigma(M)\}.$$

By Theorem 18.47, $M \subset \mathcal{H}_\sigma(M) = \ell^\infty(\mathcal{M}, \mathbb{R})$ which implies that every $f \in M$ is $\mathcal{M}$ – measurable. This then implies $\sigma(M) \subset \mathcal{M}$ and therefore

$$\ell^\infty(\sigma(M), \mathbb{R}) \subset \ell^\infty(\mathcal{M}, \mathbb{R}) = \mathcal{H}_\sigma(M).$$

■

Definition 18.50 (Multiplicative System). *A collection of bounded real or complex valued functions, M, on a set X is called a **multiplicative system** if $f \cdot g \in M$ whenever f and g are in M.*

Theorem 18.51 (Dynkin's Multiplicative System Theorem). *Suppose $M \subset \ell^\infty(X, \mathbb{R})$ is a multiplicative system, then*

$$\mathcal{H}(M) = \mathcal{H}_\sigma(M) = \ell^\infty(\sigma(M), \mathbb{R}). \tag{18.12}$$

In words, the smallest subspace of bounded real valued functions on X which contains M that is closed under bounded convergence is the same as the space of bounded real valued $\sigma(M)$ – measurable functions on X.

Proof. We begin by proving $\mathcal{H} := \mathcal{H}(M)$ is a σ – function algebra. To do this, for any $f \in \mathcal{H}$ let

$$\mathcal{H}_f := \{g \in \mathcal{H} : fg \in \mathcal{H}\} \subset \mathcal{H}$$

and notice that $\mathcal{H}_f$ is a linear subspace of $\ell^\infty(X, \mathbb{R})$ which is closed under bounded convergence. Moreover if $f \in M$, $M \subset \mathcal{H}_f$ since M is multiplicative. Therefore $\mathcal{H}_f = \mathcal{H}$ and we have shown that $fg \in \mathcal{H}$ whenever $f \in M$ and $g \in \mathcal{H}$. Given this it now follows that $M \subset \mathcal{H}_f$ for any $f \in \mathcal{H}$ and by the same reasoning just used, $\mathcal{H}_f = \mathcal{H}$. Since $f \in \mathcal{H}$ is arbitrary, we have shown $fg \in \mathcal{H}$ for all $f, g \in \mathcal{H}$, i.e. $\mathcal{H}$ is an algebra. Since it is harder to be an algebra of functions containing M (see Exercise 18.13) than it is to be a subspace of functions containing M it follows that $\mathcal{H}(M) \subset \mathcal{H}_\sigma(M)$. But as we have just seen $\mathcal{H}(M)$ is a σ – function algebra which contains M so we must have $\mathcal{H}_\sigma(M) \subset \mathcal{H}(M)$ because $\mathcal{H}_\sigma(M)$ is by definition the smallest such σ – function algebra. Hence $\mathcal{H}_\sigma(M) = \mathcal{H}(M)$. The assertion that $\mathcal{H}_\sigma(M) = \ell^\infty(\sigma(M), \mathbb{R})$ has already been proved in Theorem 18.49. ∎

Theorem 18.52 (Complex Multiplicative System Theorem). *Suppose $\mathcal{H}$ is a complex linear subspace of $\ell^\infty(X, \mathbb{C})$ such that: $1 \in \mathcal{H}$, $\mathcal{H}$ is closed under complex conjugation, and $\mathcal{H}$ is closed under bounded convergence. If $M \subset \mathcal{H}$ is multiplicative system which is closed under conjugation, then $\mathcal{H}$ contains all bounded complex valued $\sigma(M)$-measurable functions, i.e. $\ell^\infty(\sigma(M), \mathbb{C}) \subset \mathcal{H}$.*

Proof. Let $M_0 = \mathrm{span}_{\mathbb{C}}(M \cup \{1\})$ be the complex span of M. As the reader should verify, M_0 is an algebra, $M_0 \subset \mathcal{H}$, M_0 is closed under complex conjugation and that $\sigma(M_0) = \sigma(M)$. Let $\mathcal{H}^{\mathbb{R}} := \mathcal{H} \cap \ell^\infty(X, \mathbb{R})$ and $M_0^{\mathbb{R}} = M \cap \ell^\infty(X, \mathbb{R})$. Then (you verify) $M_0^{\mathbb{R}}$ is a multiplicative system, $M_0^{\mathbb{R}} \subset \mathcal{H}^{\mathbb{R}}$ and $\mathcal{H}^{\mathbb{R}}$ is a linear space containing 1 which is closed under bounded convergence. Therefore by Theorem 18.51, $\ell^\infty\left(\sigma\left(M_0^{\mathbb{R}}\right), \mathbb{R}\right) \subset \mathcal{H}^{\mathbb{R}}$. Since $\mathcal{H}$ and M_0 are complex linear spaces closed under complex conjugation, for any $f \in \mathcal{H}$ or $f \in M_0$, the functions $\mathrm{Re}\, f = \frac{1}{2}\left(f + \bar{f}\right)$ and $\mathrm{Im}\, f = \frac{1}{2i}\left(f - \bar{f}\right)$ are in $\mathcal{H}$ (M_0) or M_0 respectively. Therefore $\mathcal{H} = \mathcal{H}^{\mathbb{R}} + i\mathcal{H}^{\mathbb{R}}$, $M_0 = M_0^{\mathbb{R}} + iM_0^{\mathbb{R}}$, $\sigma\left(M_0^{\mathbb{R}}\right) = \sigma(M_0) = \sigma(M)$ and

$$\begin{aligned} \ell^\infty(\sigma(M), \mathbb{C}) &= \ell^\infty\left(\sigma\left(M_0^{\mathbb{R}}\right), \mathbb{R}\right) + i\ell^\infty\left(\sigma\left(M_0^{\mathbb{R}}\right), \mathbb{R}\right) \\ &\subset \mathcal{H}^{\mathbb{R}} + i\mathcal{H}^{\mathbb{R}} = \mathcal{H}. \end{aligned}$$

∎

Exercise 18.7 (Algebra analogue of Theorem 18.47). Call a function algebra $\mathcal{H} \subset \ell^\infty(X, \mathbb{R})$ a **simple function algebra** if the range of each function $f \in \mathcal{H}$ is a finite subset of $\mathbb{R}$. Prove there is a one to one correspondence between algebras $\mathcal{A}$ on a set X and simple function algebras $\mathcal{H}$ on X.

Definition 18.53. *A collection of subsets, $\mathcal{C}$, of X is a* ***multiplicative class****(or a π –* ***class****) if $\mathcal{C}$ is closed under finite intersections.*

Corollary 18.54. *Suppose $\mathcal{H}$ is a subspace of $\ell^\infty(X, \mathbb{R})$ which is closed under bounded convergence and $1 \in \mathcal{H}$. If $\mathcal{C} \subset 2^X$ is a multiplicative class such that $1_A \in \mathcal{H}$ for all $A \in \mathcal{C}$, then $\mathcal{H}$ contains all bounded $\sigma(\mathcal{C})$ – measurable functions.*

Proof. Let $M = \{1\} \cup \{1_A : A \in \mathcal{C}\}$. Then $M \subset \mathcal{H}$ is a multiplicative system and the proof is completed with an application of Theorem 18.51. ■

Corollary 18.55. *Suppose that (X, d) is a metric space and $\mathcal{B}_X = \sigma(\tau_d)$ is the Borel σ – algebra on X and $\mathcal{H}$ is a subspace of $\ell^\infty(X, \mathbb{R})$ such that $BC(X, \mathbb{R}) \subset \mathcal{H}$ and $\mathcal{H}$ is closed under bounded convergence*[1]*. Then $\mathcal{H}$ contains all bounded $\mathcal{B}_X$ – measurable real valued functions on X. (This may be stated as follows: the smallest vector space of bounded functions which is closed under bounded convergence and contains $BC(X, \mathbb{R})$ is the space of bounded $\mathcal{B}_X$ – measurable real valued functions on X.)*

Proof. Let $V \in \tau_d$ be an open subset of X and for $n \in \mathbb{N}$ let

$$f_n(x) := \min(n \cdot d_{V^c}(x), 1) \text{ for all } x \in X.$$

Notice that $f_n = \phi_n \circ d_{V^c}$ where $\phi_n(t) = \min(nt, 1)$ (see Figure 18.3) which is continuous and hence $f_n \in BC(X, \mathbb{R})$ for all n. Furthermore, f_n converges boundedly to $1_{d_{V^c} > 0} = 1_V$ as $n \to \infty$ and therefore $1_V \in \mathcal{H}$ for all $V \in \tau$. Since τ is a π – class, the result now follows by an application of Corollary 18.54.

Plots of ϕ_1, ϕ_2 and ϕ_3.

[1] Recall that $BC(X, \mathbb{R})$ are the bounded continuous functions on X.

■

Here are some more variants of Corollary 18.55.

Proposition 18.56. *Let (X,d) be a metric space, $\mathcal{B}_X = \sigma(\tau_d)$ be the Borel σ – algebra and assume there exists compact sets $K_k \subset X$ such that $K_k^o \uparrow X$. Suppose that $\mathcal{H}$ is a subspace of $\ell^\infty(X,\mathbb{R})$ such that $C_c(X,\mathbb{R}) \subset \mathcal{H}$ ($C_c(X,\mathbb{R})$ is the space of continuous functions with compact support) and $\mathcal{H}$ is closed under bounded convergence. Then $\mathcal{H}$ contains all bounded $\mathcal{B}_X$ – measurable real valued functions on X.*

Proof. Let k and n be positive integers and set $\psi_{n,k}(x) = \min(1, n \cdot d_{(K_k^o)^c}(x))$. Then $\psi_{n,k} \in C_c(X,\mathbb{R})$ and $\{\psi_{n,k} \neq 0\} \subset K_k^o$. Let $\mathcal{H}_{n,k}$ denote those bounded $\mathcal{B}_X$ – measurable functions, $f : X \to \mathbb{R}$, such that $\psi_{n,k} f \in \mathcal{H}$. It is easily seen that $\mathcal{H}_{n,k}$ is closed under bounded convergence and that $\mathcal{H}_{n,k}$ contains $BC(X,\mathbb{R})$ and therefore by Corollary 18.55, $\psi_{n,k} f \in \mathcal{H}$ for all bounded measurable functions $f : X \to \mathbb{R}$. Since $\psi_{n,k} f \to 1_{K_k^o} f$ boundedly as $n \to \infty$, $1_{K_k^o} f \in \mathcal{H}$ for all k and similarly $1_{K_k^o} f \to f$ boundedly as $k \to \infty$ and therefore $f \in \mathcal{H}$. ■

Lemma 18.57. *Suppose that (X,τ) is a locally compact second countable Hausdorff space.*[2] *Then:*

1. *every open subset $U \subset X$ is σ – compact. In fact U is still a locally compact second countable Hausdorff space.*
2. *If $F \subset X$ is a closed set, there exist open sets $V_n \subset X$ such that $V_n \downarrow F$ as $n \to \infty$.*
3. *To each open set $U \subset X$ there exists $f_n \prec U$ (i.e. $f_n \in C_c(U,[0,1])$) such that $\lim_{n\to\infty} f_n = 1_U$.*
4. *$\mathcal{B}_X = \sigma(C_c(X,\mathbb{R}))$, i.e. the σ – algebra generated by $C_c(X)$ is the Borel σ – algebra on X.*

Proof.

1. Let U be an open subset of X, $\mathcal{V}$ be a countable base for τ and
$$\mathcal{V}^U := \{W \in \mathcal{V} : \bar{W} \subset U \text{ and } \bar{W} \text{ is compact}\}.$$
For each $x \in U$, by Proposition 12.7, there exists an open neighborhood V of x such that $\bar{V} \subset U$ and $\bar{V}$ is compact. Since $\mathcal{V}$ is a base for the topology τ, there exists $W \in \mathcal{V}$ such that $x \in W \subset V$. Because $\bar{W} \subset \bar{V}$, it follows that $\bar{W}$ is compact and hence $W \in \mathcal{V}^U$. As $x \in U$ was arbitrary, $U = \cup \mathcal{V}^U$. This shows $\mathcal{V}^U$ is a countable basis for the topology on U and that U is still locally compact.
Let $\{W_n\}_{n=1}^\infty$ be an enumeration of $\mathcal{V}^U$ and set $K_n := \cup_{k=1}^n \bar{W}_k$. Then $K_n \uparrow U$ as $n \to \infty$ and K_n is compact for each n. This shows U is σ – compact. (See Exercise 11.7.)

[2] For example any separable locally compact metric space and in particular any open subset of $\mathbb{R}^n$.

2. Let $\{K_n\}_{n=1}^\infty$ be compact subsets of F^c such that $K_n \uparrow F^c$ as $n \to \infty$ and set $V_n := K_n^c = X \setminus K_n$. Then $V_n \downarrow F$ and by Proposition 12.5, V_n is open for each n.
3. Let $U \subset X$ be an open set and $\{K_n\}_{n=1}^\infty$ be compact subsets of U such that $K_n \uparrow U$. By Urysohn's Lemma 12.8, there exist $f_n \prec U$ such that $f_n = 1$ on K_n. These functions satisfy, $1_U = \lim_{n\to\infty} f_n$.
4. By item 3., 1_U is $\sigma(C_c(X,\mathbb{R}))$ – measurable for all $U \in \tau$ and hence $\tau \subset \sigma(C_c(X,\mathbb{R}))$. Therefore $\mathcal{B}_X = \sigma(\tau) \subset \sigma(C_c(X,\mathbb{R}))$. The converse inclusion holds because continuous functions are always Borel measurable.

■

Here is a variant of Corollary 18.55.

Corollary 18.58. *Suppose that (X,τ) is a second countable locally compact Hausdorff space and $\mathcal{B}_X = \sigma(\tau)$ is the Borel σ – algebra on X. If $\mathcal{H}$ is a subspace of $\ell^\infty(X,\mathbb{R})$ which is closed under bounded convergence and contains $C_c(X,\mathbb{R})$, then $\mathcal{H}$ contains all bounded $\mathcal{B}_X$ – measurable real valued functions on X.*

Proof. By Item 3. of Lemma 18.57, for every $U \in \tau$ the characteristic function, 1_U, may be written as a bounded pointwise limit of functions from $C_c(X,\mathbb{R})$. Therefore $1_U \in \mathcal{H}$ for all $U \in \tau$. Since τ is a π – class, the proof is finished with an application of Corollary 18.54 ■

18.4 Product σ – Algebras

Let $\{(X_\alpha, \mathcal{M}_\alpha)\}_{\alpha\in A}$ be a collection of measurable spaces $X = X_A = \prod_{\alpha\in A} X_\alpha$ and $\pi_\alpha : X_A \to X_\alpha$ be the canonical projection map as in Notation 2.2.

Definition 18.59 (Product σ – Algebra). *The **product σ – algebra,** $\otimes_{\alpha\in A}\mathcal{M}_\alpha$, is the smallest σ – algebra on X such that each π_α for $\alpha \in A$ is measurable, i.e.*

$$\otimes_{\alpha\in A}\mathcal{M}_\alpha := \sigma(\pi_\alpha : \alpha \in A) = \sigma\left(\cup_\alpha \pi_\alpha^{-1}(\mathcal{M}_\alpha)\right).$$

Applying Proposition 18.25 in this setting implies the following proposition.

Proposition 18.60. *Suppose Y is a measurable space and $f : Y \to X = X_A$ is a map. Then f is measurable iff $\pi_\alpha \circ f : Y \to X_\alpha$ is measurable for all $\alpha \in A$. In particular if $A = \{1, 2, \ldots, n\}$ so that $X = X_1 \times X_2 \times \cdots \times X_n$ and $f(y) = (f_1(y), f_2(y), \ldots, f_n(y)) \in X_1 \times X_2 \times \cdots \times X_n$, then $f : Y \to X_A$ is measurable iff $f_i : Y \to X_i$ is measurable for all i.*

Proposition 18.61. *Suppose that* $(X_\alpha, \mathcal{M}_\alpha)_{\alpha\in A}$ *is a collection of measurable spaces and* $\mathcal{E}_\alpha \subset \mathcal{M}_\alpha$ *generates* $\mathcal{M}_\alpha$ *for each* $\alpha \in A$*, then*

$$\otimes_{\alpha\in A}\mathcal{M}_\alpha = \sigma\left(\cup_{\alpha\in A}\pi_\alpha^{-1}(\mathcal{E}_\alpha)\right) \tag{18.13}$$

Moreover, suppose that A *is either finite or countably infinite,* $X_\alpha \in \mathcal{E}_\alpha$ *for each* $\alpha \in A$*, and* $\mathcal{M}_\alpha = \sigma(\mathcal{E}_\alpha)$ *for each* $\alpha \in A$*. Then the product* σ *– algebra satisfies*

$$\otimes_{\alpha\in A}\mathcal{M}_\alpha = \sigma\left(\left\{\prod_{\alpha\in A} E_\alpha : E_\alpha \in \mathcal{E}_\alpha \text{ for all } \alpha \in A\right\}\right). \tag{18.14}$$

In particular if $A = \{1, 2, \ldots, n\}$ *, then* $X = X_1 \times X_2 \times \cdots \times X_n$ *and*

$$\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \cdots \otimes \mathcal{M}_n = \sigma(\mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_n),$$

where $\mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_n$ *is as defined in Notation 10.26.*

Proof. Since $\cup_\alpha \pi_\alpha^{-1}(\mathcal{E}_\alpha) \subset \cup_\alpha \pi_\alpha^{-1}(\mathcal{M}_\alpha)$, it follows that

$$\mathcal{F} := \sigma\left(\cup_\alpha \pi_\alpha^{-1}(\mathcal{E}_\alpha)\right) \subset \sigma\left(\cup_\alpha \pi_\alpha^{-1}(\mathcal{M}_\alpha)\right) = \otimes_{\alpha\in A}\mathcal{M}_\alpha.$$

Conversely,

$$\mathcal{F} \supset \sigma(\pi_\alpha^{-1}(\mathcal{E}_\alpha)) = \pi_\alpha^{-1}(\sigma(\mathcal{E}_\alpha)) = \pi_\alpha^{-1}(\mathcal{M}_\alpha)$$

holds for all α implies that

$$\cup_\alpha \pi_\alpha^{-1}(\mathcal{M}_\alpha) \subset \mathcal{F}$$

and hence that $\otimes_{\alpha\in A}\mathcal{M}_\alpha \subset \mathcal{F}$. We now prove Eq. (18.14). Since we are assuming that $X_\alpha \in \mathcal{E}_\alpha$ for each $\alpha \in A$, we see that

$$\cup_\alpha \pi_\alpha^{-1}(\mathcal{E}_\alpha) \subset \left\{\prod_{\alpha\in A} E_\alpha : E_\alpha \in \mathcal{E}_\alpha \text{ for all } \alpha \in A\right\}$$

and therefore by Eq. (18.13)

$$\otimes_{\alpha\in A}\mathcal{M}_\alpha = \sigma\left(\cup_\alpha \pi_\alpha^{-1}(\mathcal{E}_\alpha)\right) \subset \sigma\left(\left\{\prod_{\alpha\in A} E_\alpha : E_\alpha \in \mathcal{E}_\alpha \text{ for all } \alpha \in A\right\}\right).$$

This last statement is true independent as to whether A is countable or not. For the reverse inclusion it suffices to notice that since A is countable,

$$\prod_{\alpha\in A} E_\alpha = \cap_{\alpha\in A}\pi_\alpha^{-1}(E_\alpha) \in \otimes_{\alpha\in A}\mathcal{M}_\alpha$$

and hence

$$\sigma\left(\left\{\prod_{\alpha\in A} E_\alpha : E_\alpha \in \mathcal{E}_\alpha \text{ for all } \alpha \in A\right\}\right) \subset \otimes_{\alpha\in A}\mathcal{M}_\alpha.$$

■

Remark 18.62. One can not relax the assumption that $X_\alpha \in \mathcal{E}_\alpha$ in the second part of Proposition 18.61. For example, if $X_1 = X_2 = \{1,2\}$ and $\mathcal{E}_1 = \mathcal{E}_2 = \{\{1\}\}$, then $\sigma(\mathcal{E}_1 \times \mathcal{E}_2) = \{\emptyset, X_1 \times X_2, \{(1,1)\}\}$ while $\sigma(\sigma(\mathcal{E}_1) \times \sigma(\mathcal{E}_2)) = 2^{X_1 \times X_2}$.

Theorem 18.63. *Let $\{X_\alpha\}_{\alpha \in A}$ be a sequence of sets where A is at most countable. Suppose for each $\alpha \in A$ we are given a countable set $\mathcal{E}_\alpha \subset 2^{X_\alpha}$. Let $\tau_\alpha = \tau(\mathcal{E}_\alpha)$ be the topology on X_α generated by $\mathcal{E}_\alpha$ and X be the product space $\prod_{\alpha \in A} X_\alpha$ with equipped with the product topology $\tau := \otimes_{\alpha \in A} \tau(\mathcal{E}_\alpha)$. Then the Borel σ – algebra $\mathcal{B}_X = \sigma(\tau)$ is the same as the product σ – algebra:*

$$\mathcal{B}_X = \otimes_{\alpha \in A} \mathcal{B}_{X_\alpha},$$

where $\mathcal{B}_{X_\alpha} = \sigma(\tau(\mathcal{E}_\alpha)) = \sigma(\mathcal{E}_\alpha)$ for all $\alpha \in A$.

In particular if $A = \{1, 2, \ldots, n\}$ and each (X_i, τ_i) is a second countable topological space, then

$$\mathcal{B}_X := \sigma(\tau_1 \otimes \tau_2 \otimes \cdots \otimes \tau_n) = \sigma(\mathcal{B}_{X_1} \times \cdots \times \mathcal{B}_{X_n}) =: \mathcal{B}_{X_1} \otimes \cdots \otimes \mathcal{B}_{X_n}.$$

Proof. By Proposition 10.25, the topology τ may be described as the smallest topology containing $\mathcal{E} = \cup_{\alpha \in A} \pi_\alpha^{-1}(\mathcal{E}_\alpha)$. Now $\mathcal{E}$ is the countable union of countable sets so is still countable. Therefore by Proposition 18.17 and Proposition 18.61,

$$\begin{aligned}\mathcal{B}_X &= \sigma(\tau) = \sigma(\tau(\mathcal{E})) = \sigma(\mathcal{E}) = \otimes_{\alpha \in A} \sigma(\mathcal{E}_\alpha) \\ &= \otimes_{\alpha \in A} \sigma(\tau_\alpha) = \otimes_{\alpha \in A} \mathcal{B}_{X_\alpha}.\end{aligned}$$

■

Corollary 18.64. *If (X_i, d_i) are separable metric spaces for $i = 1, \ldots, n$, then*

$$\mathcal{B}_{X_1} \otimes \cdots \otimes \mathcal{B}_{X_n} = \mathcal{B}_{(X_1 \times \cdots \times X_n)}$$

where $\mathcal{B}_{X_i}$ is the Borel σ – algebra on X_i and $\mathcal{B}_{(X_1 \times \cdots \times X_n)}$ is the Borel σ – algebra on $X_1 \times \cdots \times X_n$ equipped with the metric topology associated to the metric $d(x,y) = \sum_{i=1}^n d_i(x_i, y_i)$ where $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$.

Proof. This is a combination of the results in Lemma 10.28, Exercise 10.12 and Theorem 18.63. ■

Because all norms on finite dimensional spaces are equivalent, the usual Euclidean norm on $\mathbb{R}^m \times \mathbb{R}^n$ is equivalent to the "product" norm defined by

$$\|(x,y)\|_{\mathbb{R}^m \times \mathbb{R}^n} = \|x\|_{\mathbb{R}^m} + \|y\|_{\mathbb{R}^n}.$$

Hence by Lemma 10.28, the Euclidean topology on $\mathbb{R}^{m+n}$ is the same as the product topology on $\mathbb{R}^{m+n} \cong \mathbb{R}^m \times \mathbb{R}^n$. Here we are identifying $\mathbb{R}^m \times \mathbb{R}^n$ with $\mathbb{R}^{m+n}$ by the map

$$(x,y) \in \mathbb{R}^m \times \mathbb{R}^n \to (x_1, \ldots, x_m, y_1, \ldots, y_n) \in \mathbb{R}^{m+n}.$$

These comments along with Corollary 18.64 proves the following result.

Corollary 18.65. *After identifying $\mathbb{R}^m \times \mathbb{R}^n$ with $\mathbb{R}^{m+n}$ as above and letting $\mathcal{B}_{\mathbb{R}^n}$ denote the Borel σ –algebra on $\mathbb{R}^n$, we have*

$$\mathcal{B}_{\mathbb{R}^{m+n}} = \mathcal{B}_{\mathbb{R}^n} \otimes \mathcal{B}_{\mathbb{R}^m} \text{ and } \mathcal{B}_{\mathbb{R}^n} = \overbrace{\mathcal{B}_{\mathbb{R}} \otimes \cdots \otimes \mathcal{B}_{\mathbb{R}}}^{n\text{–times}}.$$

18.4.1 Factoring of Measurable Maps

Lemma 18.66. *Suppose that $(Y, \mathcal{F})$ is a measurable space and $F : X \to Y$ is a map. Then to every $(\sigma(F), \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, $H : X \to \bar{\mathbb{R}}$, there is a $(\mathcal{F}, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function $h : Y \to \bar{\mathbb{R}}$ such that $H = h \circ F$.*

Proof. First suppose that $H = 1_A$ where $A \in \sigma(F) = F^{-1}(\mathcal{F})$. Let $B \in \mathcal{F}$ such that $A = F^{-1}(B)$ then $1_A = 1_{F^{-1}(B)} = 1_B \circ F$ and hence the Lemma is valid in this case with $h = 1_B$. More generally if $H = \sum a_i 1_{A_i}$ is a simple function, then there exists $B_i \in \mathcal{F}$ such that $1_{A_i} = 1_{B_i} \circ F$ and hence $H = h \circ F$ with $h := \sum a_i 1_{B_i}$ – a simple function on $\bar{\mathbb{R}}$. For general $(\sigma(F), \mathcal{F})$ – measurable function, H, from $X \to \bar{\mathbb{R}}$, choose simple functions H_n converging to H. Let h_n be simple functions on $\bar{\mathbb{R}}$ such that $H_n = h_n \circ F$. Then it follows that

$$H = \lim_{n\to\infty} H_n = \limsup_{n\to\infty} H_n = \limsup_{n\to\infty} h_n \circ F = h \circ F$$

where $h := \limsup_{n\to\infty} h_n$ – a measurable function from Y to $\bar{\mathbb{R}}$. ∎

The following is an immediate corollary of Proposition 18.25 and Lemma 18.66.

Corollary 18.67. *Let X and A be sets, and suppose for $\alpha \in A$ we are give a measurable space $(Y_\alpha, \mathcal{F}_\alpha)$ and a function $f_\alpha : X \to Y_\alpha$. Let $Y := \prod_{\alpha\in A} Y_\alpha$, $\mathcal{F} := \otimes_{\alpha\in A}\mathcal{F}_\alpha$ be the product σ – algebra on Y and $\mathcal{M} := \sigma(f_\alpha : \alpha \in A)$ be the smallest σ – algebra on X such that each f_α is measurable. Then the function $F : X \to Y$ defined by $[F(x)]_\alpha := f_\alpha(x)$ for each $\alpha \in A$ is $(\mathcal{M}, \mathcal{F})$ – measurable and a function $H : X \to \bar{\mathbb{R}}$ is $(\mathcal{M}, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable iff there exists a $(\mathcal{F}, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function h from Y to $\bar{\mathbb{R}}$ such that $H = h \circ F$.*

18.5 Exercises

Exercise 18.8. Prove Corollary 18.23. **Hint:** See Exercise 18.3.

Exercise 18.9. If $\mathcal{M}$ is the σ – algebra generated by $\mathcal{E} \subset 2^X$, then $\mathcal{M}$ is the union of the σ – algebras generated by countable subsets $\mathcal{F} \subset \mathcal{E}$.

Exercise 18.10. Let $(X, \mathcal{M})$ be a measure space and $f_n : X \to \mathbb{F}$ be a sequence of measurable functions on X. Show that $\{x : \lim_{n\to\infty} f_n(x) \text{ exists } \textbf{in } \mathbb{F}\} \in \mathcal{M}$.

Exercise 18.11. Show that every monotone function $f : \mathbb{R} \to \mathbb{R}$ is $(\mathcal{B}_{\mathbb{R}}, \mathcal{B}_{\mathbb{R}})$ – measurable.

Exercise 18.12. Show by example that the supremum of an uncountable family of measurable functions need not be measurable. (Folland problem 2.6 on p. 48.)

Exercise 18.13. Let $X = \{1, 2, 3, 4\}$, $A = \{1, 2\}$, $B = \{2, 3\}$ and $M := \{1_A, 1_B\}$. Show $\mathcal{H}_\sigma(M) \neq \mathcal{H}(M)$ in this case.

19

Measures and Integration

Definition 19.1. *A* ***measure*** μ *on a measurable space* $(X, \mathcal{M})$ *is a function* $\mu : \mathcal{M} \to [0, \infty]$ *such that*

1. $\mu(\emptyset) = 0$ *and*
2. *(Finite Additivity) If* $\{A_i\}_{i=1}^n \subset \mathcal{M}$ *are pairwise disjoint, i.e.* $A_i \cap A_j = \emptyset$ *when* $i \neq j$*, then*
$$\mu(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i).$$
3. *(Continuity) If* $A_n \in \mathcal{M}$ *and* $A_n \uparrow A$*, then* $\mu(A_n) \uparrow \mu(A)$*.*

We call a triple $(X, \mathcal{M}, \mu)$*, where* $(X, \mathcal{M})$ *is a measurable space and* $\mu : \mathcal{M} \to [0, \infty]$ *is a measure, a* ***measure space.***

Remark 19.2. Properties 2) and 3) in Definition 19.1 are equivalent to the following condition. If $\{A_i\}_{i=1}^\infty \subset \mathcal{M}$ are pairwise disjoint then

$$\mu(\bigcup_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mu(A_i). \tag{19.1}$$

To prove this assume that Properties 2) and 3) in Definition 19.1 hold and $\{A_i\}_{i=1}^\infty \subset \mathcal{M}$ are pairwise disjoint. Letting $B_n := \bigcup\limits_{i=1}^n A_i \uparrow B := \bigcup\limits_{i=1}^\infty A_i$, we have

$$\mu(\bigcup_{i=1}^\infty A_i) = \mu(B) \overset{(3)}{=} \lim_{n\to\infty} \mu(B_n) \overset{(2)}{=} \lim_{n\to\infty} \sum_{i=1}^n \mu(A_i) = \sum_{i=1}^\infty \mu(A_i).$$

Conversely, if Eq. (19.1) holds we may take $A_j = \emptyset$ for all $j > n$ to see that Property 2) of Definition 19.1 holds. Also if $A_n \uparrow A$, let $B_n := A_n \setminus A_{n-1}$ with $A_0 := \emptyset$. Then $\{B_n\}_{n=1}^\infty$ are pairwise disjoint, $A_n = \cup_{j=1}^n B_j$ and $A = \cup_{j=1}^\infty B_j$. So if Eq. (19.1) holds we have

$$\mu(A) = \mu\left(\cup_{j=1}^{\infty} B_j\right) = \sum_{j=1}^{\infty} \mu(B_j)$$

$$= \lim_{n\to\infty} \sum_{j=1}^{n} \mu(B_j) = \lim_{n\to\infty} \mu(\cup_{j=1}^{n} B_j) = \lim_{n\to\infty} \mu(A_n).$$

Proposition 19.3 (Basic properties of measures). *Suppose that* $(X, \mathcal{M}, \mu)$ *is a measure space and* $E, F \in \mathcal{M}$ *and* $\{E_j\}_{j=1}^{\infty} \subset \mathcal{M}$, *then :*

1. $\mu(E) \le \mu(F)$ *if* $E \subset F$.
2. $\mu(\cup E_j) \le \sum \mu(E_j)$.
3. *If* $\mu(E_1) < \infty$ *and* $E_j \downarrow E$, *i.e.* $E_1 \supset E_2 \supset E_3 \supset \dots$ *and* $E = \cap_j E_j$, *then* $\mu(E_j) \downarrow \mu(E)$ *as* $j \to \infty$.

Proof.

1. Since $F = E \cup (F \setminus E)$,
$$\mu(F) = \mu(E) + \mu(F \setminus E) \ge \mu(E).$$
2. Let $\widetilde{E}_j = E_j \setminus (E_1 \cup \dots \cup E_{j-1})$ so that the $\tilde{E}_j$'s are pair-wise disjoint and $E = \cup \widetilde{E}_j$. Since $\tilde{E}_j \subset E_j$ it follows from Remark 19.2 and part (1), that
$$\mu(E) = \sum \mu(\widetilde{E}_j) \le \sum \mu(E_j).$$
3. Define $D_i := E_1 \setminus E_i$ then $D_i \uparrow E_1 \setminus E$ which implies that
$$\mu(E_1) - \mu(E) = \lim_{i\to\infty} \mu(D_i) = \mu(E_1) - \lim_{i\to\infty} \mu(E_i)$$
which shows that $\lim_{i\to\infty} \mu(E_i) = \mu(E)$.

■

Definition 19.4. *A set* $E \subset X$ *is a* ***null*** *set if* $E \in \mathcal{M}$ *and* $\mu(E) = 0$. *If* P *is some "property" which is either true or false for each* $x \in X$, *we will use the terminology* P *a.e. (to be read* P *almost everywhere) to mean*

$$E := \{x \in X : P \text{ is false for } x\}$$

is a null set. For example if f *and* g *are two measurable functions on* $(X, \mathcal{M}, \mu)$, $f = g$ *a.e. means that* $\mu(f \ne g) = 0$.

Definition 19.5. *A measure space* $(X, \mathcal{M}, \mu)$ *is* ***complete*** *if every subset of a null set is in* $\mathcal{M}$, *i.e. for all* $F \subset X$ *such that* $F \subset E \in \mathcal{M}$ *with* $\mu(E) = 0$ *implies that* $F \in \mathcal{M}$.

Proposition 19.6 (Completion of a Measure). *Let $(X, \mathcal{M}, \mu)$ be a measure space. Set*

$$\begin{aligned}
\mathcal{N} = \mathcal{N}^\mu &:= \{N \subset X : \exists\, F \in \mathcal{M} \ni N \subset F \text{ and } \mu(F) = 0\}, \\
\bar{\mathcal{M}} = \bar{\mathcal{M}}^\mu &:= \{A \cup N : A \in \mathcal{M} \text{ and } N \in \mathcal{N}\} \text{ and} \\
\bar{\mu}(A \cup N) &:= \mu(A) \text{ for } A \in \mathcal{M} \text{ and } N \in \mathcal{N},
\end{aligned}$$

*see Fig. 19.1. Then $\bar{\mathcal{M}}$ is a σ – algebra, $\bar{\mu}$ is a well defined measure on $\bar{\mathcal{M}}$, $\bar{\mu}$ is the unique measure on $\bar{\mathcal{M}}$ which extends μ on $\mathcal{M}$, and $(X, \bar{\mathcal{M}}, \bar{\mu})$ is complete measure space. The σ-algebra, $\bar{\mathcal{M}}$, is called the **completion** of $\mathcal{M}$ relative to μ and $\bar{\mu}$, is called the **completion of** μ.*

Proof. Clearly $X, \emptyset \in \bar{\mathcal{M}}$. Let $A \in \mathcal{M}$ and $N \in \mathcal{N}$ and choose $F \in \mathcal{M}$

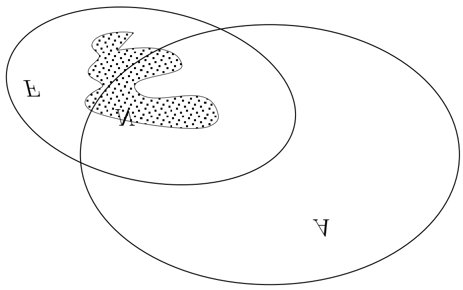


Fig. 19.1. Completing a σ – algebra.

such that $N \subset F$ and $\mu(F) = 0$. Since $N^c = (F \setminus N) \cup F^c$,

$$\begin{aligned}
(A \cup N)^c &= A^c \cap N^c = A^c \cap (F \setminus N \cup F^c) \\
&= [A^c \cap (F \setminus N)] \cup [A^c \cap F^c]
\end{aligned}$$

where $[A^c \cap (F \setminus N)] \in \mathcal{N}$ and $[A^c \cap F^c] \in \mathcal{M}$. Thus $\bar{\mathcal{M}}$ is closed under complements. If $A_i \in \mathcal{M}$ and $N_i \subset F_i \in \mathcal{M}$ such that $\mu(F_i) = 0$ then $\cup(A_i \cup N_i) = (\cup A_i) \cup (\cup N_i) \in \bar{\mathcal{M}}$ since $\cup A_i \in \mathcal{M}$ and $\cup N_i \subset \cup F_i$ and $\mu(\cup F_i) \leq \sum \mu(F_i) = 0$. Therefore, $\bar{\mathcal{M}}$ is a σ – algebra. Suppose $A \cup N_1 = B \cup N_2$ with $A, B \in \mathcal{M}$ and $N_1, N_2, \in \mathcal{N}$. Then $A \subset A \cup N_1 \subset A \cup N_1 \cup F_2 = B \cup F_2$ which shows that

$$\mu(A) \leq \mu(B) + \mu(F_2) = \mu(B).$$

Similarly, we show that $\mu(B) \leq \mu(A)$ so that $\mu(A) = \mu(B)$ and hence $\bar{\mu}(A \cup N) := \mu(A)$ is well defined. It is left as an exercise to show $\bar{\mu}$ is a measure, i.e. that it is countable additive. ■

Many theorems in the sequel will require some control on the size of a measure μ. The relevant notion for our purposes (and most purposes) is that of a σ – finite measure defined next.

Definition 19.7. *Suppose X is a set, $\mathcal{E} \subset \mathcal{M} \subset 2^X$ and $\mu : \mathcal{M} \to [0, \infty]$ is a function. The function μ is σ – finite on $\mathcal{E}$ if there exists $E_n \in \mathcal{E}$ such that $\mu(E_n) < \infty$ and $X = \cup_{n=1}^{\infty} E_n$. If $\mathcal{M}$ is a σ – algebra and μ is a measure on $\mathcal{M}$ which is σ – finite on $\mathcal{M}$ we will say $(X, \mathcal{M}, \mu)$ is a σ – finite measure space.*

The reader should check that if μ is a finitely additive measure on an algebra, $\mathcal{M}$, then μ is σ – finite on $\mathcal{M}$ iff there exists $X_n \in \mathcal{M}$ such that $X_n \uparrow X$ and $\mu(X_n) < \infty$.

19.1 Example of Measures

Most σ – algebras and σ -additive measures are somewhat difficult to describe and define. However, one special case is fairly easy to understand. Namely suppose that $\mathcal{F} \subset 2^X$ is a countable or finite partition of X and $\mathcal{M} \subset 2^X$ is the σ – algebra which consists of the collection of sets $A \subset X$ such that

$$A = \cup \{\alpha \in \mathcal{F} : \alpha \subset A\}. \tag{19.2}$$

It is easily seen that $\mathcal{M}$ is a σ – algebra.

Any measure $\mu : \mathcal{M} \to [0, \infty]$ is determined uniquely by its values on $\mathcal{F}$. Conversely, if we are given any function $\lambda : \mathcal{F} \to [0, \infty]$ we may define, for $A \in \mathcal{M}$,

$$\mu(A) = \sum_{\alpha \in \mathcal{F} \ni \alpha \subset A} \lambda(\alpha) = \sum_{\alpha \in \mathcal{F}} \lambda(\alpha) 1_{\alpha \subset A}$$

where $1_{\alpha \subset A}$ is one if $\alpha \subset A$ and zero otherwise. We may check that μ is a measure on $\mathcal{M}$. Indeed, if $A = \coprod_{i=1}^{\infty} A_i$ and $\alpha \in \mathcal{F}$, then $\alpha \subset A$ iff $\alpha \subset A_i$ for one and hence exactly one A_i. Therefore $1_{\alpha \subset A} = \sum_{i=1}^{\infty} 1_{\alpha \subset A_i}$ and hence

$$\begin{aligned}
\mu(A) &= \sum_{\alpha \in \mathcal{F}} \lambda(\alpha) 1_{\alpha \subset A} = \sum_{\alpha \in \mathcal{F}} \lambda(\alpha) \sum_{i=1}^{\infty} 1_{\alpha \subset A_i} \\
&= \sum_{i=1}^{\infty} \sum_{\alpha \in \mathcal{F}} \lambda(\alpha) 1_{\alpha \subset A_i} = \sum_{i=1}^{\infty} \mu(A_i)
\end{aligned}$$

as desired. Thus we have shown that there is a one to one correspondence between measures μ on $\mathcal{M}$ and functions $\lambda : \mathcal{F} \to [0, \infty]$.

The construction of measures will be covered in Chapters **??** – **??** below. However, let us record here the existence of an interesting class of measures.

Theorem 19.8. *To every right continuous non-decreasing function $F : \mathbb{R} \to \mathbb{R}$ there exists a unique measure μ_F on $\mathcal{B}_{\mathbb{R}}$ such that*

$$\mu_F((a, b]) = F(b) - F(a) \ \forall \ -\infty < a \le b < \infty \tag{19.3}$$

Moreover, if $A \in \mathcal{B}_{\mathbb{R}}$ then

$$\mu_F(A) = \inf\left\{\sum_{i=1}^{\infty}(F(b_i) - F(a_i)) : A \subset \cup_{i=1}^{\infty}(a_i, b_i]\right\} \tag{19.4}$$

$$= \inf\left\{\sum_{i=1}^{\infty}(F(b_i) - F(a_i)) : A \subset \coprod_{i=1}^{\infty}(a_i, b_i]\right\}. \tag{19.5}$$

In fact the map $F \to \mu_F$ is a one to one correspondence between right continuous functions F with $F(0) = 0$ on one hand and measures μ on $\mathcal{B}_{\mathbb{R}}$ such that $\mu(J) < \infty$ on any bounded set $J \in \mathcal{B}_{\mathbb{R}}$ on the other.

Proof. See Section **??** below or Theorem **??** below. ∎

Example 19.9. The most important special case of Theorem 19.8 is when $F(x) = x$, in which case we write m for μ_F. The measure m is called Lebesgue measure.

Theorem 19.10. *Lebesgue measure m is invariant under translations, i.e. for $B \in \mathcal{B}_{\mathbb{R}}$ and $x \in \mathbb{R}$,*

$$m(x + B) = m(B). \tag{19.6}$$

Moreover, m is the unique measure on $\mathcal{B}_{\mathbb{R}}$ such that $m((0, 1]) = 1$ and Eq. (19.6) holds for $B \in \mathcal{B}_{\mathbb{R}}$ and $x \in \mathbb{R}$. Moreover, m has the scaling property

$$m(\lambda B) = |\lambda|\, m(B) \tag{19.7}$$

where $\lambda \in \mathbb{R}$, $B \in \mathcal{B}_{\mathbb{R}}$ and $\lambda B := \{\lambda x : x \in B\}$.

Proof. Let $m_x(B) := m(x + B)$, then one easily shows that m_x is a measure on $\mathcal{B}_{\mathbb{R}}$ such that $m_x((a, b]) = b - a$ for all $a < b$. Therefore, $m_x = m$ by the uniqueness assertion in Theorem 19.8. For the converse, suppose that m is translation invariant and $m((0, 1]) = 1$. Given $n \in \mathbb{N}$, we have

$$(0, 1] = \cup_{k=1}^{n}(\frac{k-1}{n}, \frac{k}{n}] = \cup_{k=1}^{n}\left(\frac{k-1}{n} + (0, \frac{1}{n}]\right).$$

Therefore,

$$1 = m((0, 1]) = \sum_{k=1}^{n} m\left(\frac{k-1}{n} + (0, \frac{1}{n}]\right)$$

$$= \sum_{k=1}^{n} m((0, \frac{1}{n}]) = n \cdot m((0, \frac{1}{n}]).$$

That is to say

$$m((0, \frac{1}{n}]) = 1/n.$$

Similarly, $m((0,\frac{l}{n}]) = l/n$ for all $l,n \in \mathbb{N}$ and therefore by the translation invariance of m,

$$m((a,b]) = b-a \text{ for all } a,b \in \mathbb{Q} \text{ with } a<b.$$

Finally for $a,b \in \mathbb{R}$ such that $a<b$, choose $a_n, b_n \in \mathbb{Q}$ such that $b_n \downarrow b$ and $a_n \uparrow a$, then $(a_n, b_n] \downarrow (a,b]$ and thus

$$m((a,b]) = \lim_{n\to\infty} m((a_n,b_n]) = \lim_{n\to\infty} (b_n - a_n) = b-a,$$

i.e. m is Lebesgue measure. To prove Eq. (19.7) we may assume that $\lambda \neq 0$ since this case is trivial to prove. Now let $m_\lambda(B) := |\lambda|^{-1} m(\lambda B)$. It is easily checked that m_λ is again a measure on $\mathcal{B}_{\mathbb{R}}$ which satisfies

$$m_\lambda((a,b]) = \lambda^{-1} m\left((\lambda a, \lambda b]\right) = \lambda^{-1}(\lambda b - \lambda a) = b-a$$

if $\lambda > 0$ and

$$m_\lambda((a,b]) = |\lambda|^{-1} m\left([\lambda b, \lambda a)\right) = -\left|\lambda\right|^{-1}(\lambda b - \lambda a) = b-a$$

if $\lambda < 0$. Hence $m_\lambda = m$. ■

We are now going to develop integration theory relative to a measure. The integral defined in the case for Lebesgue measure, m, will be an extension of the standard Riemann integral on $\mathbb{R}$.

19.1.1 ADD: Examples of Measures

BRUCE: ADD details.

1. Product measure for the flipping of a coin.
2. Haar Measure
3. Measure on embedded submanifolds, i.e. Hausdorff measure.
4. Wiener measure.
5. Gibbs states.
6. Measure associated to self-adjoint operators and classifying them.

19.2 Integrals of Simple functions

Let $(X, \mathcal{M}, \mu)$ be a fixed measure space in this section.

Definition 19.11. *Let $\mathbb{F} = \mathbb{C}$ or $[0,\infty)$ and suppose that $\phi : X \to \mathbb{F}$ is a simple function as in Definition 18.41. If $\mathbb{F} = \mathbb{C}$ assume further that $\mu(\phi^{-1}(\{y\})) < \infty$ for all $y \neq 0$ in $\mathbb{C}$. For such functions ϕ, define $I_\mu(\phi)$ by*

$$I_\mu(\phi) = \sum_{y\in\mathbb{F}} y\mu(\phi^{-1}(\{y\})).$$

Proposition 19.12. *Let $\lambda \in \mathbb{F}$ and ϕ and ψ be two simple functions, then I_μ satisfies:*

1.
$$I_\mu(\lambda\phi) = \lambda I_\mu(\phi). \tag{19.8}$$

2.
$$I_\mu(\phi + \psi) = I_\mu(\psi) + I_\mu(\phi).$$

3. *If ϕ and ψ are non-negative simple functions such that $\phi \le \psi$ then*
$$I_\mu(\phi) \le I_\mu(\psi).$$

Proof. Let us write $\{\phi = y\}$ for the set $\phi^{-1}(\{y\}) \subset X$ and $\mu(\phi = y)$ for $\mu(\{\phi = y\}) = \mu(\phi^{-1}(\{y\}))$ so that

$$I_\mu(\phi) = \sum_{y\in\mathbb{F}} y\mu(\phi = y).$$

We will also write $\{\phi = a, \psi = b\}$ for $\phi^{-1}(\{a\}) \cap \psi^{-1}(\{b\})$. This notation is more intuitive for the purposes of this proof. Suppose that $\lambda \in \mathbb{F}$ then

$$\begin{aligned} I_\mu(\lambda\phi) &= \sum_{y\in\mathbb{F}} y\ \mu(\lambda\phi = y) = \sum_{y\in\mathbb{F}} y\ \mu(\phi = y/\lambda) \\ &= \sum_{z\in\mathbb{F}} \lambda z\ \mu(\phi = z) = \lambda I_\mu(\phi) \end{aligned}$$

provided that $\lambda \neq 0$. The case $\lambda = 0$ is clear, so we have proved 1. Suppose that ϕ and ψ are two simple functions, then

$$\begin{aligned} I_\mu(\phi + \psi) &= \sum_{z\in\mathbb{F}} z\ \mu(\phi + \psi = z) \\ &= \sum_{z\in\mathbb{F}} z\ \mu\left(\cup_{w\in\mathbb{F}} \{\phi = w,\ \psi = z - w\}\right) \\ &= \sum_{z\in\mathbb{F}} z \sum_{w\in\mathbb{F}} \mu(\phi = w,\ \psi = z - w) \\ &= \sum_{z,w\in\mathbb{F}} (z + w)\mu(\phi = w,\ \psi = z) \\ &= \sum_{z\in\mathbb{F}} z\ \mu(\psi = z) + \sum_{w\in\mathbb{F}} w\ \mu(\phi = w) \\ &= I_\mu(\psi) + I_\mu(\phi). \end{aligned}$$

which proves 2. For 3. if ϕ and ψ are non-negative simple functions such that $\phi \le \psi$

$$\begin{aligned} I_\mu(\phi) &= \sum_{a\geq 0} a\mu(\phi = a) = \sum_{a,b\geq 0} a\mu(\phi = a, \psi = b) \\ &\leq \sum_{a,b\geq 0} b\mu(\phi = a, \psi = b) = \sum_{b\geq 0} b\mu(\psi = b) = I_\mu(\psi), \end{aligned}$$

wherein the third inequality we have used $\{\phi = a, \psi = b\} = \emptyset$ if $a > b$. ■

19.3 Integrals of positive functions

Definition 19.13. *Let $L^+ = L^+(\mathcal{M}) = \{f : X \to [0,\infty] : f$ is measurable$\}$. Define*

$$\int_X f(x)\, d\mu(x) = \int_X f d\mu := \sup\{I_\mu(\phi) : \phi \text{ is simple and } \phi \leq f\}.$$

*We say the $f \in L^+$ is **integrable** if $\int_X f d\mu < \infty$. If $A \in \mathcal{M}$, let*

$$\int_A f(x)\, d\mu(x) = \int_A f d\mu := \int_X 1_A f \ d\mu.$$

Remark 19.14. Because of item 3. of Proposition 19.12, if ϕ is a non-negative simple function, $\int_X \phi d\mu = I_\mu(\phi)$ so that $\int_X$ is an extension of I_μ. This extension still has the monotonicity property if I_μ : namely if $0 \leq f \leq g$ then

$$\begin{aligned} \int_X f d\mu &= \sup\{I_\mu(\phi) : \phi \text{ is simple and } \phi \leq f\} \\ &\leq \sup\{I_\mu(\phi) : \phi \text{ is simple and } \phi \leq g\} \leq \int_X g d\mu. \end{aligned}$$

Similarly if $c > 0$,

$$\int_X c f d\mu = c\int_X f d\mu.$$

Also notice that if f is integrable, then $\mu(\{f = \infty\}) = 0$.

Lemma 19.15 (Sums as Integrals). *Let X be a set and $\rho : X \to [0,\infty]$ be a function, let $\mu = \sum_{x\in X} \rho(x)\delta_x$ on $\mathcal{M} = 2^X$, i.e.*

$$\mu(A) = \sum_{x\in A} \rho(x).$$

If $f : X \to [0,\infty]$ is a function (which is necessarily measurable), then

$$\int_X f d\mu = \sum_X f\rho.$$

Proof. Suppose that $\phi : X \to [0,\infty)$ is a simple function, then $\phi = \sum_{z\in[0,\infty)} z1_{\{\phi=z\}}$ and

$$\sum_X \phi\rho = \sum_{x\in X} \rho(x) \sum_{z\in[0,\infty)} z1_{\{\phi=z\}}(x) = \sum_{z\in[0,\infty)} z \sum_{x\in X} \rho(x)1_{\{\phi=z\}}(x)$$
$$= \sum_{z\in[0,\infty)} z\mu(\{\phi=z\}) = \int_X \phi d\mu.$$

So if $\phi : X \to [0,\infty)$ is a simple function such that $\phi \le f$, then

$$\int_X \phi d\mu = \sum_X \phi\rho \le \sum_X f\rho.$$

Taking the sup over ϕ in this last equation then shows that

$$\int_X f d\mu \le \sum_X f\rho.$$

For the reverse inequality, let $\Lambda \subset\subset X$ be a finite set and $N \in (0,\infty)$. Set $f^N(x) = \min\{N, f(x)\}$ and let $\phi_{N,\Lambda}$ be the simple function given by $\phi_{N,\Lambda}(x) := 1_\Lambda(x)f^N(x)$. Because $\phi_{N,\Lambda}(x) \le f(x)$,

$$\sum_\Lambda f^N\rho = \sum_X \phi_{N,\Lambda}\rho = \int_X \phi_{N,\Lambda} d\mu \le \int_X f d\mu.$$

Since $f^N \uparrow f$ as $N \to \infty$, we may let $N \to \infty$ in this last equation to concluded

$$\sum_\Lambda f\rho \le \int_X f d\mu.$$

Since Λ is arbitrary, this implies

$$\sum_X f\rho \le \int_X f d\mu.$$

■

Theorem 19.16 (Monotone Convergence Theorem). *Suppose $f_n \in L^+$ is a sequence of functions such that $f_n \uparrow f$ (f is necessarily in L^+) then*

$$\int f_n \uparrow \int f \text{ as } n \to \infty.$$

Proof. Since $f_n \le f_m \le f$, for all $n \le m < \infty$,

$$\int f_n \le \int f_m \le \int f$$

from which if follows $\int f_n$ is increasing in n and

$$\lim_{n\to\infty} \int f_n \le \int f. \tag{19.9}$$

For the opposite inequality, let $\phi : X \to [0, \infty)$ be a simple function such that $0 \le \phi \le f$, $\alpha \in (0,1)$ and $X_n := \{f_n \ge \alpha\phi\}$. Notice that $X_n \uparrow X$ and $f_n \ge \alpha 1_{X_n}\phi$ and so by definition of $\int f_n$,

$$\int f_n \ge \int \alpha 1_{X_n}\phi = \alpha \int 1_{X_n}\phi. \tag{19.10}$$

Then using the continuity property of μ,

$$\begin{aligned}
\lim_{n\to\infty} \int 1_{X_n}\phi &= \lim_{n\to\infty} \int 1_{X_n} \sum_{y>0} y 1_{\{\phi=y\}} \\
&= \lim_{n\to\infty} \sum_{y>0} y\mu(X_n \cap \{\phi = y\}) = \sum_{y>0} y \lim_{n\to\infty} \mu(X_n \cap \{\phi = y\}) \\
&= \sum_{y>0} y \lim_{n\to\infty} \mu(\{\phi = y\}) = \int \phi.
\end{aligned}$$

This identity allows us to let $n \to \infty$ in Eq. (19.10) to conclude

$$\int_X \phi \le \frac{1}{\alpha} \lim_{n\to\infty} \int f_n.$$

Since this is true for all non-negative simple functions ϕ with $\phi \le f$;

$$\int f = \sup \left\{ \int_X \phi : \phi \text{ is simple and } \phi \le f \right\} \le \frac{1}{\alpha} \lim_{n\to\infty} \int f_n.$$

Because $\alpha \in (0,1)$ was arbitrary, it follows that $\int f \le \lim_{n\to\infty} \int f_n$ which combined with Eq. (19.9) proves the theorem. ∎

The following simple lemma will be use often in the sequel.

Lemma 19.17 (Chebyshev's Inequality). *Suppose that $f \ge 0$ is a measurable function, then for any $\varepsilon > 0$,*

$$\mu(f \ge \varepsilon) \le \frac{1}{\varepsilon} \int_X f d\mu. \tag{19.11}$$

In particular if $\int_X f d\mu < \infty$ then $\mu(f = \infty) = 0$ (i.e. $f < \infty$ a.e.) and the set $\{f > 0\}$ is σ – finite.

Proof. Since $1_{\{f\ge\varepsilon\}} \le 1_{\{f\ge\varepsilon\}} \frac{1}{\varepsilon} f \le \frac{1}{\varepsilon} f$,

$$\mu(f \ge \varepsilon) = \int_X 1_{\{f\ge\varepsilon\}} d\mu \le \int_X 1_{\{f\ge\varepsilon\}} \frac{1}{\varepsilon} f d\mu \le \frac{1}{\varepsilon} \int_X f d\mu.$$

If $M := \int_X f d\mu < \infty$, then

$$\mu(f = \infty) \leq \mu(f \geq n) \leq \frac{M}{n} \to 0 \text{ as } n \to \infty$$

and $\{f \geq 1/n\} \uparrow \{f > 0\}$ with $\mu(f \geq 1/n) \leq nM < \infty$ for all n. ■

Corollary 19.18. *If $f_n \in L^+$ is a sequence of functions then*

$$\int \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int f_n.$$

In particular, if $\sum_{n=1}^{\infty} \int f_n < \infty$ then $\sum_{n=1}^{\infty} f_n < \infty$ a.e.

Proof. First off we show that

$$\int (f_1 + f_2) = \int f_1 + \int f_2$$

by choosing non-negative simple function ϕ_n and ψ_n such that $\phi_n \uparrow f_1$ and $\psi_n \uparrow f_2$. Then $(\phi_n + \psi_n)$ is simple as well and $(\phi_n + \psi_n) \uparrow (f_1 + f_2)$ so by the monotone convergence theorem,

$$\begin{aligned}\int (f_1 + f_2) &= \lim_{n\to\infty} \int (\phi_n + \psi_n) = \lim_{n\to\infty} \left(\int \phi_n + \int \psi_n \right) \\ &= \lim_{n\to\infty} \int \phi_n + \lim_{n\to\infty} \int \psi_n = \int f_1 + \int f_2.\end{aligned}$$

Now to the general case. Let $g_N := \sum\limits_{n=1}^{N} f_n$ and $g = \sum\limits_{1}^{\infty} f_n$, then $g_N \uparrow g$ and so again by monotone convergence theorem and the additivity just proved,

$$\begin{aligned}\sum_{n=1}^{\infty} \int f_n &:= \lim_{N\to\infty} \sum_{n=1}^{N} \int f_n = \lim_{N\to\infty} \int \sum_{n=1}^{N} f_n \\ &= \lim_{N\to\infty} \int g_N = \int g =: \int \sum_{n=1}^{\infty} f_n.\end{aligned}$$

■

Remark 19.19. It is in the proof of this corollary (i.e. the linearity of the integral) that we really make use of the assumption that all of our functions are measurable. In fact the definition $\int f d\mu$ makes sense for **all** functions $f : X \to [0, \infty]$ not just measurable functions. Moreover the monotone convergence theorem holds in this generality with no change in the proof. However, in the proof of Corollary 19.18, we use the approximation Theorem 18.42 which relies heavily on the measurability of the functions to be approximated.

The following Lemma and the next Corollary are simple applications of Corollary 19.18.

Lemma 19.20 (The First Borell – Carntelli Lemma). *Let $(X, \mathcal{M}, \mu)$ be a measure space, $A_n \in \mathcal{M}$, and set*

$$\{A_n \text{ i.o.}\} = \{x \in X : x \in A_n \text{ for infinitely many } n\text{'s}\} = \bigcap_{N=1}^{\infty} \bigcup_{n \geq N} A_n.$$

If $\sum_{n=1}^{\infty} \mu(A_n) < \infty$ then $\mu(\{A_n \text{ i.o.}\}) = 0$.

Proof. (First Proof.) Let us first observe that

$$\{A_n \text{ i.o.}\} = \left\{x \in X : \sum_{n=1}^{\infty} 1_{A_n}(x) = \infty\right\}.$$

Hence if $\sum_{n=1}^{\infty} \mu(A_n) < \infty$ then

$$\infty > \sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} \int_X 1_{A_n}\, d\mu = \int_X \sum_{n=1}^{\infty} 1_{A_n}\, d\mu$$

implies that $\sum_{n=1}^{\infty} 1_{A_n}(x) < \infty$ for μ - a.e. x. That is to say $\mu(\{A_n \text{ i.o.}\}) = 0$. (Second Proof.) Of course we may give a strictly measure theoretic proof of this fact:

$$\begin{aligned}\mu(A_n \text{ i.o.}) &= \lim_{N\to\infty} \mu\left(\bigcup_{n \geq N} A_n\right)\\ &\leq \lim_{N\to\infty} \sum_{n \geq N} \mu(A_n)\end{aligned}$$

and the last limit is zero since $\sum_{n=1}^{\infty} \mu(A_n) < \infty$. ∎

Corollary 19.21. *Suppose that $(X, \mathcal{M}, \mu)$ is a measure space and $\{A_n\}_{n=1}^{\infty} \subset \mathcal{M}$ is a collection of sets such that $\mu(A_i \cap A_j) = 0$ for all $i \neq j$, then*

$$\mu\left(\cup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

Proof. Since

$$\begin{aligned}\mu\left(\cup_{n=1}^{\infty} A_n\right) &= \int_X 1_{\cup_{n=1}^{\infty} A_n}\, d\mu \text{ and}\\ \sum_{n=1}^{\infty} \mu(A_n) &= \int_X \sum_{n=1}^{\infty} 1_{A_n}\, d\mu\end{aligned}$$

it suffices to show

$$\sum_{n=1}^{\infty} 1_{A_n} = 1_{\cup_{n=1}^{\infty} A_n} \ \mu - \text{a.e.} \tag{19.12}$$

Now $\sum_{n=1}^{\infty} 1_{A_n} \geq 1_{\cup_{n=1}^{\infty} A_n}$ and $\sum_{n=1}^{\infty} 1_{A_n}(x) \neq 1_{\cup_{n=1}^{\infty} A_n}(x)$ iff $x \in A_i \cap A_j$ for some $i \neq j$, that is

$$\left\{ x : \sum_{n=1}^{\infty} 1_{A_n}(x) \neq 1_{\cup_{n=1}^{\infty} A_n}(x) \right\} = \cup_{i<j} A_i \cap A_j$$

and the latter set has measure 0 being the countable union of sets of measure zero. This proves Eq. (19.12) and hence the corollary. ■

Notation 19.22 *If m is Lebesgue measure on $\mathcal{B}_{\mathbb{R}}$, f is a non-negative Borel measurable function and $a < b$ with $a, b \in \bar{\mathbb{R}}$, we will often write $\int_a^b f(x)\, dx$ or $\int_a^b f dm$ for $\int_{(a,b] \cap \mathbb{R}} f dm$.*

Example 19.23. Suppose $-\infty < a < b < \infty$, $f \in C([a,b],[0,\infty))$ and m be Lebesgue measure on $\mathbb{R}$. Also let $\pi_k = \{a = a_0^k < a_1^k < \cdots < a_{n_k}^k = b\}$ be a sequence of refining partitions (i.e. $\pi_k \subset \pi_{k+1}$ for all k) such that

$$\operatorname{mesh}(\pi_k) := \max\{\left|a_j^k - a_{j-1}^{k+1}\right| : j = 1, \ldots, n_k\} \to 0 \text{ as } k \to \infty.$$

For each k, let

$$f_k(x) = f(a) 1_{\{a\}} + \sum_{l=0}^{n_k - 1} \min\left\{f(x) : a_l^k \leq x \leq a_{l+1}^k\right\} 1_{(a_l^k, a_{l+1}^k]}(x)$$

then $f_k \uparrow f$ as $k \to \infty$ and so by the monotone convergence theorem,

$$\begin{aligned} \int_a^b f dm &:= \int_{[a,b]} f dm = \lim_{k \to \infty} \int_a^b f_k \ dm \\ &= \lim_{k \to \infty} \sum_{l=0}^{n_k - 1} \min\left\{f(x) : a_l^k \leq x \leq a_{l+1}^k\right\} m\left((a_l^k, a_{l+1}^k]\right) \\ &= \int_a^b f(x) dx. \end{aligned}$$

The latter integral being the Riemann integral.

We can use the above result to integrate some non-Riemann integrable functions:

Example 19.24. For all $\lambda > 0$,

$$\int_0^{\infty} e^{-\lambda x} dm(x) = \lambda^{-1} \text{ and } \int_{\mathbb{R}} \frac{1}{1 + x^2} dm(x) = \pi.$$

The proof of these identities are similar. By the monotone convergence theorem, Example 19.23 and the fundamental theorem of calculus for Riemann integrals (or see Theorem 8.13 above or Theorem 19.40 below),

$$\begin{aligned}\int_0^\infty e^{-\lambda x} dm(x) &= \lim_{N\to\infty}\int_0^N e^{-\lambda x} dm(x) = \lim_{N\to\infty}\int_0^N e^{-\lambda x} dx \\ &= -\lim_{N\to\infty}\frac{1}{\lambda}e^{-\lambda x}|_0^N = \lambda^{-1}\end{aligned}$$

and

$$\begin{aligned}\int_{\mathbb{R}}\frac{1}{1+x^2}dm(x) &= \lim_{N\to\infty}\int_{-N}^N\frac{1}{1+x^2}dm(x) = \lim_{N\to\infty}\int_{-N}^N\frac{1}{1+x^2}dx \\ &= \lim_{N\to\infty}\left[\tan^{-1}(N) - \tan^{-1}(-N)\right] = \pi.\end{aligned}$$

Let us also consider the functions x^{-p},

$$\begin{aligned}\int_{(0,1]}\frac{1}{x^p}\,dm(x) &= \lim_{n\to\infty}\int_0^1 1_{(\frac{1}{n},1]}(x)\frac{1}{x^p}dm(x) \\ &= \lim_{n\to\infty}\int_{\frac{1}{n}}^1\frac{1}{x^p}dx = \lim_{n\to\infty}\left.\frac{x^{-p+1}}{1-p}\right|_{1/n}^1 \\ &= \begin{cases}\frac{1}{1-p} & \text{if } p<1 \\ \infty & \text{if } p>1\end{cases}\end{aligned}$$

If $p=1$ we find

$$\int_{(0,1]}\frac{1}{x^p}\,dm(x) = \lim_{n\to\infty}\int_{\frac{1}{n}}^1\frac{1}{x}dx = \lim_{n\to\infty}\ln(x)|_{1/n}^1 = \infty.$$

Example 19.25. Let $\{r_n\}_{n=1}^\infty$ be an enumeration of the points in $\mathbb{Q}\cap[0,1]$ and define

$$f(x) = \sum_{n=1}^\infty 2^{-n}\frac{1}{\sqrt{|x-r_n|}}$$

with the convention that

$$\frac{1}{\sqrt{|x-r_n|}} = 5 \text{ if } x = r_n.$$

Since, By Theorem 19.40,

$$\begin{aligned}\int_0^1\frac{1}{\sqrt{|x-r_n|}}\,dx &= \int_{r_n}^1\frac{1}{\sqrt{x-r_n}}dx + \int_0^{r_n}\frac{1}{\sqrt{r_n-x}}dx \\ &= 2\sqrt{x-r_n}|_{r_n}^1 - 2\sqrt{r_n-x}|_0^{r_n} = 2\left(\sqrt{1-r_n}-\sqrt{r_n}\right) \\ &\le 4,\end{aligned}$$

we find

$$\int_{[0,1]} f(x)dm(x) = \sum_{n=1}^{\infty} 2^{-n} \int_{[0,1]} \frac{1}{\sqrt{|x-r_n|}} dx \leq \sum_{n=1}^{\infty} 2^{-n} 4 = 4 < \infty.$$

In particular, $m(f = \infty) = 0$, i.e. that $f < \infty$ for almost every $x \in [0,1]$ and this implies that

$$\sum_{n=1}^{\infty} 2^{-n} \frac{1}{\sqrt{|x-r_n|}} < \infty \text{ for a.e. } x \in [0,1].$$

This result is somewhat surprising since the singularities of the summands form a dense subset of $[0,1]$.

Proposition 19.26. *Suppose that $f \geq 0$ is a measurable function. Then $\int_X f d\mu = 0$ iff $f = 0$ a.e. Also if $f, g \geq 0$ are measurable functions such that $f \leq g$ a.e. then $\int f d\mu \leq \int g d\mu$. In particular if $f = g$ a.e. then $\int f d\mu = \int g d\mu$.*

Proof. If $f = 0$ a.e. and $\phi \leq f$ is a simple function then $\phi = 0$ a.e. This implies that $\mu(\phi^{-1}(\{y\})) = 0$ for all $y > 0$ and hence $\int_X \phi d\mu = 0$ and therefore $\int_X f d\mu = 0$. Conversely, if $\int f d\mu = 0$, then by (Lemma 19.17),

$$\mu(f \geq 1/n) \leq n \int f d\mu = 0 \text{ for all } n.$$

Therefore, $\mu(f > 0) \leq \sum_{n=1}^{\infty} \mu(f \geq 1/n) = 0$, i.e. $f = 0$ a.e. For the second assertion let E be the exceptional set where $f > g$, i.e. $E := \{x \in X : f(x) > g(x)\}$. By assumption E is a null set and $1_{E^c} f \leq 1_{E^c} g$ everywhere. Because $g = 1_{E^c} g + 1_E g$ and $1_E g = 0$ a.e.,

$$\int g d\mu = \int 1_{E^c} g d\mu + \int 1_E g d\mu = \int 1_{E^c} g d\mu$$

and similarly $\int f d\mu = \int 1_{E^c} f d\mu$. Since $1_{E^c} f \leq 1_{E^c} g$ everywhere,

$$\int f d\mu = \int 1_{E^c} f d\mu \leq \int 1_{E^c} g d\mu = \int g d\mu.$$

■

Corollary 19.27. *Suppose that $\{f_n\}$ is a sequence of non-negative measurable functions and f is a measurable function such that $f_n \uparrow f$ off a null set, then*

$$\int f_n \uparrow \int f \text{ as } n \to \infty.$$

Proof. Let $E \subset X$ be a null set such that $f_n 1_{E^c} \uparrow f 1_{E^c}$ as $n \to \infty$. Then by the monotone convergence theorem and Proposition 19.26,

$$\int f_n = \int f_n 1_{E^c} \uparrow \int f 1_{E^c} = \int f \text{ as } n \to \infty.$$

■

Lemma 19.28 (Fatou's Lemma). *If $f_n : X \to [0,\infty]$ is a sequence of measurable functions then*

$$\int \liminf_{n\to\infty} f_n \le \liminf_{n\to\infty} \int f_n$$

Proof. Define $g_k := \inf_{n\ge k} f_n$ so that $g_k \uparrow \liminf_{n\to\infty} f_n$ as $k \to \infty$. Since $g_k \le f_n$ for all $k \le n$,

$$\int g_k \le \int f_n \text{ for all } n \ge k$$

and therefore

$$\int g_k \le \lim\inf_{n\to\infty} \int f_n \text{ for all } k.$$

We may now use the monotone convergence theorem to let $k \to \infty$ to find

$$\int \lim\inf_{n\to\infty} f_n = \int \lim_{k\to\infty} g_k \overset{\text{MCT}}{=} \lim_{k\to\infty} \int g_k \le \lim\inf_{n\to\infty} \int f_n.$$

■

19.4 Integrals of Complex Valued Functions

Definition 19.29. *A measurable function $f : X \to \bar{\mathbb{R}}$ is* ***integrable*** *if $f_+ := f 1_{\{f\ge 0\}}$ and $f_- = -f\ 1_{\{f\le 0\}}$ are* ***integrable****. We write $\mathrm{L}^1(\mu;\mathbb{R})$ for the space of real valued integrable functions. For $f \in \mathrm{L}^1(\mu;\mathbb{R})$, let*

$$\int f d\mu = \int f_+ d\mu - \int f_- d\mu$$

Convention: If $f, g : X \to \bar{\mathbb{R}}$ are two measurable functions, let $f+g$ denote the collection of measurable functions $h : X \to \bar{\mathbb{R}}$ such that $h(x) = f(x)+g(x)$ whenever $f(x)+g(x)$ is well defined, i.e. is not of the form $\infty-\infty$ or $-\infty+\infty$. We use a similar convention for $f-g$. Notice that if $f,g \in \mathrm{L}^1(\mu;\mathbb{R})$ and $h_1, h_2 \in f+g$, then $h_1 = h_2$ a.e. because $|f| < \infty$ and $|g| < \infty$ a.e.

Notation 19.30 (Abuse of notation) *We will sometimes denote the integral $\int_X f d\mu$ by $\mu(f)$. With this notation we have $\mu(A) = \mu(1_A)$ for all $A \in \mathcal{M}$.*

Remark 19.31. Since

$$f_\pm \le |f| \le f_+ + f_-,$$

a measurable function f is **integrable** iff $\int |f|\ d\mu < \infty$. Hence

$$\mathrm{L}^1(\mu;\mathbb{R}) := \left\{ f : X \to \bar{\mathbb{R}} :\ f \text{ is measurable and } \int_X |f|\ d\mu < \infty \right\}.$$

If $f, g \in \mathrm{L}^1(\mu;\mathbb{R})$ and $f = g$ a.e. then $f_\pm = g_\pm$ a.e. and so it follows from Proposition 19.26 that $\int f d\mu = \int g d\mu$. In particular if $f, g \in \mathrm{L}^1(\mu;\mathbb{R})$ we may define

$$\int_X (f+g)\, d\mu = \int_X h d\mu$$

where h is any element of $f + g$.

Proposition 19.32. *The map*

$$f \in \mathrm{L}^1(\mu;\mathbb{R}) \to \int_X f d\mu \in \mathbb{R}$$

is linear and has the monotonicity property: $\int f d\mu \leq \int g d\mu$ *for all* $f, g \in \mathrm{L}^1(\mu;\mathbb{R})$ *such that* $f \leq g$ *a.e.*

Proof. Let $f, g \in \mathrm{L}^1(\mu;\mathbb{R})$ and $a, b \in \mathbb{R}$. By modifying f and g on a null set, we may assume that f, g are real valued functions. We have $af + bg \in \mathrm{L}^1(\mu;\mathbb{R})$ because

$$|af + bg| \leq |a|\,|f| + |b|\,|g| \in \mathrm{L}^1(\mu;\mathbb{R}).$$

If $a < 0$, then

$$(af)_+ = -af_- \text{ and } (af)_- = -af_+$$

so that

$$\int af = -a\int f_- + a\int f_+ = a(\int f_+ - \int f_-) = a\int f.$$

A similar calculation works for $a > 0$ and the case $a = 0$ is trivial so we have shown that

$$\int af = a\int f.$$

Now set $h = f + g$. Since $h = h_+ - h_-$,

$$h_+ - h_- = f_+ - f_- + g_+ - g_-$$

or

$$h_+ + f_- + g_- = h_- + f_+ + g_+.$$

Therefore,

$$\int h_+ + \int f_- + \int g_- = \int h_- + \int f_+ + \int g_+$$

and hence

$$\int h = \int h_+ - \int h_- = \int f_+ + \int g_+ - \int f_- - \int g_- = \int f + \int g.$$

Finally if $f_+ - f_- = f \leq g = g_+ - g_-$ then $f_+ + g_- \leq g_+ + f_-$ which implies that

$$\int f_+ + \int g_- \le \int g_+ + \int f_-$$

or equivalently that

$$\int f = \int f_+ - \int f_- \le \int g_+ - \int g_- = \int g.$$

The monotonicity property is also a consequence of the linearity of the integral, the fact that $f \le g$ a.e. implies $0 \le g - f$ a.e. and Proposition 19.26. ∎

Definition 19.33. *A measurable function* $f : X \to \mathbb{C}$ *is* ***integrable*** *if* $\int_X |f| \; d\mu < \infty$. *Analogously to the real case, let*

$$\mathrm{L}^1(\mu;\mathbb{C}) := \left\{ f : X \to \mathbb{C} : \; f \text{ is measurable and } \int_X |f| \; d\mu < \infty \right\}.$$

denote the complex valued integrable functions. Because, $\max(|\mathrm{Re}\, f|, |\mathrm{Im}\, f|) \le |f| \le \sqrt{2}\max(|\mathrm{Re}\, f|, |\mathrm{Im}\, f|)$, $\int |f| \; d\mu < \infty$ *iff*

$$\int |\mathrm{Re}\, f| \, d\mu + \int |\mathrm{Im}\, f| \, d\mu < \infty.$$

For $f \in \mathrm{L}^1(\mu;\mathbb{C})$ *define*

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

It is routine to show the integral is still linear on $\mathrm{L}^1(\mu;\mathbb{C})$ (prove!). In the remainder of this section, let $\mathrm{L}^1(\mu)$ be either $\mathrm{L}^1(\mu;\mathbb{C})$ or $\mathrm{L}^1(\mu;\mathbb{R})$. If $A \in \mathcal{M}$ and $f \in \mathrm{L}^1(\mu;\mathbb{C})$ or $f : X \to [0,\infty]$ is a measurable function, let

$$\int_A f d\mu := \int_X 1_A f d\mu.$$

Proposition 19.34. *Suppose that* $f \in \mathrm{L}^1(\mu;\mathbb{C})$, *then*

$$\left| \int_X f d\mu \right| \le \int_X |f| \, d\mu. \tag{19.13}$$

Proof. Start by writing $\int_X f \; d\mu = Re^{i\theta}$ with $R \ge 0$. We may assume that $R = \left|\int_X f d\mu\right| > 0$ since otherwise there is nothing to prove. Since

$$R = e^{-i\theta} \int_X f \; d\mu = \int_X e^{-i\theta} f \; d\mu = \int_X \mathrm{Re}\left(e^{-i\theta} f\right) d\mu + i \int_X \mathrm{Im}\left(e^{-i\theta} f\right) d\mu,$$

it must be that $\int_X \mathrm{Im}\left[e^{-i\theta} f\right] d\mu = 0$. Using the monotonicity in Proposition 19.26,

$$\left| \int_X f d\mu \right| = \int_X \mathrm{Re}\left(e^{-i\theta} f\right) d\mu \le \int_X \left|\mathrm{Re}\left(e^{-i\theta} f\right)\right| d\mu \le \int_X |f| \, d\mu.$$

∎

Proposition 19.35. *Let $f, g \in \mathrm{L}^1(\mu)$, then*

1. *The set $\{f \neq 0\}$ is σ – finite, in fact $\{|f| \geq \frac{1}{n}\} \uparrow \{f \neq 0\}$ and $\mu(|f| \geq \frac{1}{n}) < \infty$ for all n.*
2. *The following are equivalent*
   a) $\int_E f = \int_E g$ for all $E \in \mathcal{M}$
   b) $\int\limits_X |f - g| = 0$
   c) $f = g$ a.e.

Proof. 1. By Chebyshev's inequality, Lemma 19.17,

$$\mu(|f| \geq \frac{1}{n}) \leq n \int_X |f| \, d\mu < \infty$$

for all n. 2. (a) $\Longrightarrow$ (c) Notice that

$$\int_E f = \int_E g \Leftrightarrow \int_E (f - g) = 0$$

for all $E \in \mathcal{M}$. Taking $E = \{\mathrm{Re}(f - g) > 0\}$ and using $1_E \mathrm{Re}(f - g) \geq 0$, we learn that

$$0 = \mathrm{Re} \int_E (f - g) d\mu = \int 1_E \mathrm{Re}(f - g) \Longrightarrow 1_E \mathrm{Re}(f - g) = 0 \text{ a.e.}$$

This implies that $1_E = 0$ a.e. which happens iff

$$\mu(\{\mathrm{Re}(f - g) > 0\}) = \mu(E) = 0.$$

Similar $\mu(\mathrm{Re}(f-g) < 0) = 0$ so that $\mathrm{Re}(f-g) = 0$ a.e. Similarly, $\mathrm{Im}(f-g) = 0$ a.e and hence $f - g = 0$ a.e., i.e. $f = g$ a.e. (c) $\Longrightarrow$ (b) is clear and so is (b) $\Longrightarrow$ (a) since

$$\left| \int_E f - \int_E g \right| \leq \int |f - g| = 0.$$

■

Definition 19.36. *Let $(X, \mathcal{M}, \mu)$ be a measure space and $L^1(\mu) = L^1(X, \mathcal{M}, \mu)$ denote the set of $\mathrm{L}^1(\mu)$ functions modulo the equivalence relation; $f \sim g$ iff $f = g$ a.e. We make this into a normed space using the norm*

$$\|f - g\|_{L^1} = \int |f - g| \, d\mu$$

and into a metric space using $\rho_1(f, g) = \|f - g\|_{L^1}$.

Warning: in the future we will often not make much of a distinction between $L^1(\mu)$ and $\mathrm{L}^1(\mu)$. On occasion this can be dangerous and this danger will be pointed out when necessary.

Remark 19.37. More generally we may define $L^p(\mu) = L^p(X, \mathcal{M}, \mu)$ for $p \in [1, \infty)$ as the set of measurable functions f such that

$$\int_X |f|^p \, d\mu < \infty$$

modulo the equivalence relation; $f \sim g$ iff $f = g$ a.e.

We will see in Chapter 21 that

$$\|f\|_{L^p} = \left(\int |f|^p \, d\mu \right)^{1/p} \quad \text{for } f \in L^p(\mu)$$

is a norm and $(L^p(\mu), \|\cdot\|_{L^p})$ is a Banach space in this norm.

Theorem 19.38 (Dominated Convergence Theorem). *Suppose $f_n, g_n, g \in \mathrm{L}^1(\mu)$, $f_n \to f$ a.e., $|f_n| \leq g_n \in \mathrm{L}^1(\mu)$, $g_n \to g$ a.e. and $\int_X g_n d\mu \to \int_X g d\mu$. Then $f \in \mathrm{L}^1(\mu)$ and*

$$\int_X f d\mu = \lim_{h \to \infty} \int_X f_n d\mu.$$

(In most typical applications of this theorem $g_n = g \in \mathrm{L}^1(\mu)$ for all n.)

Proof. Notice that $|f| = \lim_{n\to\infty} |f_n| \leq \lim_{n\to\infty} |g_n| \leq g$ a.e. so that $f \in \mathrm{L}^1(\mu)$. By considering the real and imaginary parts of f separately, it suffices to prove the theorem in the case where f is real. By Fatou's Lemma,

$$\begin{aligned}
\int_X (g \pm f) d\mu &= \int_X \liminf_{n\to\infty} (g_n \pm f_n) \, d\mu \leq \liminf_{n\to\infty} \int_X (g_n \pm f_n) \, d\mu \\
&= \lim_{n\to\infty} \int_X g_n d\mu + \liminf_{n\to\infty} \left(\pm \int_X f_n d\mu \right) \\
&= \int_X g d\mu + \liminf_{n\to\infty} \left(\pm \int_X f_n d\mu \right)
\end{aligned}$$

Since $\liminf_{n\to\infty}(-a_n) = -\limsup_{n\to\infty} a_n$, we have shown,

$$\int_X g d\mu \pm \int_X f d\mu \leq \int_X g d\mu + \begin{cases} \liminf_{n\to\infty} \int_X f_n d\mu \\ -\limsup_{n\to\infty} \int_X f_n d\mu \end{cases}$$

and therefore

$$\limsup_{n\to\infty} \int_X f_n d\mu \leq \int_X f d\mu \leq \liminf_{n\to\infty} \int_X f_n d\mu.$$

This shows that $\lim\limits_{n\to\infty} \int_X f_n d\mu$ exists and is equal to $\int_X f d\mu$. ∎

Exercise 19.1. Give another proof of Proposition 19.34 by first proving Eq. (19.13) with f being a cylinder function in which case the triangle inequality for complex numbers will do the trick. Then use the approximation Theorem 18.42 along with the dominated convergence Theorem 19.38 to handle the general case.

Corollary 19.39. *Let $\{f_n\}_{n=1}^\infty \subset \mathrm{L}^1(\mu)$ be a sequence such that $\sum_{n=1}^\infty \|f_n\|_{\mathrm{L}^1(\mu)} < \infty$, then $\sum_{n=1}^\infty f_n$ is convergent a.e. and*

$$\int_X \left(\sum_{n=1}^\infty f_n\right) d\mu = \sum_{n=1}^\infty \int_X f_n d\mu.$$

Proof. The condition $\sum_{n=1}^\infty \|f_n\|_{\mathrm{L}^1(\mu)} < \infty$ is equivalent to $\sum_{n=1}^\infty |f_n| \in \mathrm{L}^1(\mu)$. Hence $\sum_{n=1}^\infty f_n$ is almost everywhere convergent and if $S_N := \sum_{n=1}^N f_n$, then

$$|S_N| \leq \sum_{n=1}^N |f_n| \leq \sum_{n=1}^\infty |f_n| \in \mathrm{L}^1(\mu).$$

So by the dominated convergence theorem,

$$\begin{aligned}\int_X \left(\sum_{n=1}^\infty f_n\right) d\mu &= \int_X \lim_{N\to\infty} S_N d\mu = \lim_{N\to\infty}\int_X S_N d\mu \\ &= \lim_{N\to\infty}\sum_{n=1}^N \int_X f_n d\mu = \sum_{n=1}^\infty \int_X f_n d\mu.\end{aligned}$$

■

Theorem 19.40 (The Fundamental Theorem of Calculus). *Suppose $-\infty < a < b < \infty$, $f \in C((a,b),\mathbb{R}) \cap L^1((a,b),m)$ and $F(x) := \int_a^x f(y)dm(y)$. Then*

1. *$F \in C([a,b],\mathbb{R}) \cap C^1((a,b),\mathbb{R})$.*
2. *$F'(x) = f(x)$ for all $x \in (a,b)$.*
3. *If $G \in C([a,b],\mathbb{R}) \cap C^1((a,b),\mathbb{R})$ is an anti-derivative of f on (a,b) (i.e. $f = G'|_{(a,b)}$) then*

$$\int_a^b f(x)dm(x) = G(b) - G(a).$$

Proof. Since $F(x) := \int_{\mathbb{R}} 1_{(a,x)}(y)f(y)dm(y)$, $\lim_{x\to z} 1_{(a,x)}(y) = 1_{(a,z)}(y)$ for m – a.e. y and $\left|1_{(a,x)}(y)f(y)\right| \leq 1_{(a,b)}(y)\,|f(y)|$ is an L^1 – function, it follows from the dominated convergence Theorem 19.38 that F is continuous on $[a,b]$. Simple manipulations show,

$$\begin{aligned}\left|\frac{F(x+h)-F(x)}{h} - f(x)\right| &= \frac{1}{|h|}\begin{cases}\left|\int_x^{x+h}[f(y)-f(x)]\,dm(y)\right| & \text{if } h>0 \\ \left|\int_{x+h}^x [f(y)-f(x)]\,dm(y)\right| & \text{if } h<0\end{cases} \\ &\leq \frac{1}{|h|}\begin{cases}\int_x^{x+h}|f(y)-f(x)|\,dm(y) \text{ if } h>0 \\ \int_{x+h}^x |f(y)-f(x)|\,dm(y) \text{ if } h<0\end{cases} \\ &\leq \sup\{|f(y)-f(x)| : y \in [x-|h|, x+|h|]\}\end{aligned}$$

and the latter expression, by the continuity of f, goes to zero as $h \to 0$. This shows $F' = f$ on (a, b). For the converse direction, we have by assumption that $G'(x) = F'(x)$ for $x \in (a, b)$. Therefore by the mean value theorem, $F - G = C$ for some constant C. Hence

$$\begin{aligned}\int_a^b f(x)dm(x) &= F(b) = F(b) - F(a) \\ &= (G(b) + C) - (G(a) + C) = G(b) - G(a).\end{aligned}$$

■

Example 19.41. The following limit holds,

$$\lim_{n\to\infty} \int_0^n (1 - \frac{x}{n})^n dm(x) = 1.$$

Let $f_n(x) = (1 - \frac{x}{n})^n 1_{[0,n]}(x)$ and notice that $\lim_{n\to\infty} f_n(x) = e^{-x}$. We will now show

$$0 \le f_n(x) \le e^{-x} \text{ for all } x \ge 0.$$

It suffices to consider $x \in [0, n]$. Let $g(x) = e^x f_n(x)$, then for $x \in (0, n)$,

$$\frac{d}{dx} \ln g(x) = 1 + n \frac{1}{(1 - \frac{x}{n})}(-\frac{1}{n}) = 1 - \frac{1}{(1 - \frac{x}{n})} \le 0$$

which shows that $\ln g(x)$ and hence $g(x)$ is decreasing on $[0, n]$. Therefore $g(x) \le g(0) = 1$, i.e.

$$0 \le f_n(x) \le e^{-x}.$$

From Example 19.24, we know

$$\int_0^\infty e^{-x} dm(x) = 1 < \infty,$$

so that e^{-x} is an integrable function on $[0, \infty)$. Hence by the dominated convergence theorem,

$$\begin{aligned}\lim_{n\to\infty} \int_0^n (1 - \frac{x}{n})^n dm(x) &= \lim_{n\to\infty} \int_0^\infty f_n(x) dm(x) \\ &= \int_0^\infty \lim_{n\to\infty} f_n(x) dm(x) = \int_0^\infty e^{-x} dm(x) = 1.\end{aligned}$$

Example 19.42 (Integration of Power Series). Suppose $R > 0$ and $\{a_n\}_{n=0}^\infty$ is a sequence of complex numbers such that $\sum_{n=0}^\infty |a_n| r^n < \infty$ for all $r \in (0, R)$. Then

$$\int_\alpha^\beta \left(\sum_{n=0}^\infty a_n x^n \right) dm(x) = \sum_{n=0}^\infty a_n \int_\alpha^\beta x^n dm(x) = \sum_{n=0}^\infty a_n \frac{\beta^{n+1} - \alpha^{n+1}}{n+1}$$

for all $-R < \alpha < \beta < R$. Indeed this follows from Corollary 19.39 since

$$\sum_{n=0}^{\infty} \int_{\alpha}^{\beta} |a_n| \, |x|^n \, dm(x) \leq \sum_{n=0}^{\infty} \left(\int_0^{|\beta|} |a_n| \, |x|^n \, dm(x) + \int_0^{|\alpha|} |a_n| \, |x|^n \, dm(x) \right)$$
$$\leq \sum_{n=0}^{\infty} |a_n| \frac{|\beta|^{n+1} + |\alpha|^{n+1}}{n+1} \leq 2r \sum_{n=0}^{\infty} |a_n| \, r^n < \infty$$

where $r = \max(|\beta|, |\alpha|)$.

Corollary 19.43 (Differentiation Under the Integral). *Suppose that $J \subset \mathbb{R}$ is an open interval and $f : J \times X \to \mathbb{C}$ is a function such that*

1. *$x \to f(t, x)$ is measurable for each $t \in J$.*
2. *$f(t_0, \cdot) \in L^1(\mu)$ for some $t_0 \in J$.*
3. *$\frac{\partial f}{\partial t}(t, x)$ exists for all (t, x).*
4. *There is a function $g \in \mathrm{L}^1(\mu)$ such that $\left| \frac{\partial f}{\partial t}(t, \cdot) \right| \leq g \in \mathrm{L}^1(\mu)$ for each $t \in J$.*
   Then $f(t, \cdot) \in \mathrm{L}^1(\mu)$ for all $t \in J$ (i.e. $\int_X |f(t, x)| \, d\mu(x) < \infty$), $t \to \int_X f(t, x) d\mu(x)$ is a differentiable function on J and

$$\frac{d}{dt} \int_X f(t, x) d\mu(x) = \int_X \frac{\partial f}{\partial t}(t, x) d\mu(x).$$

Proof. (The proof is essentially the same as for sums.) By considering the real and imaginary parts of f separately, we may assume that f is real. Also notice that

$$\frac{\partial f}{\partial t}(t, x) = \lim_{n \to \infty} n(f(t + n^{-1}, x) - f(t, x))$$

and therefore, for $x \to \frac{\partial f}{\partial t}(t, x)$ is a sequential limit of measurable functions and hence is measurable for all $t \in J$. By the mean value theorem,

$$|f(t, x) - f(t_0, x)| \leq g(x) \, |t - t_0| \text{ for all } t \in J \tag{19.14}$$

and hence

$$|f(t, x)| \leq |f(t, x) - f(t_0, x)| + |f(t_0, x)| \leq g(x) \, |t - t_0| + |f(t_0, x)| \, .$$

This shows $f(t, \cdot) \in \mathrm{L}^1(\mu)$ for all $t \in J$. Let $G(t) := \int_X f(t, x) d\mu(x)$, then

$$\frac{G(t) - G(t_0)}{t - t_0} = \int_X \frac{f(t, x) - f(t_0, x)}{t - t_0} d\mu(x).$$

By assumption,

$$\lim_{t \to t_0} \frac{f(t, x) - f(t_0, x)}{t - t_0} = \frac{\partial f}{\partial t}(t, x) \text{ for all } x \in X$$

and by Eq. (19.14),

$$\left|\frac{f(t,x)-f(t_0,x)}{t-t_0}\right| \le g(x) \text{ for all } t\in J \text{ and } x\in X.$$

Therefore, we may apply the dominated convergence theorem to conclude

$$\begin{aligned}\lim_{n\to\infty}\frac{G(t_n)-G(t_0)}{t_n-t_0} &= \lim_{n\to\infty}\int_X \frac{f(t_n,x)-f(t_0,x)}{t_n-t_0}d\mu(x)\\ &= \int_X \lim_{n\to\infty}\frac{f(t_n,x)-f(t_0,x)}{t_n-t_0}d\mu(x)\\ &= \int_X \frac{\partial f}{\partial t}(t_0,x)d\mu(x)\end{aligned}$$

for **all** sequences $t_n \in J\setminus\{t_0\}$ such that $t_n \to t_0$. Therefore, $\dot G(t_0) = \lim_{t\to t_0}\frac{G(t)-G(t_0)}{t-t_0}$ exists and

$$\dot G(t_0) = \int_X \frac{\partial f}{\partial t}(t_0,x)d\mu(x).$$

■

Example 19.44. Recall from Example 19.24 that

$$\lambda^{-1} = \int_{[0,\infty)} e^{-\lambda x}dm(x) \text{ for all } \lambda>0.$$

Let $\varepsilon>0$. For $\lambda\ge 2\varepsilon>0$ and $n\in\mathbb{N}$ there exists $C_n(\varepsilon)<\infty$ such that

$$0\le\left(-\frac{d}{d\lambda}\right)^n e^{-\lambda x} = x^n e^{-\lambda x}\le C(\varepsilon)e^{-\varepsilon x}.$$

Using this fact, Corollary 19.43 and induction gives

$$\begin{aligned}n!\lambda^{-n-1} &= \left(-\frac{d}{d\lambda}\right)^n\lambda^{-1} = \int_{[0,\infty)}\left(-\frac{d}{d\lambda}\right)^n e^{-\lambda x}dm(x)\\ &= \int_{[0,\infty)} x^n e^{-\lambda x}dm(x).\end{aligned}$$

That is $n! = \lambda^n\int_{[0,\infty)} x^n e^{-\lambda x}dm(x)$. Recall that

$$\Gamma(t) := \int_{[0,\infty)} x^{t-1}e^{-x}dx \text{ for } t>0.$$

(The reader should check that $\Gamma(t)<\infty$ for all $t>0$.) We have just shown that $\Gamma(n+1)=n!$ for all $n\in\mathbb{N}$.

Remark 19.45. Corollary 19.43 may be generalized by allowing the hypothesis to hold for $x \in X \setminus E$ where $E \in \mathcal{M}$ is a **fixed** null set, i.e. E must be independent of t. Consider what happens if we formally apply Corollary 19.43 to $g(t) := \int_0^\infty 1_{x \le t} dm(x)$,

$$\dot{g}(t) = \frac{d}{dt} \int_0^\infty 1_{x \le t} dm(x) \stackrel{?}{=} \int_0^\infty \frac{\partial}{\partial t} 1_{x \le t} dm(x).$$

The last integral is zero since $\frac{\partial}{\partial t} 1_{x \le t} = 0$ unless $t = x$ in which case it is not defined. On the other hand $g(t) = t$ so that $\dot{g}(t) = 1$. (The reader should decide which hypothesis of Corollary 19.43 has been violated in this example.)

19.5 Measurability on Complete Measure Spaces

In this subsection we will discuss a couple of measurability results concerning completions of measure spaces.

Proposition 19.46. *Suppose that $(X, \mathcal{M}, \mu)$ is a complete measure space*[1] *and $f : X \to \mathbb{R}$ is measurable.*

1. *If $g : X \to \mathbb{R}$ is a function such that $f(x) = g(x)$ for μ – a.e. x, then g is measurable.*
2. *If $f_n : X \to \mathbb{R}$ are measurable and $f : X \to \mathbb{R}$ is a function such that $\lim_{n\to\infty} f_n = f$, μ - a.e., then f is measurable as well.*

Proof. 1. Let $E = \{x : f(x) \neq g(x)\}$ which is assumed to be in $\mathcal{M}$ and $\mu(E) = 0$. Then $g = 1_{E^c} f + 1_E g$ since $f = g$ on E^c. Now $1_{E^c} f$ is measurable so g will be measurable if we show $1_E g$ is measurable. For this consider,

$$(1_E g)^{-1}(A) = \begin{cases} E^c \cup (1_E g)^{-1}(A \setminus \{0\}) & \text{if } 0 \in A \\ (1_E g)^{-1}(A) & \text{if } 0 \notin A \end{cases} \tag{19.15}$$

Since $(1_E g)^{-1}(B) \subset E$ if $0 \notin B$ and $\mu(E) = 0$, it follow by completeness of $\mathcal{M}$ that $(1_E g)^{-1}(B) \in \mathcal{M}$ if $0 \notin B$. Therefore Eq. (19.15) shows that $1_E g$ is measurable. 2. Let $E = \{x : \lim_{n\to\infty} f_n(x) \neq f(x)\}$ by assumption $E \in \mathcal{M}$ and $\mu(E) = 0$. Since $g := 1_E f = \lim_{n\to\infty} 1_{E^c} f_n$, g is measurable. Because $f = g$ on E^c and $\mu(E) = 0$, $f = g$ a.e. so by part 1. f is also measurable. ∎

The above results are in general false if $(X, \mathcal{M}, \mu)$ is not complete. For example, let $X = \{0, 1, 2\}$, $\mathcal{M} = \{\{0\}, \{1, 2\}, X, \phi\}$ and $\mu = \delta_0$. Take $g(0) = 0$, $g(1) = 1$, $g(2) = 2$, then $g = 0$ a.e. yet g is not measurable.

[1] Recall this means that if $N \subset X$ is a set such that $N \subset A \in \mathcal{M}$ and $\mu(A) = 0$, then $N \in \mathcal{M}$ as well.

Lemma 19.47. *Suppose that* $(X, \mathcal{M}, \mu)$ *is a measure space and* $\bar{\mathcal{M}}$ *is the completion of* $\mathcal{M}$ *relative to* μ *and* $\bar{\mu}$ *is the extension of* μ *to* $\bar{\mathcal{M}}$. *Then a function* $f : X \to \mathbb{R}$ *is* $(\bar{\mathcal{M}}, \mathcal{B} = \mathcal{B}_{\mathbb{R}})$ *– measurable iff there exists a function* $g : X \to \mathbb{R}$ *that is* $(\mathcal{M}, \mathcal{B})$ *– measurable such* $E = \{x : f(x) \neq g(x)\} \in \bar{\mathcal{M}}$ *and* $\bar{\mu}(E) = 0$, *i.e.* $f(x) = g(x)$ *for* $\bar{\mu}$ *– a.e.* x. *Moreover for such a pair* f *and* g, $f \in L^1(\bar{\mu})$ *iff* $g \in L^1(\mu)$ *and in which case*

$$\int_X f d\bar{\mu} = \int_X g d\mu.$$

Proof. Suppose first that such a function g exists so that $\bar{\mu}(E) = 0$. Since g is also $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable, we see from Proposition 19.46 that f is $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable. Conversely if f is $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable, by considering $f_\pm$ we may assume that $f \geq 0$. Choose $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable simple function $\phi_n \geq 0$ such that $\phi_n \uparrow f$ as $n \to \infty$. Writing

$$\phi_n = \sum a_k 1_{A_k}$$

with $A_k \in \bar{\mathcal{M}}$, we may choose $B_k \in \mathcal{M}$ such that $B_k \subset A_k$ and $\bar{\mu}(A_k \setminus B_k) = 0$. Letting

$$\tilde{\phi}_n := \sum a_k 1_{B_k}$$

we have produced a $(\mathcal{M}, \mathcal{B})$ – measurable simple function $\tilde{\phi}_n \geq 0$ such that $E_n := \{\phi_n \neq \tilde{\phi}_n\}$ has zero $\bar{\mu}$ – measure. Since $\bar{\mu}(\cup_n E_n) \leq \sum_n \bar{\mu}(E_n)$, there exists $F \in \mathcal{M}$ such that $\cup_n E_n \subset F$ and $\mu(F) = 0$. It now follows that

$$1_F \tilde{\phi}_n = 1_F \phi_n \uparrow g := 1_F f \text{ as } n \to \infty.$$

This shows that $g = 1_F f$ is $(\mathcal{M}, \mathcal{B})$ – measurable and that $\{f \neq g\} \subset F$ has $\bar{\mu}$ – measure zero. Since $f = g$, $\bar{\mu}$ – a.e., $\int_X f d\bar{\mu} = \int_X g d\bar{\mu}$ so to prove Eq. (19.16) it suffices to prove

$$\int_X g d\bar{\mu} = \int_X g d\mu. \tag{19.16}$$

Because $\bar{\mu} = \mu$ on $\mathcal{M}$, Eq. (19.16) is easily verified for non-negative $\mathcal{M}$ – measurable simple functions. Then by the monotone convergence theorem and the approximation Theorem 18.42 it holds for all $\mathcal{M}$ – measurable functions $g : X \to [0, \infty]$. The rest of the assertions follow in the standard way by considering $(\operatorname{Re} g)_\pm$ and $(\operatorname{Im} g)_\pm$. ∎

19.6 Comparison of the Lebesgue and the Riemann Integral

For the rest of this chapter, let $-\infty < a < b < \infty$ and $f : [a, b] \to \mathbb{R}$ be a bounded function. A partition of $[a, b]$ is a finite subset $\pi \subset [a, b]$ containing $\{a, b\}$. To each partition

$$\pi = \{a = t_0 < t_1 < \cdots < t_n = b\} \tag{19.17}$$

of $[a, b]$ let

$$\mathrm{mesh}(\pi) := \max\{|t_j - t_{j-1}| : j = 1, \ldots, n\},$$

$$M_j = \sup\{f(x) : t_j \le x \le t_{j-1}\}, \;\; m_j = \inf\{f(x) : t_j \le x \le t_{j-1}\}$$

$$G_\pi = f(a)1_{\{a\}} + \sum_1^n M_j 1_{(t_{j-1}, t_j]}, \;\; g_\pi = f(a)1_{\{a\}} + \sum_1^n m_j 1_{(t_{j-1}, t_j]} \text{ and}$$

$$S_\pi f = \sum M_j (t_j - t_{j-1}) \text{ and } s_\pi f = \sum m_j (t_j - t_{j-1}).$$

Notice that

$$S_\pi f = \int_a^b G_\pi dm \text{ and } s_\pi f = \int_a^b g_\pi dm.$$

The upper and lower Riemann integrals are defined respectively by

$$\overline{\int_a^b} f(x)dx = \inf_\pi S_\pi f \text{ and } \underline{\int_b^a} f(x)dx = \sup_\pi \; s_\pi f.$$

Definition 19.48. *The function f is* ***Riemann integrable*** *iff $\overline{\int_a^b} f = \underline{\int_a^b} f \in \mathbb{R}$ and which case the Riemann integral $\int_a^b f$ is defined to be the common value:*

$$\int_a^b f(x)dx = \overline{\int_a^b} f(x)dx = \underline{\int_a^b} f(x)dx.$$

The proof of the following Lemma is left to the reader as Exercise 19.20.

Lemma 19.49. *If π' and π are two partitions of $[a, b]$ and $\pi \subset \pi'$ then*

$$G_\pi \ge G_{\pi'} \ge f \ge g_{\pi'} \ge g_\pi \text{ and}$$
$$S_\pi f \ge S_{\pi'} f \ge s_{\pi'} f \ge s_\pi f.$$

There exists an increasing sequence of partitions $\{\pi_k\}_{k=1}^\infty$ such that $\mathrm{mesh}(\pi_k) \downarrow 0$ *and*

$$S_{\pi_k} f \downarrow \overline{\int_a^b} f \text{ and } s_{\pi_k} f \uparrow \underline{\int_a^b} f \text{ as } k \to \infty.$$

If we let

$$G := \lim_{k\to\infty} G_{\pi_k} \text{ and } g := \lim_{k\to\infty} g_{\pi_k} \tag{19.18}$$

then by the dominated convergence theorem,

$$\int_{[a,b]} g dm = \lim_{k\to\infty} \int_{[a,b]} g_{\pi_k} = \lim_{k\to\infty} s_{\pi_k} f = \underline{\int_a^b} f(x)dx \tag{19.19}$$

and

$$\int_{[a,b]} G dm = \lim_{k\to\infty} \int_{[a,b]} G_{\pi_k} = \lim_{k\to\infty} S_{\pi_k} f = \overline{\int_a^b} f(x)dx. \tag{19.20}$$

Notation 19.50 *For $x \in [a,b]$, let*

$$H(x) = \limsup_{y\to x} f(y) := \lim_{\varepsilon\downarrow 0} \sup\{f(y) : |y-x| \le \varepsilon,\ y \in [a,b]\} \text{ and}$$
$$h(x) = \liminf_{y\to x} f(y) := \lim_{\varepsilon\downarrow 0} \inf\{f(y) : |y-x| \le \varepsilon,\ y \in [a,b]\}.$$

Lemma 19.51. *The functions $H, h : [a,b] \to \mathbb{R}$ satisfy:*

1. *$h(x) \le f(x) \le H(x)$ for all $x \in [a,b]$ and $h(x) = H(x)$ iff f is continuous at x.*
2. *If $\{\pi_k\}_{k=1}^\infty$ is any increasing sequence of partitions such that* $\mathrm{mesh}(\pi_k) \downarrow 0$ *and G and g are defined as in Eq. (19.18), then*
$$G(x) = H(x) \ge f(x) \ge h(x) = g(x) \quad \forall\ x \notin \pi := \cup_{k=1}^\infty \pi_k. \tag{19.21}$$
(Note π is a countable set.)
3. *H and h are Borel measurable.*

Proof. Let $G_k := G_{\pi_k} \downarrow G$ and $g_k := g_{\pi_k} \uparrow g$.

1. It is clear that $h(x) \le f(x) \le H(x)$ for all x and $H(x) = h(x)$ iff $\lim_{y\to x} f(y)$ exists and is equal to $f(x)$. That is $H(x) = h(x)$ iff f is continuous at x.
2. For $x \notin \pi$,
$$G_k(x) \ge H(x) \ge f(x) \ge h(x) \ge g_k(x)\ \forall\ k$$
and letting $k \to \infty$ in this equation implies
$$G(x) \ge H(x) \ge f(x) \ge h(x) \ge g(x)\ \forall\ x \notin \pi. \tag{19.22}$$
Moreover, given $\varepsilon > 0$ and $x \notin \pi$,
$$\sup\{f(y) : |y-x| \le \varepsilon,\ y \in [a,b]\} \ge G_k(x)$$
for all k large enough, since eventually $G_k(x)$ is the supremum of $f(y)$ over some interval contained in $[x-\varepsilon, x+\varepsilon]$. Again letting $k \to \infty$ implies $\sup_{|y-x|\le\varepsilon} f(y) \ge G(x)$ and therefore, that
$$H(x) = \limsup_{y\to x} f(y) \ge G(x)$$
for all $x \notin \pi$. Combining this equation with Eq. (19.22) then implies $H(x) = G(x)$ if $x \notin \pi$. A similar argument shows that $h(x) = g(x)$ if $x \notin \pi$ and hence Eq. (19.21) is proved.
3. The functions G and g are limits of measurable functions and hence measurable. Since $H = G$ and $h = g$ except possibly on the countable set π, both H and h are also Borel measurable. (You justify this statement.)

∎

Theorem 19.52. *Let $f : [a, b] \to \mathbb{R}$ be a bounded function. Then*

$$\overline{\int_a^b} f = \int_{[a,b]} H dm \text{ and } \underline{\int_a^b} f = \int_{[a,b]} h dm \tag{19.23}$$

and the following statements are equivalent:

1. *$H(x) = h(x)$ for m -a.e. x,*
2. *the set*
$$E := \{x \in [a, b] : f \text{ is discontinuous at } x\}$$
is an $\bar{m}$ – null set.
3. *f is Riemann integrable.*

If f is Riemann integrable then f is Lebesgue measurable[2], i.e. f is $\mathcal{L}/\mathcal{B}$ – measurable where $\mathcal{L}$ is the Lebesgue σ – algebra and $\mathcal{B}$ is the Borel σ – algebra on $[a, b]$. Moreover if we let $\bar{m}$ denote the completion of m, then

$$\int_{[a,b]} H dm = \int_a^b f(x)dx = \int_{[a,b]} f d\bar{m} = \int_{[a,b]} h dm. \tag{19.24}$$

Proof. Let $\{\pi_k\}_{k=1}^\infty$ be an increasing sequence of partitions of $[a, b]$ as described in Lemma 19.49 and let G and g be defined as in Lemma 19.51. Since $m(\pi) = 0$, $H = G$ a.e., Eq. (19.23) is a consequence of Eqs. (19.19) and (19.20). From Eq. (19.23), f is Riemann integrable iff

$$\int_{[a,b]} H dm = \int_{[a,b]} h dm$$

and because $h \le f \le H$ this happens iff $h(x) = H(x)$ for m - a.e. x. Since $E = \{x : H(x) \ne h(x)\}$, this last condition is equivalent to E being a m – null set. In light of these results and Eq. (19.21), the remaining assertions including Eq. (19.24) are now consequences of Lemma 19.47. ■

Notation 19.53 *In view of this theorem we will often write $\int_a^b f(x)dx$ for $\int_a^b f dm$.*

19.7 Determining Classes of Measures

Definition 19.54 (σ – finite). *Let X be a set and $\mathcal{E} \subset \mathcal{F} \subset 2^X$. We say that a function $\mu : \mathcal{F} \to [0, \infty]$ is σ – **finite on** $\mathcal{E}$ if there exist $X_n \in \mathcal{E}$ such that $X_n \uparrow X$ and $\mu(X_n) < \infty$ for all n.*

[2] f need not be Borel measurable.

Theorem 19.55 (Uniqueness). *Suppose that $\mathcal{C} \subset 2^X$ is a π – class (see Definition 18.53), $\mathcal{M} = \sigma(\mathcal{C})$ and μ and ν are two measure on $\mathcal{M}$. If μ and ν are σ – finite on $\mathcal{C}$ and $\mu = \nu$ on $\mathcal{C}$, then $\mu = \nu$ on $\mathcal{M}$.*

Proof. We begin first with the special case where $\mu(X) < \infty$ and therefore also

$$\nu(X) = \lim_{n\to\infty} \nu(X_n) = \lim_{n\to\infty} \mu(X_n) = \mu(X) < \infty.$$

Let

$$\mathcal{H} := \{f \in \ell^\infty(\mathcal{M}, \mathbb{R}) : \mu(f) = \nu(f)\}.$$

Then $\mathcal{H}$ is a linear subspace which is closed under bounded convergence (by the dominated convergence theorem), contains 1 and contains the multiplicative system, $M := \{1_C : C \in \mathcal{C}\}$. Therefore, by Theorem 18.51 or Corollary 18.54, $\mathcal{H} = \ell^\infty(\mathcal{M}, \mathbb{R})$ and hence $\mu = \nu$. For the general case, let $X_n^1, X_n^2 \in \mathcal{C}$ be chosen so that $X_n^1 \uparrow X$ and $X_n^2 \uparrow X$ as $n \to \infty$ and $\mu(X_n^1) + \nu(X_n^2) < \infty$ for all n. Then $X_n := X_n^1 \cap X_n^2 \in \mathcal{C}$ increases to X and $\nu(X_n) = \mu(X_n) < \infty$ for all n. For each $n \in \mathbb{N}$, define two measures μ_n and ν_n on $\mathcal{M}$ by

$$\mu_n(A) := \mu(A \cap X_n) \text{ and } \nu_n(A) = \nu(A \cap X_n).$$

Then, as the reader should verify, μ_n and ν_n are finite measure on $\mathcal{M}$ such that $\mu_n = \nu_n$ on $\mathcal{C}$. Therefore, by the special case just proved, $\mu_n = \nu_n$ on $\mathcal{M}$. Finally, using the continuity properties of measures,

$$\mu(A) = \lim_{n\to\infty} \mu(A \cap X_n) = \lim_{n\to\infty} \nu(A \cap X_n) = \nu(A)$$

for all $A \in \mathcal{M}$. ∎

As an immediate consequence we have the following corollaries.

Corollary 19.56. *Suppose that (X, τ) is a topological space, $\mathcal{B}_X = \sigma(\tau)$ is the Borel σ – algebra on X and μ and ν are two measures on $\mathcal{B}_X$ which are σ – finite on τ. If $\mu = \nu$ on τ then $\mu = \nu$ on $\mathcal{B}_X$, i.e. $\mu \equiv \nu$.*

Corollary 19.57. *Suppose that μ and ν are two measures on $\mathcal{B}_{\mathbb{R}^n}$ which are finite on bounded sets and such that $\mu(A) = \nu(A)$ for all sets A of the form*

$$A = (a, b] = (a_1, b_1] \times \cdots \times (a_n, b_n]$$

with $a, b \in \mathbb{R}^n$ and $a < b$, i.e. $a_i < b_i$ for all i. Then $\mu = \nu$ on $\mathcal{B}_{\mathbb{R}^n}$.

Proposition 19.58. *Suppose that (X, d) is a metric space, μ and ν are two measures on $\mathcal{B}_X := \sigma(\tau_d)$ which are finite on bounded measurable subsets of X and*

$$\int_X f d\mu = \int_X f d\nu \tag{19.25}$$

for all $f \in BC_b(X, \mathbb{R})$ where

$$BC_b(X, \mathbb{R}) = \{f \in BC(X, \mathbb{R}) : \operatorname{supp}(f) \text{ is bounded}\}. \tag{19.26}$$

Then $\mu \equiv \nu$.

Proof. To prove this fix a $o \in X$ and let

$$\psi_R(x) = ([R+1-d(x,o)] \wedge 1) \vee 0 \tag{19.27}$$

so that $\psi_R \in BC_b(X,[0,1])$, $\mathrm{supp}(\psi_R) \subset B(o,R+2)$ and $\psi_R \uparrow 1$ as $R \to \infty$. Let $\mathcal{H}_R$ denote the space of bounded real valued $\mathcal{B}_X$ – measurable functions f such that

$$\int_X \psi_R f d\mu = \int_X \psi_R f d\nu. \tag{19.28}$$

Then $\mathcal{H}_R$ is closed under bounded convergence and because of Eq. (19.25) contains $BC(X,\mathbb{R})$. Therefore by Corollary 18.55, $\mathcal{H}_R$ contains all bounded measurable functions on X. Take $f = 1_A$ in Eq. (19.28) with $A \in \mathcal{B}_X$, and then use the monotone convergence theorem to let $R \to \infty$. The result is $\mu(A) = \nu(A)$ for all $A \in \mathcal{B}_X$. ∎

Here is another version of Proposition 19.58.

Proposition 19.59. *Suppose that (X,d) is a metric space, μ and ν are two measures on $\mathcal{B}_X = \sigma(\tau_d)$ which are both finite on compact sets. Further assume there exists compact sets $K_k \subset X$ such that $K_k^o \uparrow X$. If*

$$\int_X f d\mu = \int_X f d\nu \tag{19.29}$$

for all $f \in C_c(X,\mathbb{R})$ then $\mu \equiv \nu$.

Proof. Let $\psi_{n,k}$ be defined as in the proof of Proposition 18.56 and let $\mathcal{H}_{n,k}$ denote those bounded $\mathcal{B}_X$ – measurable functions, $f : X \to \mathbb{R}$ such that

$$\int_X f \psi_{n,k} d\mu = \int_X f \psi_{n,k} d\nu.$$

By assumption $BC(X,\mathbb{R}) \subset \mathcal{H}_{n,k}$ and one easily checks that $\mathcal{H}_{n,k}$ is closed under bounded convergence. Therefore, by Corollary 18.55, $\mathcal{H}_{n,k}$ contains all bounded measurable function. In particular for $A \in \mathcal{B}_X$,

$$\int_X 1_A \psi_{n,k} d\mu = \int_X 1_A \psi_{n,k} d\nu.$$

Letting $n \to \infty$ in this equation, using the dominated convergence theorem, one shows

$$\int_X 1_A 1_{K_k^o} d\mu = \int_X 1_A 1_{K_k^o} d\nu$$

holds for k. Finally using the monotone convergence theorem we may let $k \to \infty$ to conclude

$$\mu(A) = \int_X 1_A d\mu = \int_X 1_A d\nu = \nu(A)$$

for all $A \in \mathcal{B}_X$. ∎

19.8 Exercises

Exercise 19.2. Let μ be a measure on an algebra $\mathcal{A} \subset 2^X$, then $\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B)$ for all $A, B \in \mathcal{A}$.

Exercise 19.3 (From problem 12 on p. 27 of Folland.). Let $(X, \mathcal{M}, \mu)$ be a finite measure space and for $A, B \in \mathcal{M}$ let $\rho(A, B) = \mu(A \Delta B)$ where $A \Delta B = (A \setminus B) \cup (B \setminus A)$. It is clear that $\rho(A, B) = \rho(B, A)$. Show:

1. ρ satisfies the triangle inequality:
$$\rho(A, C) \leq \rho(A, B) + \rho(B, C) \text{ for all } A, B, C \in \mathcal{M}.$$
2. Define $A \sim B$ iff $\mu(A \Delta B) = 0$ and notice that $\rho(A, B) = 0$ iff $A \sim B$. Show "$\sim$ " is an equivalence relation.
3. Let $\mathcal{M}/\sim$ denote $\mathcal{M}$ modulo the equivalence relation, $\sim$, and let $[A] := \{B \in \mathcal{M} : B \sim A\}$. Show that $\bar{\rho}([A], [B]) := \rho(A, B)$ is gives a well defined metric on $\mathcal{M}/\sim$.
4. Similarly show $\tilde{\mu}([A]) = \mu(A)$ is a well defined function on $\mathcal{M}/\sim$ and show $\tilde{\mu} : (\mathcal{M}/\sim) \to \mathbb{R}_+$ is $\bar{\rho}$ – continuous.

Exercise 19.4. Suppose that $\mu_n : \mathcal{M} \to [0, \infty]$ are measures on $\mathcal{M}$ for $n \in \mathbb{N}$. Also suppose that $\mu_n(A)$ is increasing in n for all $A \in \mathcal{M}$. Prove that $\mu : \mathcal{M} \to [0, \infty]$ defined by $\mu(A) := \lim_{n\to\infty} \mu_n(A)$ is also a measure.

Exercise 19.5. Now suppose that Λ is some index set and for each $\lambda \in \Lambda$, $\mu_\lambda : \mathcal{M} \to [0, \infty]$ is a measure on $\mathcal{M}$. Define $\mu : \mathcal{M} \to [0, \infty]$ by $\mu(A) = \sum_{\lambda\in\Lambda} \mu_\lambda(A)$ for each $A \in \mathcal{M}$. Show that μ is also a measure.

Exercise 19.6. Let $(X, \mathcal{M}, \mu)$ be a measure space and $\rho : X \to [0, \infty]$ be a measurable function. For $A \in \mathcal{M}$, set $\nu(A) := \int_A \rho d\mu$.

1. Show $\nu : \mathcal{M} \to [0, \infty]$ is a measure.
2. Let $f : X \to [0, \infty]$ be a measurable function, show
$$\int_X f d\nu = \int_X f \rho d\mu. \tag{19.30}$$
Hint: first prove the relationship for characteristic functions, then for simple functions, and then for general positive measurable functions.
3. Show that a measurable function $f : X \to \mathbb{C}$ is in $L^1(\nu)$ iff $|f| \rho \in L^1(\mu)$ and if $f \in L^1(\nu)$ then Eq. (19.30) still holds.

Notation 19.60 *It is customary to informally describe ν defined in Exercise 19.6 by writing $d\nu = \rho d\mu$.*

Exercise 19.7. Let $(X, \mathcal{M}, \mu)$ be a measure space, $(Y, \mathcal{F})$ be a measurable space and $f : X \to Y$ be a measurable map. Define a function $\nu : \mathcal{F} \to [0, \infty]$ by $\nu(A) := \mu(f^{-1}(A))$ for all $A \in \mathcal{F}$.

1. Show ν is a measure. (We will write $\nu = f_*\mu$ or $\nu = \mu \circ f^{-1}$.)
2. Show
$$\int_Y g d\nu = \int_X (g \circ f)\, d\mu \tag{19.31}$$
for all measurable functions $g : Y \to [0, \infty]$. **Hint:** see the hint from Exercise 19.6.
3. Show a measurable function $g : Y \to \mathbb{C}$ is in $L^1(\nu)$ iff $g \circ f \in L^1(\mu)$ and that Eq. (19.31) holds for all $g \in \mathrm{L}^1(\nu)$.

Exercise 19.8. Let $F : \mathbb{R} \to \mathbb{R}$ be a C^1-function such that $F'(x) > 0$ for all $x \in \mathbb{R}$ and $\lim_{x\to\pm\infty} F(x) = \pm\infty$. (Notice that F is strictly increasing so that $F^{-1} : \mathbb{R} \to \mathbb{R}$ exists and moreover, by the inverse function theorem that F^{-1} is a C^1 – function.) Let m be Lebesgue measure on $\mathcal{B}_\mathbb{R}$ and
$$\nu(A) = m(F(A)) = m(\left(F^{-1}\right)^{-1}(A)) = \left(F_*^{-1} m\right)(A)$$
for all $A \in \mathcal{B}_\mathbb{R}$. Show $d\nu = F' dm$. Use this result to prove the change of variable formula,
$$\int_\mathbb{R} h \circ F \cdot F' dm = \int_\mathbb{R} h dm \tag{19.32}$$
which is valid for all Borel measurable functions $h : \mathbb{R} \to [0, \infty]$.

Hint: Start by showing $d\nu = F' dm$ on sets of the form $A = (a, b]$ with $a, b \in \mathbb{R}$ and $a < b$. Then use the uniqueness assertions in Theorem 19.8 (or see Corollary 19.57) to conclude $d\nu = F' dm$ on all of $\mathcal{B}_\mathbb{R}$. To prove Eq. (19.32) apply Exercise 19.7 with $g = h \circ F$ and $f = F^{-1}$.

Exercise 19.9. Let $(X, \mathcal{M}, \mu)$ be a measure space and $\{A_n\}_{n=1}^\infty \subset \mathcal{M}$, show
$$\mu(\{A_n \text{ a.a.}\}) \le \liminf_{n\to\infty} \mu(A_n)$$
and if $\mu(\cup_{m\ge n} A_m) < \infty$ for some n, then
$$\mu(\{A_n \text{ i.o.}\}) \ge \limsup_{n\to\infty} \mu(A_n).$$

Exercise 19.10. BRUCE: Delete this exercise which is contained in Lemma 19.17. Suppose $(X, \mathcal{M}, \mu)$ be a measure space and $f : X \to [0\, \infty]$ be a measurable function such that $\int_X f d\mu < \infty$. Show $\mu(\{f = \infty\}) = 0$ and the set $\{f > 0\}$ is σ – finite.

Exercise 19.11. Folland 2.13 on p. 52. **Hint:** "Fatou times two."

Exercise 19.12. Folland 2.14 on p. 52. BRUCE: delete this exercise

Exercise 19.13. Give examples of measurable functions $\{f_n\}$ on $\mathbb{R}$ such that f_n decreases to 0 uniformly yet $\int f_n dm = \infty$ for all n. Also give an example of a sequence of measurable functions $\{g_n\}$ on $[0, 1]$ such that $g_n \to 0$ while $\int g_n dm = 1$ for all n.

Exercise 19.14. Folland 2.19 on p. 59. (This problem is essentially covered in the previous exercise.)

Exercise 19.15. Suppose $\{a_n\}_{n=-\infty}^{\infty} \subset \mathbb{C}$ is a summable sequence (i.e. $\sum_{n=-\infty}^{\infty} |a_n| < \infty$), then $f(\theta) := \sum_{n=-\infty}^{\infty} a_n e^{in\theta}$ is a continuous function for $\theta \in \mathbb{R}$ and

$$a_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in\theta}d\theta.$$

Exercise 19.16. For any function $f \in L^1(m)$, show $x \in \mathbb{R} \to \int_{(-\infty,x]} f(t)\,dm(t)$ is continuous in x. Also find a finite measure, μ, on $\mathcal{B}_{\mathbb{R}}$ such that $x \to \int_{(-\infty,x]} f(t)\,d\mu(t)$ is not continuous.

Exercise 19.17. Folland 2.28 on p. 60.

Exercise 19.18. Folland 2.31b and 2.31e on p. 60. (The answer in 2.13b is wrong by a factor of -1 and the sum is on $k = 1$ to ∞. In part e, s should be taken to be a. You may also freely use the Taylor series expansion

$$(1-z)^{-1/2} = \sum_{n=0}^{\infty} \frac{(2n-1)!!}{2^n n!} z^n = \sum_{n=0}^{\infty} \frac{(2n)!}{4^n (n!)^2} z^n \text{ for } |z| < 1.$$

Exercise 19.19. There exists a meager (see Definition 13.4 and Proposition 13.3) subsets of $\mathbb{R}$ which have full Lebesgue measure, i.e. whose complement is a Lebesgue null set. (This is Folland 5.27. **Hint:** Consider the generalized Cantor sets discussed on p. 39 of Folland.)

Exercise 19.20. Prove Lemma 19.49.

20

Multiple Integrals

In this chapter we will introduce iterated integrals and product measures. We are particularly interested in when it is permissible to interchange the order of integration in multiple integrals.

Example 20.1. As an example let $X = [1, \infty)$ and $Y = [0, 1]$ equipped with their Borel σ - algebras and let $\mu = \nu = m$, where m is Lebesgue measure. The iterated integrals of the function $f(x, y) := e^{-xy} - 2e^{-2xy}$ satisfy,

$$\int_0^1 \left[\int_1^\infty (e^{-xy} - 2e^{-2xy}) dx \right] dy = \int_0^1 e^{-y} \left(\frac{1 - e^{-y}}{y} \right) dy \in (0, \infty)$$

and

$$\int_1^\infty \left[\int_0^1 (e^{-xy} - 2e^{-2xy}) dy \right] dx = -\int_1^\infty e^{-x} \left[\frac{1 - e^{-x}}{x} \right] dx \in (-\infty, 0)$$

and therefore are not equal. Hence it is not always true that order of integration is irrelevant.

Lemma 20.2. *Let $\mathbb{F}$ be either $[0, \infty)$, $\mathbb{R}$ or $\mathbb{C}$. Suppose $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ are two measurable spaces and $f : X \times Y \to \mathbb{F}$ is a $(\mathcal{M} \otimes \mathcal{N}, \mathcal{B}_{\mathbb{F}})$ – measurable function, then for each $y \in Y$,*

$$x \to f(x, y) \text{ is } (\mathcal{M}, \mathcal{B}_{\mathbb{F}}) \text{ measurable,} \tag{20.1}$$

for each $x \in X$,

$$y \to f(x, y) \text{ is } (\mathcal{N}, \mathcal{B}_{\mathbb{F}}) \text{ measurable.} \tag{20.2}$$

Proof. Suppose that $E = A \times B \in \mathcal{E} := \mathcal{M} \times \mathcal{N}$ and $f = 1_E$. Then

$$f(x, y) = 1_{A \times B}(x, y) = 1_A(x) 1_B(y)$$

from which it follows that Eqs. (20.1) and (20.2) for this function. Let $\mathcal{H}$ be the collection of all bounded $(\mathcal{M} \otimes \mathcal{N}, \mathcal{B}_{\mathbb{F}})$ – measurable functions on $X \times Y$ such that Eqs. (20.1) and (20.2) hold, here we assume $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Because measurable functions are closed under taking linear combinations and pointwise limits, $\mathcal{H}$ is linear subspace of $\ell^\infty(\mathcal{M}\otimes\mathcal{N},\mathbb{F})$ which is closed under bounded convergence and contain $1_E \in \mathcal{H}$ for all E in the π – class, $\mathcal{E}$. Therefore by by Corollary 18.54, that $\mathcal{H} = \ell^\infty(\mathcal{M}\otimes\mathcal{N},\mathbb{F})$.

For the general $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_\mathbb{R})$ – measurable functions $f : X\times Y \to \mathbb{F}$ and $M \in \mathbb{N}$, let $f_M := 1_{|f|\le M} f \in \ell^\infty(\mathcal{M}\otimes\mathcal{N},\mathbb{F})$. Then Eqs. (20.1) and (20.2) hold with f replaced by f_M and hence for f as well by letting $M\to\infty$. ■

Notation 20.3 (Iterated Integrals) *If $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are two measure spaces and $f : X\times Y \to \mathbb{C}$ is a $\mathcal{M}\otimes\mathcal{N}$ – measurable function, the* ***iterated integrals*** *of f (when they make sense) are:*

$$\int_X d\mu(x)\int_Y d\nu(y) f(x,y) := \int_X \left[\int_Y f(x,y)d\nu(y)\right] d\mu(x)$$

and

$$\int_Y d\nu(y)\int_X d\mu(x) f(x,y) := \int_Y \left[\int_X f(x,y)d\mu(x)\right] d\nu(y).$$

Notation 20.4 *Suppose that $f : X \to \mathbb{C}$ and $g : Y \to \mathbb{C}$ are functions, let $f\otimes g$ denote the function on $X\times Y$ given by*

$$f\otimes g(x,y) = f(x)g(y).$$

Notice that if f, g are measurable, then $f\otimes g$ is $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_\mathbb{C})$ – measurable. To prove this let $F(x,y) = f(x)$ and $G(x,y) = g(y)$ so that $f\otimes g = F\cdot G$ will be measurable provided that F and G are measurable. Now $F = f\circ\pi_1$ where $\pi_1 : X\times Y \to X$ is the projection map. This shows that F is the composition of measurable functions and hence measurable. Similarly one shows that G is measurable.

20.1 Fubini-Tonelli's Theorem and Product Measure

Theorem 20.5. *Suppose $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are σ-finite measure spaces and f is a nonnegative $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_\mathbb{R})$ – measurable function, then for each $y\in Y$,*

$$x \to f(x,y) \text{ is } \mathcal{M} – \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{20.3}$$

for each $x\in X$,

$$y \to f(x,y) \text{ is } \mathcal{N} – \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{20.4}$$

$$x \to \int_Y f(x,y)d\nu(y) \text{ is } \mathcal{M} – \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{20.5}$$

$$y \to \int_X f(x,y)d\mu(x) \text{ is } \mathcal{N} – \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{20.6}$$

and

$$\int_X d\mu(x) \int_Y d\nu(y) f(x,y) = \int_Y d\nu(y) \int_X d\mu(x) f(x,y). \tag{20.7}$$

Proof. Suppose that $E = A \times B \in \mathcal{E} := \mathcal{M} \times \mathcal{N}$ and $f = 1_E$. Then

$$f(x,y) = 1_{A\times B}(x,y) = 1_A(x) 1_B(y)$$

and one sees that Eqs. (20.3) and (20.4) hold. Moreover

$$\int_Y f(x,y) d\nu(y) = \int_Y 1_A(x) 1_B(y) d\nu(y) = 1_A(x)\nu(B),$$

so that Eq. (20.5) holds and we have

$$\int_X d\mu(x) \int_Y d\nu(y) f(x,y) = \nu(B)\mu(A). \tag{20.8}$$

Similarly,

$$\int_X f(x,y) d\mu(x) = \mu(A) 1_B(y) \text{ and}$$
$$\int_Y d\nu(y) \int_X d\mu(x) f(x,y) = \nu(B)\mu(A)$$

from which it follows that Eqs. (20.6) and (20.7) hold in this case as well. For the moment let us further assume that $\mu(X) < \infty$ and $\nu(Y) < \infty$ and let $\mathcal{H}$ be the collection of all bounded $(\mathcal{M} \otimes \mathcal{N}, \mathcal{B}_{\mathbb{R}})$ – measurable functions on $X \times Y$ such that Eqs. (20.3) – (20.7) hold. Using the fact that measurable functions are closed under pointwise limits and the dominated convergence theorem (the dominating function always being a constant), one easily shows that $\mathcal{H}$ closed under bounded convergence. Since we have just verified that $1_E \in \mathcal{H}$ for all E in the π – class, $\mathcal{E}$, it follows by Corollary 18.54 that $\mathcal{H}$ is the space of all bounded $(\mathcal{M} \otimes \mathcal{N}, \mathcal{B}_{\mathbb{R}})$ – measurable functions on $X \times Y$. Finally if $f : X \times Y \to [0, \infty]$ is a $(\mathcal{M} \otimes \mathcal{N}, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, let $f_M = M \wedge f$ so that $f_M \uparrow f$ as $M \to \infty$ and Eqs. (20.3) – (20.7) hold with f replaced by f_M for all $M \in \mathbb{N}$. Repeated use of the monotone convergence theorem allows us to pass to the limit $M \to \infty$ in these equations to deduce the theorem in the case μ and ν are finite measures. For the σ – finite case, choose $X_n \in \mathcal{M}$, $Y_n \in \mathcal{N}$ such that $X_n \uparrow X$, $Y_n \uparrow Y$, $\mu(X_n) < \infty$ and $\nu(Y_n) < \infty$ for all $m, n \in \mathbb{N}$. Then define $\mu_m(A) = \mu(X_m \cap A)$ and $\nu_n(B) = \nu(Y_n \cap B)$ for all $A \in \mathcal{M}$ and $B \in \mathcal{N}$ or equivalently $d\mu_m = 1_{X_m} d\mu$ and $d\nu_n = 1_{Y_n} d\nu$. By what we have just proved Eqs. (20.3) – (20.7) with μ replaced by μ_m and ν by ν_n for all $(\mathcal{M} \otimes \mathcal{N}, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable functions, $f : X \times Y \to [0, \infty]$. The validity of Eqs. (20.3) – (20.7) then follows by passing to the limits $m \to \infty$ and then $n \to \infty$ making use of the monotone convergence theorem in the form,

$$\int_X u d\mu_m = \int_X u 1_{X_m} d\mu \uparrow \int_X u d\mu \text{ as } m \to \infty$$

and

$$\int_Y v d\mu_n = \int_Y v 1_{Y_n} d\mu \uparrow \int_Y v d\mu \text{ as } n \to \infty$$

for all $u \in L^+(X, \mathcal{M})$ and $v \in L^+(Y, \mathcal{N})$. ■

Corollary 20.6. *Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ – finite measure spaces. Then there exists a unique measure π on $\mathcal{M} \otimes \mathcal{N}$ such that $\pi(A \times B) = \mu(A)\nu(B)$ for all $A \in \mathcal{M}$ and $B \in \mathcal{N}$. Moreover π is given by*

$$\pi(E) = \int_X d\mu(x) \int_Y d\nu(y) 1_E(x, y) = \int_Y d\nu(y) \int_X d\mu(x) 1_E(x, y) \tag{20.9}$$

for all $E \in \mathcal{M} \otimes \mathcal{N}$ and π is σ – finite.

Proof. Notice that any measure π such that $\pi(A \times B) = \mu(A)\nu(B)$ for all $A \in \mathcal{M}$ and $B \in \mathcal{N}$ is necessarily σ – finite. Indeed, let $X_n \in \mathcal{M}$ and $Y_n \in \mathcal{N}$ be chosen so that $\mu(X_n) < \infty$, $\nu(Y_n) < \infty$, $X_n \uparrow X$ and $Y_n \uparrow Y$, then $X_n \times Y_n \in \mathcal{M} \otimes \mathcal{N}$, $X_n \times Y_n \uparrow X \times Y$ and $\pi(X_n \times Y_n) < \infty$ for all n. The uniqueness assertion is a consequence of Theorem 19.55 or see Theorem **??** below with $\mathcal{E} = \mathcal{M} \times \mathcal{N}$. For the existence, it suffices to observe, using the monotone convergence theorem, that π defined in Eq. (20.9) is a measure on $\mathcal{M} \otimes \mathcal{N}$. Moreover this measure satisfies $\pi(A \times B) = \mu(A)\nu(B)$ for all $A \in \mathcal{M}$ and $B \in \mathcal{N}$ from Eq. (20.8 ■

Notation 20.7 *The measure π is called the product measure of μ and ν and will be denoted by $\mu \otimes \nu$.*

Theorem 20.8 (Tonelli's Theorem). *Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ – finite measure spaces and $\pi = \mu \otimes \nu$ is the product measure on $\mathcal{M} \otimes \mathcal{N}$. If $f \in L^+(X \times Y, \mathcal{M} \otimes \mathcal{N})$, then $f(\cdot, y) \in L^+(X, \mathcal{M})$ for all $y \in Y$, $f(x, \cdot) \in L^+(Y, \mathcal{N})$ for all $x \in X$,*

$$\int_Y f(\cdot, y) d\nu(y) \in L^+(X, \mathcal{M}), \int_X f(x, \cdot) d\mu(x) \in L^+(Y, \mathcal{N})$$

and

$$\int_{X \times Y} f \, d\pi = \int_X d\mu(x) \int_Y d\nu(y) f(x, y) \tag{20.10}$$

$$= \int_Y d\nu(y) \int_X d\mu(x) f(x, y). \tag{20.11}$$

Proof. By Theorem 20.5 and Corollary 20.6, the theorem holds when $f = 1_E$ with $E \in \mathcal{M} \otimes \mathcal{N}$. Using the linearity of all of the statements, the theorem is also true for non-negative simple functions. Then using the monotone convergence theorem repeatedly along with the approximation Theorem 18.42, one deduces the theorem for general $f \in L^+(X \times Y, \mathcal{M} \otimes \mathcal{N})$. ■

The following convention will be in force for the rest of this chapter.

Convention: If $(X, \mathcal{M}, \mu)$ is a measure space and $f : X \to \mathbb{C}$ is a measurable but non-integrable function, i.e. $\int_X |f| \, d\mu = \infty$, by convention we will define $\int_X f d\mu := 0$. However if f is a non-negative function (i.e. $f : X \to [0, \infty]$) is a non-integrable function we will still write $\int_X f d\mu = \infty$.

Theorem 20.9 (Fubini's Theorem). *Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ – finite measure spaces, $\pi = \mu \otimes \nu$ is the product measure on $\mathcal{M} \otimes \mathcal{N}$ and $f : X \times Y \to \mathbb{C}$ is a $\mathcal{M} \otimes \mathcal{N}$ – measurable function. Then the following three conditions are equivalent:*

$$\int_{X \times Y} |f| \, d\pi < \infty, \text{ i.e. } f \in L^1(\pi), \tag{20.12}$$

$$\int_X \left(\int_Y |f(x, y)| \, d\nu(y) \right) d\mu(x) < \infty \text{ and} \tag{20.13}$$

$$\int_Y \left(\int_X |f(x, y)| \, d\mu(x) \right) d\nu(y) < \infty. \tag{20.14}$$

If any one (and hence all) of these condition hold, then $f(x, \cdot) \in L^1(\nu)$ for μ a.e. x, $f(\cdot, y) \in L^1(\mu)$ for ν a.e. y, $\int_Y f(\cdot, y) dv(y) \in L^1(\mu)$, $\int_X f(x, \cdot) d\mu(x) \in L^1(\nu)$ and Eqs. (20.10) and (20.11) are still valid.

Proof. The equivalence of Eqs. (20.12) – (20.14) is a direct consequence of Tonelli's Theorem 20.8. Now suppose $f \in L^1(\pi)$ is a real valued function and let

$$E := \left\{ x \in X : \int_Y |f(x, y)| \, d\nu(y) = \infty \right\}. \tag{20.15}$$

Then by Tonelli's theorem, $x \to \int_Y |f(x, y)| \, d\nu(y)$ is measurable and hence $E \in \mathcal{M}$. Moreover Tonelli's theorem implies

$$\int_X \left[\int_Y |f(x, y)| \, d\nu(y) \right] d\mu(x) = \int_{X \times Y} |f| \, d\pi < \infty$$

which implies that $\mu(E) = 0$. Let $f_\pm$ be the positive and negative parts of f, then using the above convention we have

$$\begin{aligned} \int_Y f(x, y) \, d\nu(y) &= \int_Y 1_E(x) f(x, y) \, d\nu(y) \\ &= \int_Y 1_E(x) [f_+(x, y) - f_-(x, y)] \, d\nu(y) \\ &= \int_Y 1_E(x) f_+(x, y) \, d\nu(y) - \int_Y 1_E(x) f_-(x, y) \, d\nu(y). \end{aligned} \tag{20.16}$$

Noting that $1_E(x) f_\pm(x, y) = (1_E \otimes 1_Y \cdot f_\pm)(x, y)$ is a positive $\mathcal{M} \otimes \mathcal{N}$ – measurable function, it follows from another application of Tonelli's theorem that $x \to \int_Y f(x,y)\, d\nu(y)$ is $\mathcal{M}$ – measurable, being the difference of two measurable functions. Moreover

$$\int_X \left| \int_Y f(x,y)\, d\nu(y) \right| d\mu(x) \le \int_X \left[\int_Y |f(x,y)|\, d\nu(y) \right] d\mu(x) < \infty,$$

which shows $\int_Y f(\cdot,y) dv(y) \in L^1(\mu)$. Integrating Eq. (20.16) on x and using Tonelli's theorem repeatedly implies,

$$\begin{aligned}
&\int_X \left[\int_Y f(x,y)\, d\nu(y) \right] d\mu(x) \\
&\quad = \int_X d\mu(x) \int_Y d\nu(y)\, 1_E(x)\, f_+(x,y) - \int_X d\mu(x) \int_Y d\nu(y)\, 1_E(x)\, f_-(x,y) \\
&\quad = \int_Y d\nu(y) \int_X d\mu(x)\, 1_E(x)\, f_+(x,y) - \int_Y d\nu(y) \int_X d\mu(x)\, 1_E(x)\, f_-(x,y) \\
&\quad = \int_Y d\nu(y) \int_X d\mu(x)\, f_+(x,y) - \int_Y d\nu(y) \int_X d\mu(x)\, f_-(x,y) \\
&\quad = \int_{X\times Y} f_+ d\pi - \int_{X\times Y} f_- d\pi = \int_{X\times Y} (f_+ - f_-)\, d\pi = \int_{X\times Y} f d\pi \quad (20.17)
\end{aligned}$$

which proves Eq. (20.10) holds.

Now suppose that $f = u + iv$ is complex valued and again let E be as in Eq. (20.15). Just as above we still have $E \in \mathcal{M}$ and $\mu(E) < \infty$. By our convention,

$$\begin{aligned}
\int_Y f(x,y)\, d\nu(y) &= \int_Y 1_E(x)\, f(x,y)\, d\nu(y) = \int_Y 1_E(x) \left[u(x,y) + iv(x,y) \right] d\nu(y) \\
&= \int_Y 1_E(x)\, u(x,y)\, d\nu(y) + i \int_Y 1_E(x)\, v(x,y)\, d\nu(y)
\end{aligned}$$

which is measurable in x by what we have just proved. Similarly one shows $\int_Y f(\cdot,y)\, d\nu(y) \in L^1(\mu)$ and Eq. (20.10) still holds by a computation similar to that done in Eq. (20.17). The assertions pertaining to Eq. (20.11) may be proved in the same way. ■

Notation 20.10 *Given $E \subset X \times Y$ and $x \in X$, let*

$$_xE := \{y \in Y : (x,y) \in E\}.$$

Similarly if $y \in Y$ is given let

$$E_y := \{x \in X : (x,y) \in E\}.$$

If $f : X \times Y \to \mathbb{C}$ is a function let $f_x = f(x,\cdot)$ and $f^y := f(\cdot,y)$ so that $f_x : Y \to \mathbb{C}$ and $f^y : X \to \mathbb{C}$.

Theorem 20.11. *Suppose $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are complete σ – finite measure spaces. Let $(X\times Y,\mathcal{L},\lambda)$ be the completion of $(X\times Y,\mathcal{M}\otimes\mathcal{N},\mu\otimes\nu)$. If f is $\mathcal{L}$ – measurable and (a) $f\geq 0$ or (b) $f\in L^1(\lambda)$ then f_x is $\mathcal{N}$ – measurable for μ a.e. x and f^y is $\mathcal{M}$ – measurable for ν a.e. y and in case (b) $f_x\in L^1(\nu)$ and $f^y\in L^1(\mu)$ for μ a.e. x and ν a.e. y respectively. Moreover,*

$$\left(x\to\int_Y f_x d\nu\right)\in L^1(\mu)\text{ and }\left(y\to\int_X f^y d\mu\right)\in L^1(\nu)$$

and

$$\int_{X\times Y} f d\lambda=\int_Y d\nu\int_X d\mu\, f=\int_X d\mu\int_Y d\nu\, f.$$

Proof. If $E\in\mathcal{M}\otimes\mathcal{N}$ is a $\mu\otimes\nu$ null set (i.e. $(\mu\otimes\nu)(E)=0$), then

$$0=(\mu\otimes\nu)(E)=\int\limits_X\nu({}_xE)d\mu(x)=\int\limits_X\mu(E_y)d\nu(y).$$

This shows that

$$\mu(\{x:\nu({}_xE)\neq 0\})=0\text{ and }\nu(\{y:\mu(E_y)\neq 0\})=0,$$

i.e. $\nu({}_xE)=0$ for μ a.e. x and $\mu(E_y)=0$ for ν a.e. y. If h is $\mathcal{L}$ measurable and $h=0$ for λ – a.e., then there exists $E\in\mathcal{M}\otimes\mathcal{N}$ such that $\{(x,y):h(x,y)\neq 0\}\subset E$ and $(\mu\otimes\nu)(E)=0$. Therefore $|h(x,y)|\leq 1_E(x,y)$ and $(\mu\otimes\nu)(E)=0$. Since

$$\begin{aligned}\{h_x\neq 0\}&=\{y\in Y:h(x,y)\neq 0\}\subset {}_xE\text{ and}\\ \{h_y\neq 0\}&=\{x\in X:h(x,y)\neq 0\}\subset E_y\end{aligned}$$

we learn that for μ a.e. x and ν a.e. y that $\{h_x\neq 0\}\in\mathcal{M}$, $\{h_y\neq 0\}\in\mathcal{N}$, $\nu(\{h_x\neq 0\})=0$ and a.e. and $\mu(\{h_y\neq 0\})=0$. This implies $\int_Y h(x,y)d\nu(y)$ exists and equals 0 for μ a.e. x and similarly that $\int_X h(x,y)d\mu(x)$ exists and equals 0 for ν a.e. y. Therefore

$$0=\int_{X\times Y}hd\lambda=\int_Y\left(\int_X hd\mu\right)d\nu=\int_X\left(\int_Y hd\nu\right)d\mu.$$

For general $f\in L^1(\lambda)$, we may choose $g\in L^1(\mathcal{M}\otimes\mathcal{N},\mu\otimes\nu)$ such that $f(x,y)=g(x,y)$ for $\lambda-$ a.e. (x,y). Define $h:=f-g$. Then $h=0$, $\lambda-$ a.e. Hence by what we have just proved and Theorem 20.8 $f=g+h$ has the following properties:

1. For μ a.e. x, $y\to f(x,y)=g(x,y)+h(x,y)$ is in $L^1(\nu)$ and

$$\int_Y f(x,y)d\nu(y)=\int_Y g(x,y)d\nu(y).$$

2. For ν a.e. y, $x \to f(x,y) = g(x,y) + h(x,y)$ is in $L^1(\mu)$ and

$$\int_X f(x,y)d\mu(x) = \int_X g(x,y)d\mu(x).$$

From these assertions and Theorem 20.8, it follows that

$$\begin{aligned}\int_X d\mu(x)\int_Y d\nu(y)f(x,y) &= \int_X d\mu(x)\int_Y d\nu(y)g(x,y)\\ &= \int_Y d\nu(y)\int_Y d\nu(x)g(x,y)\\ &= \int_{X\times Y} g(x,y)d(\mu\otimes\nu)(x,y)\\ &= \int_{X\times Y} f(x,y)d\lambda(x,y).\end{aligned}$$

Similarly it is shown that

$$\int_Y d\nu(y)\int_X d\mu(x)f(x,y) = \int_{X\times Y} f(x,y)d\lambda(x,y).$$

■

The previous theorems have obvious generalizations to products of any finite number of σ – finite measure spaces. For example the following theorem holds.

Theorem 20.12. *Suppose $\{(X_i, \mathcal{M}_i, \mu_i)\}_{i=1}^n$ are σ – finite measure spaces and $X := X_1 \times \cdots \times X_n$. Then there exists a unique measure, π, on $(X, \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n)$ such that $\pi(A_1 \times \cdots \times A_n) = \mu_1(A_1)\ldots\mu_n(A_n)$ for all $A_i \in \mathcal{M}_i$. (This measure and its completion will be denote by $\mu_1 \otimes \cdots \otimes \mu_n$.) If $f : X \to [0,\infty]$ is a $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ – measurable function then*

$$\int_X f d\pi = \int_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)})\ldots\int_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)})\ f(x_1,\ldots,x_n) \quad (20.18)$$

where σ is any permutation of $\{1,2,\ldots,n\}$. This equation also holds for any $f \in L^1(\pi)$ and moreover, $f \in L^1(\pi)$ iff

$$\int_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)})\ldots\int_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)})\ |f(x_1,\ldots,x_n)| < \infty$$

for some (and hence all) permutations, σ.

This theorem can be proved by the same methods as in the two factor case, see Exercise 20.5. Alternatively, one can use the theorems already proved and induction on n, see Exercise 20.6 in this regard.

Example 20.13. In this example we will show

$$\lim_{M\to\infty}\int_0^M \frac{\sin x}{x}dx = \pi/2. \tag{20.19}$$

To see this write $\frac{1}{x} = \int_0^\infty e^{-tx}dt$ and use Fubini-Tonelli to conclude that

$$\begin{aligned}\int_0^M \frac{\sin x}{x}dx &= \int_0^M \left[\int_0^\infty e^{-tx}\sin x\ dt\right]dx \\ &= \int_0^\infty \left[\int_0^M e^{-tx}\sin x\ dx\right]dt \\ &= \int_0^\infty \frac{1}{1+t^2}\left(1 - te^{-Mt}\sin M - e^{-Mt}\cos M\right)dt \\ &\to \int_0^\infty \frac{1}{1+t^2}dt = \frac{\pi}{2} \text{ as } M\to\infty,\end{aligned}$$

wherein we have used the dominated convergence theorem to pass to the limit.

The next example is a refinement of this result.

Example 20.14. We have

$$\int_0^\infty \frac{\sin x}{x}e^{-\Lambda x}dx = \frac{1}{2}\pi - \arctan\Lambda \text{ for all } \Lambda > 0 \tag{20.20}$$

and for$\Lambda, M\in[0,\infty)$,

$$\left|\int_0^M \frac{\sin x}{x}e^{-\Lambda x}dx - \frac{1}{2}\pi + \arctan\Lambda\right| \le C\frac{e^{-M\Lambda}}{M} \tag{20.21}$$

where $C = \max_{x\ge 0}\frac{1+x}{1+x^2} = \frac{1}{2\sqrt{2}-2} \cong 1.2$. In particular Eq. (20.19) is valid.

To verify these assertions, first notice that by the fundamental theorem of calculus,

$$|\sin x| = \left|\int_0^x \cos y dy\right| \le \left|\int_0^x |\cos y|\, dy\right| \le \left|\int_0^x 1dy\right| = |x|$$

so $\left|\frac{\sin x}{x}\right| \le 1$ for all $x\ne 0$. Making use of the identity

$$\int_0^\infty e^{-tx}dt = 1/x$$

and Fubini's theorem,

$$\begin{aligned}
\int_0^M \frac{\sin x}{x} e^{-\Lambda x} dx &= \int_0^M dx \sin x \, e^{-\Lambda x} \int_0^\infty e^{-tx} dt \\
&= \int_0^\infty dt \int_0^M dx \sin x \, e^{-(\Lambda+t)x} \\
&= \int_0^\infty \frac{1 - (\cos M + (\Lambda+t)\sin M)\, e^{-M(\Lambda+t)}}{(\Lambda+t)^2+1} dt \\
&= \int_0^\infty \frac{1}{(\Lambda+t)^2+1} dt - \int_0^\infty \frac{\cos M + (\Lambda+t)\sin M}{(\Lambda+t)^2+1} e^{-M(\Lambda+t)} dt \\
&= \frac{1}{2}\pi - \arctan \Lambda - \varepsilon(M,\Lambda) \qquad (20.22)
\end{aligned}$$

where

$$\varepsilon(M,\Lambda) = \int_0^\infty \frac{\cos M + (\Lambda+t)\sin M}{(\Lambda+t)^2+1} e^{-M(\Lambda+t)} dt.$$

Since

$$\left| \frac{\cos M + (\Lambda+t)\sin M}{(\Lambda+t)^2+1} \right| \le \frac{1+(\Lambda+t)}{(\Lambda+t)^2+1} \le C,$$

$$|\varepsilon(M,\Lambda)| \le \int_0^\infty e^{-M(\Lambda+t)} dt = C\frac{e^{-M\Lambda}}{M}.$$

This estimate along with Eq. (20.22) proves Eq. (20.21) from which Eq. (20.19) follows by taking $\Lambda \to \infty$ and Eq. (20.20) follows (using the dominated convergence theorem again) by letting $M \to \infty$.

20.2 Lebesgue Measure on $\mathbb{R}^d$ and the Change of Variables Theorem

Notation 20.15 *Let*

$$m^d := \overbrace{m \otimes \cdots \otimes m}^{d \text{ times}} \text{ on } \mathcal{B}_{\mathbb{R}^d} = \overbrace{\mathcal{B}_{\mathbb{R}} \otimes \cdots \otimes \mathcal{B}_{\mathbb{R}}}^{d \text{ times}}$$

be the d – fold product of Lebesgue measure m on $\mathcal{B}_{\mathbb{R}}$. We will also use m^d to denote its completion and let $\mathcal{L}_d$ be the completion of $\mathcal{B}_{\mathbb{R}^d}$ relative to m^d. A subset $A \in \mathcal{L}_d$ is called a Lebesgue measurable set and m^d is called d – dimensional Lebesgue measure, or just Lebesgue measure for short.

Definition 20.16. *A function $f : \mathbb{R}^d \to \mathbb{R}$ is **Lebesgue measurable** if $f^{-1}(\mathcal{B}_{\mathbb{R}}) \subset \mathcal{L}_d$.*

Notation 20.17 *I will often be sloppy in the sequel and write m for m^d and dx for $dm(x) = dm^d(x)$, i.e.*

$$\int_{\mathbb{R}^d} f(x)\,dx = \int_{\mathbb{R}^d} f dm = \int_{\mathbb{R}^d} f dm^d.$$

Hopefully the reader will understand the meaning from the context.

Theorem 20.18. *Lebesgue measure m^d is translation invariant. Moreover m^d is the unique translation invariant measure on $\mathcal{B}_{\mathbb{R}^d}$ such that $m^d((0,1]^d) = 1$.*

Proof. Let $A = J_1 \times \cdots \times J_d$ with $J_i \in \mathcal{B}_{\mathbb{R}}$ and $x \in \mathbb{R}^d$. Then

$$x + A = (x_1 + J_1) \times (x_2 + J_2) \times \cdots \times (x_d + J_d)$$

and therefore by translation invariance of m on $\mathcal{B}_{\mathbb{R}}$ we find that

$$m^d(x + A) = m(x_1 + J_1)\dots m(x_d + J_d) = m(J_1)\dots m(J_d) = m^d(A)$$

and hence $m^d(x + A) = m^d(A)$ for all $A \in \mathcal{B}_{\mathbb{R}^d}$ by Corollary 19.57. From this fact we see that the measure $m^d(x + \cdot)$ and $m^d(\cdot)$ have the same null sets. Using this it is easily seen that $m(x + A) = m(A)$ for all $A \in \mathcal{L}_d$. The proof of the second assertion is Exercise 20.7. ■

Exercise 20.1. In this problem you are asked to show there is no reasonable notion of Lebesgue measure on an infinite dimensional Hilbert space. To be more precise, suppose H is an infinite dimensional Hilbert space and m is a **countably additive** measure on $\mathcal{B}_H$ which is invariant under translations and satisfies, $m(B_0(\varepsilon)) > 0$ for all $\varepsilon > 0$. Show $m(V) = \infty$ for all non-empty open subsets $V \subset H$.

Theorem 20.19 (Change of Variables Theorem). *Let $\Omega \subset_o \mathbb{R}^d$ be an open set and $T : \Omega \to T(\Omega) \subset_o \mathbb{R}^d$ be a C^1 – diffeomorphism,*[1] *see Figure 20.1. Then for any Borel measurable function, $f : T(\Omega) \to [0, \infty]$,*

$$\int_\Omega f(T(x))\,|\det T'(x)|\,dx = \int_{T(\Omega)} f(y)\,dy, \tag{20.23}$$

where $T'(x)$ is the linear transformation on $\mathbb{R}^d$ defined by $T'(x)v := \frac{d}{dt}|_0 T(x + tv)$. More explicitly, viewing vectors in $\mathbb{R}^d$ as columns, $T'(x)$ may be represented by the matrix

$$T'(x) = \begin{bmatrix} \partial_1 T_1(x) & \dots & \partial_d T_1(x) \\ \vdots & \ddots & \vdots \\ \partial_1 T_d(x) & \dots & \partial_d T_d(x) \end{bmatrix}, \tag{20.24}$$

i.e. the i - j – matrix entry of $T'(x)$ is given by $T'(x)_{ij} = \partial_i T_j(x)$ where $T(x) = (T_1(x), \dots, T_d(x))^{\mathrm{tr}}$ and $\partial_i = \partial/\partial x_i$.

[1] That is $T : \Omega \to T(\Omega) \subset_o \mathbb{R}^d$ is a continuously differentiable bijection and the inverse map $T^{-1} : T(\Omega) \to \Omega$ is also continuously differentiable.

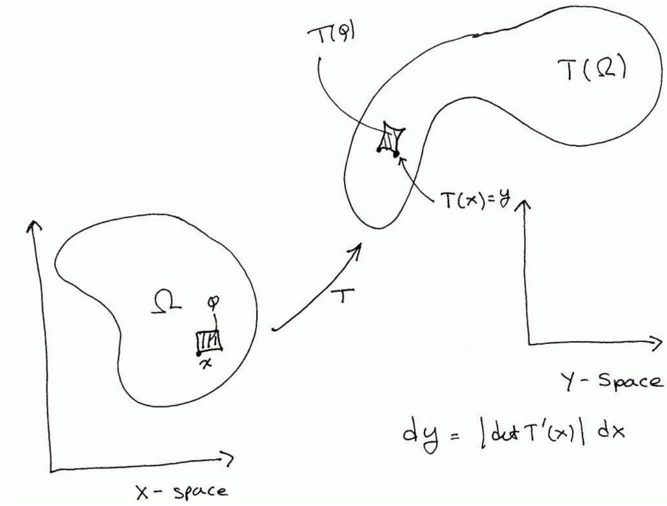


Fig. 20.1. The geometric setup of Theorem 20.19.

Remark 20.20. Theorem 20.19 is best remembered as the statement: if we make the change of variables $y = T(x)$, then $dy = |\det T'(x)| dx$. As usual, you must also change the limits of integration appropriately, i.e. if x ranges through Ω then y must range through $T(\Omega)$.

Proof. The proof will be by induction on d. The case $d = 1$ was essentially done in Exercise 19.8. Nevertheless, for the sake of completeness let us give a proof here. Suppose $d = 1$, $a < \alpha < \beta < b$ such that $[a, b]$ is a compact subinterval of Ω. Then $|\det T'| = |T'|$ and

$$\int_{[a,b]} 1_{T((\alpha,\beta])}(T(x)) |T'(x)| dx = \int_{[a,b]} 1_{(\alpha,\beta]}(x) |T'(x)| dx = \int_\alpha^\beta |T'(x)| dx.$$

If $T'(x) > 0$ on $[a, b]$, then

$$\begin{aligned}\int_\alpha^\beta |T'(x)| dx &= \int_\alpha^\beta T'(x) dx = T(\beta) - T(\alpha) \\ &= m(T((\alpha,\beta])) = \int_{T([a,b])} 1_{T((\alpha,\beta])}(y) dy\end{aligned}$$

while if $T'(x) < 0$ on $[a, b]$, then

$$\begin{aligned}\int_\alpha^\beta |T'(x)| dx &= -\int_\alpha^\beta T'(x) dx = T(\alpha) - T(\beta) \\ &= m(T((\alpha,\beta])) = \int_{T([a,b])} 1_{T((\alpha,\beta])}(y) dy.\end{aligned}$$

Combining the previous three equations shows

$$\int_{[a,b]} f(T(x))\,|T'(x)|\,dx = \int_{T([a,b])} f(y)\,dy \tag{20.25}$$

whenever f is of the form $f = 1_{T((\alpha,\beta])}$ with $a < \alpha < \beta < b$. An application of Dynkin's multiplicative system Theorem 18.51 then implies that Eq. (20.25) holds for every bounded measurable function $f : T([a,b]) \to \mathbb{R}$. (Observe that $|T'(x)|$ is continuous and hence bounded for x in the compact interval, $[a,b]$. From Exercise 10.14, $\Omega = \coprod_{n=1}^{N} (a_n, b_n)$ where $a_n, b_n \in \mathbb{R} \cup \{\pm\infty\}$ for $n = 1, 2, \dots < N$ with $N = \infty$ possible. Hence if $f : T(\Omega) \to \mathbb{R}_+$ is a Borel measurable function and $a_n < \alpha_k < \beta_k < b_n$ with $\alpha_k \downarrow a_n$ and $\beta_k \uparrow b_n$, then by what we have already proved and the monotone convergence theorem

$$\begin{aligned}
\int_\Omega 1_{(a_n,b_n)} \cdot (f \circ T) \cdot |T'| dm &= \int_\Omega \left(1_{T((a_n,b_n))} \cdot f\right) \circ T \cdot |T'| dm \\
&= \lim_{k\to\infty} \int_\Omega \left(1_{T([\alpha_k,\beta_k])} \cdot f\right) \circ T \cdot |T'|\; dm \\
&= \lim_{k\to\infty} \int_{T(\Omega)} 1_{T([\alpha_k,\beta_k])} \cdot f\; dm \\
&= \int_{T(\Omega)} 1_{T((a_n,b_n))} \cdot f\; dm.
\end{aligned}$$

Summing this equality on n, then shows Eq. (20.23) holds.

To carry out the induction step, we now suppose $d > 1$ and suppose the theorem is valid with d being replaced by $d-1$. For notational compactness, let us write vectors in $\mathbb{R}^d$ as row vectors rather than column vectors. Nevertheless, the matrix associated to the differential, $T'(x)$, will always be taken to be given as in Eq. (20.24).

Case 1. Suppose $T(x)$ has the form

$$T(x) = (x_i, T_2(x), \dots, T_d(x)) \tag{20.26}$$

or

$$T(x) = (T_1(x), \dots, T_{d-1}(x), x_i) \tag{20.27}$$

for some $i \in \{1, \dots, d\}$. For definiteness we will assume T is as in Eq. (20.26), the case of T in Eq. (20.27) may be handled similarly. For $t \in \mathbb{R}$, let $i_t : \mathbb{R}^{d-1} \to \mathbb{R}^d$ be the inclusion map defined by

$$i_t(w) := w_t := (w_1, \dots, w_{i-1}, t, w_{i+1}, \dots, w_{d-1}),$$

Ω_t be the (possibly empty) open subset of $\mathbb{R}^{d-1}$ defined by

$$\Omega_t := \left\{ w \in \mathbb{R}^{d-1} : (w_1, \dots, w_{i-1}, t, w_{i+1}, \dots, w_{d-1}) \in \Omega \right\}$$

and $T_t : \Omega_t \to \mathbb{R}^{d-1}$ be defined by

$$T_t(w) = (T_2(w_t), \ldots, T_d(w_t)),$$

see Figure 20.2. Expanding $\det T'(w_t)$ along the first row of the matrix $T'(w_t)$

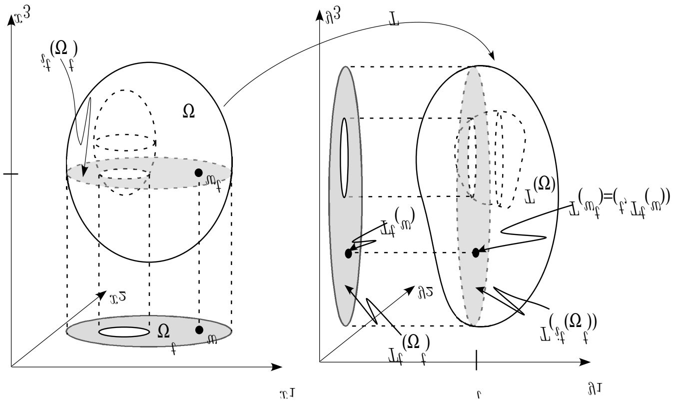


Fig. 20.2. In this picture $d = i = 3$ and Ω is an egg-shaped region with an egg-shaped hole. The picture indicates the geometry associated with the map T and slicing the set Ω along planes where $x_3 = t$.

shows

$$|\det T'(w_t)| = |\det T_t'(w)|.$$

Now by the Fubini-Tonelli Theorem and the induction hypothesis,

$$\begin{aligned}\int_\Omega f\circ T|\det T'|dm &= \int_{\mathbb{R}^d} 1_\Omega\cdot f\circ T|\det T'|dm \\ &= \int_{\mathbb{R}^d} 1_\Omega(w_t)(f\circ T)(w_t)\,|\det T'(w_t)|dwdt \\ &= \int_{\mathbb{R}}\left[\int_{\Omega_t}(f\circ T)(w_t)\,|\det T'(w_t)|dw\right]dt \\ &= \int_{\mathbb{R}}\left[\int_{\Omega_t} f(t,T_t(w))\,|\det T_t'(w)|dw\right]dt \\ &= \int_{\mathbb{R}}\left[\int_{T_t(\Omega_t)} f(t,z)\,dz\right]dt = \int_{\mathbb{R}}\left[\int_{\mathbb{R}^{d-1}} 1_{T(\Omega)}(t,z)\,f(t,z)\,dz\right]dt \\ &= \int_{T(\Omega)} f(y)\,dy\end{aligned}$$

wherein the last two equalities we have used Fubini-Tonelli along with the identity;

$$T(\Omega) = \coprod_{t\in\mathbb{R}} T(i_t(\Omega)) = \coprod_{t\in\mathbb{R}} \{(t,z) : z\in T_t(\Omega_t)\}.$$

Case 2. (Eq. (20.23) is true locally.) Suppose that $T:\Omega\to\mathbb{R}^d$ is a general map as in the statement of the theorem and $x_0\in\Omega$ is an arbitrary point. We will now show there exists an open neighborhood $W\subset\Omega$ of x_0 such that

$$\int_W f\circ T|\det T'|dm = \int_{T(W)} f dm$$

holds for all Borel measurable function, $f:T(W)\to[0,\infty]$. Let M_i be the 1-i minor of $T'(x_0)$, i.e. the determinant of $T'(x_0)$ with the first row and i^{th} – column removed. Since

$$0\neq\det T'(x_0) = \sum_{i=1}^d (-1)^{i+1}\,\partial_i T_j(x_0)\cdot M_i,$$

there must be some i such that $M_i\neq 0$. Fix an i such that $M_i\neq 0$ and let,

$$S(x) := (x_i, T_2(x), \dots, T_d(x)). \tag{20.28}$$

Observe that $|\det S'(x_0)| = |M_i| \neq 0$. Hence by the inverse function Theorem 16.25, there exist an open neighborhood W of x_0 such that $W\subset_o\Omega$ and $S(W)\subset_o\mathbb{R}^d$ and $S:W\to S(W)$ is a C^1 – diffeomorphism. Let $R:S(W)\to T(W)\subset_o\mathbb{R}^d$ to be the C^1 – diffeomorphism defined by

$$R(z) := T \circ S^{-1}(z) \text{ for all } z \in S(W).$$

Because

$$(T_1(x), \dots, T_d(x)) = T(x) = R(S(x)) = R((x_i, T_2(x), \dots, T_d(x)))$$

for all $x \in W$, if

$$(z_1, z_2, \dots, z_d) = S(x) = (x_i, T_2(x), \dots, T_d(x))$$

then

$$R(z) = \left(T_1\left(S^{-1}(z)\right), z_2, \dots, z_d\right). \tag{20.29}$$

Observe that S is a map of the form in Eq. (20.26), R is a map of the form in Eq. (20.27), $T'(x) = R'(S(x))S'(x)$ (by the chain rule) and (by the multiplicative property of the determinant)

$$|\det T'(x)| = |\det R'(S(x))|\,|\det S'(x)| \ \forall\, x \in W.$$

So if $f : T(W) \to [0,\infty]$ is a Borel measurable function, two applications of the results in Case 1. shows,

$$\begin{aligned}\int_W f \circ T \cdot |\det T'| dm &= \int_W (f \circ R \cdot |\det R'|) \circ S \cdot |\det S'| \ dm \\ &= \int_{S(W)} f \circ R \cdot |\det R'| dm = \int_{R(S(W))} f dm \\ &= \int_{T(W)} f dm\end{aligned}$$

and Case 2. is proved.

Case 3. (General Case.) Let $f : \Omega \to [0,\infty]$ be a general non-negative Borel measurable function and let

$$K_n := \{x \in \Omega : \mathrm{dist}(x, \Omega^c) \ge 1/n \text{ and } |x| \le n\}.$$

Then each K_n is a compact subset of Ω and $K_n \uparrow \Omega$ as $n \to \infty$. Using the compactness of K_n and case 2, for each $n \in \mathbb{N}$, there is a finite open cover $\mathcal{W}_n$ of K_n such that $W \subset \Omega$ and Eq. (20.23) holds with Ω replaced by W for each $W \in \mathcal{W}_n$. Let $\{W_i\}_{i=1}^\infty$ be an enumeration of $\cup_{n=1}^\infty \mathcal{W}_n$ and set $\tilde{W}_1 = W_1$ and $\tilde{W}_i := W_i \setminus (W_1 \cup \dots \cup W_{i-1})$ for all $i \ge 2$. Then $\Omega = \coprod_{i=1}^\infty \tilde{W}_i$ and by repeated use of case 2.,

$$\begin{aligned}\int_\Omega f\circ T|\det T'|dm &= \sum_{i=1}^\infty \int_\Omega 1_{\tilde{W}_i}\cdot (f\circ T)\cdot |\det T'|dm \\ &= \sum_{i=1}^\infty \int_{W_i} \left[\left(1_{T(\tilde{W}_i)}f\right)\circ T\right]\cdot |\det T'|dm \\ &= \sum_{i=1}^\infty \int_{T(W_i)} 1_{T(\tilde{W}_i)}\cdot f\, dm = \sum_{i=1}^n \int_{T(\Omega)} 1_{T(\tilde{W}_i)}\cdot f\, dm \\ &= \int_{T(\Omega)} f dm.\end{aligned}$$

∎

Remark 20.21. When $d=1$, one often learns the change of variables formula as

$$\int_a^b f(T(x))\,T'(x)\,dx = \int_{T(a)}^{T(b)} f(y)\,dy \tag{20.30}$$

where $f:[a,b]\to\mathbb{R}$ is a continuous function and T is C^1 – function defined in a neighborhood of $[a,b]$. If $T'>0$ on (a,b) then $T((a,b)) = (T(a),T(b))$ and Eq. (20.30) is implies Eq. (20.23) with $\Omega=(a,b)$. On the other hand if $T'<0$ on (a,b) then $T((a,b))=(T(b),T(a))$ and Eq. (20.30) is equivalent to

$$\int_{(a,b)} f(T(x))\,(-|T'(x)|)\,dx = -\int_{T(b)}^{T(a)} f(y)\,dy = -\int_{T((a,b))} f(y)\,dy$$

which is again implies Eq. (20.23). On the other hand Eq. Eq. (20.30) is more general than Eq. (20.23) since it does not require T to be injective. The standard proof of Eq. (20.30) is as follows. For $z\in T([a,b])$, let

$$F(z) := \int_{T(a)}^z f(y)\,dy.$$

Then by the chain rule and the fundamental theorem of calculus,

$$\begin{aligned}\int_a^b f(T(x))\,T'(x)\,dx &= \int_a^b F'(T(x))\,T'(x)\,dx = \int_a^b \frac{d}{dx}[F(T(x))]\,dx \\ &= F(T(x))\,|_a^b = \int_{T(a)}^{T(b)} f(y)\,dy.\end{aligned}$$

An application of Dynkin's multiplicative systems theorem (in the form of Corollary 18.55) now shows that Eq. (20.30) holds for all bounded measurable functions f on (a,b). Then by the usual truncation argument, it also holds for all positive measurable functions on (a,b).

Example 20.22. Continuing the setup in Theorem 20.19, if $A \in \mathcal{B}_\Omega$, then

$$\begin{aligned} m\left(T\left(A\right)\right) &= \int_{\mathbb{R}^d} 1_{T(A)}\left(y\right) dy = \int_{\mathbb{R}^d} 1_{T(A)}\left(Tx\right) \left|\det T'\left(x\right)\right| dx \\ &= \int_{\mathbb{R}^d} 1_A\left(x\right) \left|\det T'\left(x\right)\right| dx \end{aligned}$$

wherein the second equality we have made the change of variables, $y = T\left(x\right).$ Hence we have shown

$$d\left(m \circ T\right) = \left|\det T'\left(\cdot\right)\right| dm.$$

In particular if $T \in GL(d, \mathbb{R}) = GL(\mathbb{R}^d)$ – the space of $d \times d$ invertible matrices, then $m \circ T = \left|\det T\right| m$, i.e.

$$m\left(T\left(A\right)\right) = \left|\det T\right| m\left(A\right) \text{ for all} A \in \mathcal{B}_{\mathbb{R}^d}. \tag{20.31}$$

This equation also shows that $m \circ T$ and m have the same null sets and hence the equality in Eq. (20.31) is valid for any $A \in \mathcal{L}_d$.

Exercise 20.2. Show that $f \in L^1\left(T\left(\Omega\right), m^d\right)$ iff

$$\int_\Omega |f \circ T|\,|\det T'| dm < \infty$$

and if $f \in L^1\left(T\left(\Omega\right), m^d\right)$, then Eq. (20.23) holds.

Example 20.23 (Polar Coordinates). Suppose $T : (0, \infty) \times (0, 2\pi) \to \mathbb{R}^2$ is defined by

$$x = T(r, \theta) = \left(r \cos\theta, r \sin\theta\right),$$

i.e. we are making the change of variable,

$$x_1 = r\cos\theta \text{ and } x_2 = r\sin\theta \text{ for } 0 < r < \infty \text{ and } 0 < \theta < 2\pi.$$

In this case

$$T'(r, \theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

and therefore

$$dx = \left|\det T'(r, \theta)\right| dr d\theta = r dr d\theta.$$

Observing that

$$\mathbb{R}^2 \setminus T\left((0, \infty) \times (0, 2\pi)\right) = \ell := \{(x, 0) : x \geq 0\}$$

has m^2 – measure zero, it follows from the change of variables Theorem 20.19 that

$$\int_{\mathbb{R}^2} f(x) dx = \int_0^{2\pi} d\theta \int_0^\infty dr \; r \cdot f(r\left(\cos\theta, \sin\theta\right)) \tag{20.32}$$

for any Borel measurable function $f : \mathbb{R}^2 \to [0, \infty]$.

Example 20.24 (Holomorphic Change of Variables). Suppose that $f : \Omega \subset_o \mathbb{C} \cong \mathbb{R}^2 \rightarrow \mathbb{C}$ is an injective holomorphic function such that $f'(z) \neq 0$ for all $z \in \Omega$. We may express f as

$$f(x + iy) = U(x, y) + iV(x, y)$$

for all $z = x + iy \in \Omega$. Hence if we make the change of variables,

$$w = u + iv = f(x + iy) = U(x, y) + iV(x, y)$$

then

$$dudv = \left| \det \begin{bmatrix} U_x & U_y \\ V_x & V_y \end{bmatrix} \right| dxdy = \left| U_x V_y - U_y V_x \right| dxdy.$$

Recalling that U and V satisfy the Cauchy Riemann equations, $U_x = V_y$ and $U_y = -V_x$ with $f' = U_x + iV_x$, we learn

$$U_x V_y - U_y V_x = U_x^2 + V_x^2 = \left| f' \right|^2.$$

Therefore

$$dudv = \left| f'(x + iy) \right|^2 dxdy.$$

Example 20.25. In this example we will evaluate the integral

$$I := \iint_\Omega \left(x^4 - y^4 \right) dxdy$$

where

$$\Omega = \left\{ (x, y) : 1 < x^2 - y^2 < 2,\ 0 < xy < 1 \right\},$$

see Figure 20.3 We are going to do this by making the change of variables,

Fig. 20.3. The region Ω consists of the two curved rectangular regions shown.

$$(u, v) := T(x, y) = \left(x^2 - y^2, xy\right),$$

in which case

$$dudv = \left|\det \begin{bmatrix} 2x & -2y \\ y & x \end{bmatrix}\right| dxdy = 2\left(x^2 + y^2\right) dxdy$$

Notice that

$$\left(x^4 - y^4\right) = \left(x^2 - y^2\right)\left(x^2 + y^2\right) = u\left(x^2 + y^2\right) = \frac{1}{2} ududv.$$

The function T is not injective on Ω but it is injective on each of its connected components. Let D be the connected component in the first quadrant so that $\Omega = -D \cup D$ and $T(\pm D) = (1, 2) \times (0, 1)$ The change of variables theorem then implies

$$I_\pm := \iint_{\pm D} \left(x^4 - y^4\right) dxdy = \frac{1}{2} \iint_{(1,2)\times(0,1)} ududv = \frac{1}{2} \frac{u^2}{2}\Big|_1^2 \cdot 1 = \frac{3}{4}$$

and therefore $I = I_+ + I_- = 2 \cdot (3/4) = 3/2$.

Exercise 20.3 (Spherical Coordinates). Let $T : (0, \infty) \times (0, \pi) \times (0, 2\pi) \to \mathbb{R}^3$ be defined by

$$\begin{aligned} T(r, \phi, \theta) &= (r \sin\phi \cos\theta, r \sin\phi \sin\theta, r \cos\phi) \\ &= r(\sin\phi \cos\theta, \sin\phi \sin\theta, \cos\phi), \end{aligned}$$

see Figure 20.4. By making the change of variables $x = T(r, \phi, \theta)$, show

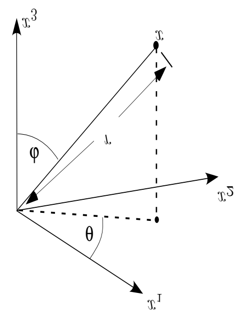


Fig. 20.4. The relation of x to (r, ϕ, θ) in spherical coordinates.

$$\int_{\mathbb{R}^3} f(x)dx = \int_0^\pi d\phi \int_0^{2\pi} d\theta \int_0^\infty dr\ r^2 \sin\phi \cdot f(T(r, \phi, \theta))$$

for any Borel measurable function, $f : \mathbb{R}^3 \to [0, \infty]$.

Lemma 20.26. *Let $a > 0$ and*

$$I_d(a) := \int_{\mathbb{R}^d} e^{-a|x|^2} dm(x).$$

Then $I_d(a) = (\pi/a)^{d/2}$.

Proof. By Tonelli's theorem and induction,

$$\begin{aligned} I_d(a) &= \int_{\mathbb{R}^{d-1}\times\mathbb{R}} e^{-a|y|^2} e^{-at^2} m_{d-1}(dy)\ dt \\ &= I_{d-1}(a) I_1(a) = I_1^d(a). \end{aligned} \tag{20.33}$$

So it suffices to compute:

$$I_2(a) = \int_{\mathbb{R}^2} e^{-a|x|^2} dm(x) = \int_{\mathbb{R}^2\setminus\{0\}} e^{-a(x_1^2+x_2^2)} dx_1 dx_2.$$

Using polar coordinates, see Eq. (20.32), we find,

$$\begin{aligned} I_2(a) &= \int_0^\infty dr\ r \int_0^{2\pi} d\theta\ e^{-ar^2} = 2\pi \int_0^\infty r e^{-ar^2} dr \\ &= 2\pi \lim_{M\to\infty} \int_0^M r e^{-ar^2} dr = 2\pi \lim_{M\to\infty} \frac{e^{-ar^2}}{-2a} \int_0^M = \frac{2\pi}{2a} = \pi/a. \end{aligned}$$

This shows that $I_2(a) = \pi/a$ and the result now follows from Eq. (20.33). ■

20.3 The Polar Decomposition of Lebesgue Measure

Let

$$S^{d-1} = \{x \in \mathbb{R}^d : |x|^2 := \sum_{i=1}^d x_i^2 = 1\}$$

be the unit sphere in $\mathbb{R}^d$ equipped with its Borel σ − algebra, $\mathcal{B}_{S^{d-1}}$ and $\Phi : \mathbb{R}^d \setminus \{0\} \to (0,\infty) \times S^{d-1}$ be defined by $\Phi(x) := (|x|, |x|^{-1} x)$. The inverse map, $\Phi^{-1} : (0,\infty) \times S^{d-1} \to \mathbb{R}^d \setminus \{0\}$, is given by $\Phi^{-1}(r,\omega) = r\omega$. Since Φ and Φ^{-1} are continuous, they are both Borel measurable. For $E \in \mathcal{B}_{S^{d-1}}$ and $a > 0$, let

$$E_a := \{r\omega : r \in (0,a] \text{ and } \omega \in E\} = \Phi^{-1}((0,a] \times E) \in \mathcal{B}_{\mathbb{R}^d}.$$

Definition 20.27. *For $E \in \mathcal{B}_{S^{d-1}}$, let $\sigma(E) := d \cdot m(E_1)$. We call σ the surface measure on S^{d-1}.*

It is easy to check that σ is a measure. Indeed if $E \in \mathcal{B}_{S^{d-1}}$, then $E_1 = \Phi^{-1}((0,1] \times E) \in \mathcal{B}_{\mathbb{R}^d}$ so that $m(E_1)$ is well defined. Moreover if $E = \coprod_{i=1}^{\infty} E_i$, then $E_1 = \coprod_{i=1}^{\infty} (E_i)_1$ and

$$\sigma(E) = d \cdot m(E_1) = \sum_{i=1}^{\infty} m((E_i)_1) = \sum_{i=1}^{\infty} \sigma(E_i).$$

The intuition behind this definition is as follows. If $E \subset S^{d-1}$ is a set and $\varepsilon > 0$ is a small number, then the volume of

$$(1, 1+\varepsilon] \cdot E = \{r\omega : r \in (1, 1+\varepsilon] \text{ and } \omega \in E\}$$

should be approximately given by $m((1, 1+\varepsilon] \cdot E) \cong \sigma(E)\varepsilon$, see Figure 20.5 below. On the other hand

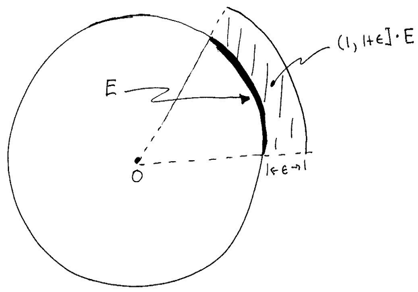


Fig. 20.5. Motivating the definition of surface measure for a sphere.

$$m((1, 1+\varepsilon]E) = m(E_{1+\varepsilon} \setminus E_1) = \left\{(1+\varepsilon)^d - 1\right\} m(E_1).$$

Therefore we expect the area of E should be given by

$$\sigma(E) = \lim_{\varepsilon \downarrow 0} \frac{\left\{(1+\varepsilon)^d - 1\right\} m(E_1)}{\varepsilon} = d \cdot m(E_1).$$

The following theorem is motivated by Example 20.23 and Exercise 20.3.

Theorem 20.28 (Polar Coordinates). *If $f : \mathbb{R}^d \to [0, \infty]$ is a $(\mathcal{B}_{R^d}, \mathcal{B})$–measurable function then*

$$\int_{\mathbb{R}^d} f(x) dm(x) = \int_{(0,\infty) \times S^{d-1}} f(r\omega) r^{d-1} \, dr d\sigma(\omega). \tag{20.34}$$

Proof. By Exercise 19.7,

$$\int_{\mathbb{R}^d} f dm = \int_{\mathbb{R}^d \setminus \{0\}} \left(f \circ \Phi^{-1}\right) \circ \Phi \; dm = \int_{(0,\infty)\times S^{d-1}} \left(f \circ \Phi^{-1}\right) \; d\left(\Phi_* m\right) \quad (20.35)$$

and therefore to prove Eq. (20.34) we must work out the measure $\Phi_* m$ on $\mathcal{B}_{(0,\infty)} \otimes \mathcal{B}_{S^{d-1}}$ defined by

$$\Phi_* m(A) := m\left(\Phi^{-1}(A)\right) \; \forall \; A \in \mathcal{B}_{(0,\infty)} \otimes \mathcal{B}_{S^{d-1}}. \quad (20.36)$$

If $A = (a, b] \times E$ with $0 < a < b$ and $E \in \mathcal{B}_{S^{d-1}}$, then

$$\Phi^{-1}(A) = \{r\omega : r \in (a, b] \text{ and } \omega \in E\} = bE_1 \setminus aE_1$$

wherein we have used $E_a = aE_1$ in the last equality. Therefore by the basic scaling properties of m and the fundamental theorem of calculus,

$$\begin{aligned}(\Phi_* m)\left((a, b] \times E\right) &= m\left(bE_1 \setminus aE_1\right) = m(bE_1) - m(aE_1) \\ &= b^d m(E_1) - a^d m(E_1) = d \cdot m(E_1) \int_a^b r^{d-1} dr. \quad (20.37)\end{aligned}$$

Letting $d\rho(r) = r^{d-1}dr$, i.e.

$$\rho(J) = \int_J r^{d-1} dr \; \forall \; J \in \mathcal{B}_{(0,\infty)}, \quad (20.38)$$

Eq. (20.37) may be written as

$$(\Phi_* m)\left((a, b] \times E\right) = \rho((a, b]) \cdot \sigma(E) = (\rho \otimes \sigma)\left((a, b] \times E\right). \quad (20.39)$$

Since

$$\mathcal{E} = \left\{(a, b] \times E : 0 < a < b \text{ and } E \in \mathcal{B}_{S^{d-1}}\right\},$$

is a π class (in fact it is an elementary class) such that $\sigma(\mathcal{E}) = \mathcal{B}_{(0,\infty)} \otimes \mathcal{B}_{S^{d-1}}$, it follows from Theorem 19.55 and Eq. (20.39) that $\Phi_* m = \rho \otimes \sigma$. Using this result in Eq. (20.35) gives

$$\int_{\mathbb{R}^d} f dm = \int_{(0,\infty)\times S^{d-1}} \left(f \circ \Phi^{-1}\right) \; d\left(\rho \otimes \sigma\right)$$

which combined with Tonelli's Theorem 20.8 proves Eq. (20.35). ■

Corollary 20.29. *The surface area $\sigma(S^{d-1})$ of the unit sphere $S^{d-1} \subset \mathbb{R}^d$ is*

$$\sigma(S^{d-1}) = \frac{2\pi^{d/2}}{\Gamma(d/2)} \quad (20.40)$$

where Γ is the gamma function given by

$$\Gamma(x) := \int_0^\infty u^{x-1} e^{-u} dr \quad (20.41)$$

Moreover, $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$ and $\Gamma(x + 1) = x\Gamma(x)$ for $x > 0$.

Proof. Using Theorem 20.28 we find

$$I_d(1) = \int_0^\infty dr\; r^{d-1} e^{-r^2} \int_{S^{d-1}} d\sigma = \sigma(S^{d-1}) \int_0^\infty r^{d-1} e^{-r^2} dr.$$

We simplify this last integral by making the change of variables $u = r^2$ so that $r = u^{1/2}$ and $dr = \frac{1}{2} u^{-1/2} du$. The result is

$$\begin{aligned}\int_0^\infty r^{d-1} e^{-r^2} dr &= \int_0^\infty u^{\frac{d-1}{2}} e^{-u} \frac{1}{2} u^{-1/2} du \\ &= \frac{1}{2} \int_0^\infty u^{\frac{d}{2}-1} e^{-u} du = \frac{1}{2} \Gamma(d/2). \qquad (20.42)\end{aligned}$$

Combing the the last two equations with Lemma 20.26 which states that $I_d(1) = \pi^{d/2}$, we conclude that

$$\pi^{d/2} = I_d(1) = \frac{1}{2} \sigma(S^{d-1}) \Gamma(d/2)$$

which proves Eq. (20.40). Example 19.24 implies $\Gamma(1) = 1$ and from Eq. (20.42),

$$\begin{aligned}\Gamma(1/2) &= 2 \int_0^\infty e^{-r^2} dr = \int_{-\infty}^\infty e^{-r^2} dr \\ &= I_1(1) = \sqrt{\pi}.\end{aligned}$$

The relation, $\Gamma(x+1) = x\Gamma(x)$ is the consequence of the following integration by parts argument:

$$\begin{aligned}\Gamma(x+1) &= \int_0^\infty e^{-u}\; u^{x+1} \frac{du}{u} = \int_0^\infty u^x \left(-\frac{d}{du}\, e^{-u}\right) du \\ &= x \int_0^\infty u^{x-1}\; e^{-u}\; du = x\; \Gamma(x).\end{aligned}$$

■

BRUCE: add Morrey's Inequality **??** here.

20.4 More proofs of the classical Weierstrass approximation Theorem 8.34

In each of these proofs we will use the reduction explained the previous proof of Theorem 8.34 to reduce to the case where $f \in C([0,1]^d)$. The first proof we will give here is based on the "weak law" of large numbers. The second will be another approximate δ – function argument.

Proof. of Theorem 8.34. Let $\mathbf{0} : = (0, 0, \ldots, 0)$, $\mathbf{1} : = (1, 1, \ldots, 1)$ and $[\mathbf{0}, \mathbf{1}] := [0, 1]^d$. By considering the real and imaginary parts of f separately, it suffices to assume $f \in C([\mathbf{0}, \mathbf{1}], \mathbb{R})$. For $x \in [0, 1]$, let ν_x be the measure on $\{0, 1\}$ such that $\nu_x(\{0\}) = 1 - x$ and $\nu_x(\{1\}) = x$. Then

$$\int_{\{0,1\}} y d\nu_x(y) = 0 \cdot (1 - x) + 1 \cdot x = x \text{ and} \tag{20.43}$$

$$\int_{\{0,1\}} (y - x)^2 d\nu_x(y) = x^2(1 - x) + (1 - x)^2 \cdot x = x(1 - x). \tag{20.44}$$

For $x \in [\mathbf{0}, \mathbf{1}]$ let $\mu_x = \nu_{x_1} \otimes \cdots \otimes \nu_{x_d}$ be the product of $\nu_{x_1}, \ldots, \nu_{x_d}$ on $\Omega := \{0, 1\}^d$. Alternatively the measure μ_x may be described by

$$\mu_x(\{\varepsilon\}) = \prod_{i=1}^{d} (1 - x_i)^{1-\varepsilon_i} x_i^{\varepsilon_i} \tag{20.45}$$

for $\varepsilon \in \Omega$. Notice that $\mu_x(\{\varepsilon\})$ is a degree d polynomial in x for each $\varepsilon \in \Omega$. For $n \in \mathbb{N}$ and $x \in [\mathbf{0}, \mathbf{1}]$, let μ_x^n denote the n – fold product of μ_x with itself on Ω^n, $X_i(\omega) = \omega_i \in \Omega \subset \mathbb{R}^d$ for $\omega \in \Omega^n$ and let

$$S_n = (S_n^1, \ldots, S_n^d) := (X_1 + X_2 + \cdots + X_n)/n,$$

so $S_n : \Omega^n \to \mathbb{R}^d$. The reader is asked to verify (Exercise 20.4) that

$$\int_{\Omega^n} S_n d\mu_x^n := \left(\int_{\Omega^n} S_n^1 d\mu_x^n, \ldots, \int_{\Omega^n} S_n^d d\mu_x^n \right) = (x_1, \ldots, x_d) = x \tag{20.46}$$

and

$$\int_{\Omega^n} |S_n - x|^2 d\mu_x^n = \frac{1}{n} \sum_{i=1}^{d} x_i(1 - x_i) \le \frac{d}{n}. \tag{20.47}$$

From these equations it follows that S_n is concentrating near x as $n \to \infty$, a manifestation of the law of large numbers. Therefore it is reasonable to expect

$$p_n(x) := \int_{\Omega^n} f(S_n) d\mu_x^n \tag{20.48}$$

should approach $f(x)$ as $n \to \infty$. Let $\varepsilon > 0$ be given, $M = \sup\{|f(x)| : x \in [0, 1]\}$ and

$$\delta_\varepsilon = \sup\{|f(y) - f(x)| : x, y \in [\mathbf{0}, \mathbf{1}] \text{ and } |y - x| \le \varepsilon\}.$$

By uniform continuity of f on $[\mathbf{0}, \mathbf{1}]$, $\lim_{\varepsilon \downarrow 0} \delta_\varepsilon = 0$. Using these definitions and the fact that $\mu_x^n(\Omega^n) = 1$,

$$\begin{aligned}
|f(x) - p_n(x)| &= \left| \int_{\Omega^n} (f(x) - f(S_n)) d\mu_x^n \right| \le \int_{\Omega^n} |f(x) - f(S_n)| d\mu_x^n \\
&\le \int_{\{|S_n - x| > \varepsilon\}} |f(x) - f(S_n)| d\mu_x^n + \int_{\{|S_n - x| \le \varepsilon\}} |f(x) - f(S_n)| d\mu_x^n \\
&\le 2M\mu_x^n(|S_n - x| > \varepsilon) + \delta_\varepsilon. \qquad (20.49)
\end{aligned}$$

By Chebyshev's inequality,

$$\mu_x^n\left(|S_n - x| > \varepsilon\right) \le \frac{1}{\varepsilon^2}\int_{\Omega^n}(S_n - x)^2 d\mu_x^n = \frac{d}{n\varepsilon^2},$$

and therefore, Eq. (20.49) yields the estimate

$$\|f - p_n\|_\infty \le \frac{2dM}{n\varepsilon^2} + \delta_\varepsilon$$

and hence

$$\limsup_{n\to\infty}\|f - p_n\|_\infty \le \delta_\varepsilon \to 0 \text{ as } \varepsilon \downarrow 0.$$

This completes the proof since, using Eq. (20.45),

$$p_n(x) = \sum_{\omega\in\Omega^n} f(S_n(\omega))\mu_x^n(\{\omega\}) = \sum_{\omega\in\Omega^n} f(S_n(\omega))\prod_{i=1}^n \mu_x(\{\omega_i\}),$$

is an nd – degree polynomial in $x \in \mathbb{R}^d$). ■

Exercise 20.4. Verify Eqs. (20.46) and (20.47). This is most easily done using Eqs. (20.43) and (20.44) and Fubini's theorem repeatedly. (Of course Fubini's theorem here is over kill since these are only finite sums after all. Nevertheless it is convenient to use this formulation.)

The second proof requires the next two lemmas.

Lemma 20.30 (Approximate δ – sequences). *Suppose that $\{Q_n\}_{n=1}^\infty$ is a sequence of positive functions on $\mathbb{R}^d$ such that*

$$\int_{\mathbb{R}^d} Q_n(x)\, dx = 1 \text{ and} \tag{20.50}$$

$$\lim_{n\to\infty}\int_{|x|\ge\varepsilon} Q_n(x)dx = 0 \text{ for all } \varepsilon > 0. \tag{20.51}$$

For $f \in BC(\mathbb{R}^d)$, $Q_n * f$ converges to f uniformly on compact subsets of $\mathbb{R}^d$.

Proof. The proof is exactly the same as the proof of Lemma 8.28, it is only necessary to replace $\mathbb{R}$ by $\mathbb{R}^d$ everywhere in the proof. ■

Define

$$Q_n : \mathbb{R}^n \to [0,\infty) \text{ by } Q_n(x) = q_n(x_1)\dots q_n(x_d). \tag{20.52}$$

where q_n is defined in Eq. (8.23).

Lemma 20.31. *The sequence $\{Q_n\}_{n=1}^\infty$ is an approximate δ – sequence, i.e. they satisfy Eqs. (20.50) and (20.51).*

Proof. The fact that Q_n integrates to one is an easy consequence of Tonelli's theorem and the fact that q_n integrates to one. Since all norms on $\mathbb{R}^d$ are equivalent, we may assume that $|x| = \max\{|x_i| : i = 1, 2, \ldots, d\}$ when proving Eq. (20.51). With this norm

$$\left\{x \in \mathbb{R}^d : |x| \geq \varepsilon\right\} = \cup_{i=1}^d \left\{x \in \mathbb{R}^d : |x_i| \geq \varepsilon\right\}$$

and therefore by Tonelli's theorem,

$$\int_{\{|x|\geq\varepsilon\}} Q_n(x)dx \leq \sum_{i=1}^d \int_{\{|x_i|\geq\varepsilon\}} Q_n(x)dx = d \int_{\{x\in\mathbb{R}|x|\geq\varepsilon\}} q_n(t)dt$$

which tends to zero as $n \to \infty$ by Lemma 8.29. ■

Proof. Proof of Theorem 8.34. Again we assume $f \in C\left(\mathbb{R}^d, \mathbb{C}\right)$ and $f \equiv 0$ on Q_d^c where $Q_d := (0,1)^d$. Let $Q_n(x)$ be defined as in Eq. (20.52). Then by Lemma 20.31 and 20.30, $p_n(x) := (Q_n * F)(x) \to F(x)$ uniformly for $x \in [\mathbf{0}, \mathbf{1}]$ as $n \to \infty$. So to finish the proof it only remains to show $p_n(x)$ is a polynomial when $x \in [\mathbf{0}, \mathbf{1}]$. For $x \in [\mathbf{0}, \mathbf{1}]$,

$$\begin{aligned} p_n(x) &= \int_{\mathbb{R}^d} Q_n(x-y)f(y)dy \\ &= \frac{1}{c_n}\int_{[\mathbf{0},\mathbf{1}]} f(y) \prod_{i=1}^d \left[c_n^{-1}(1-(x_i-y_i)^2)^n 1_{|x_i-y_i|\leq 1}\right] dy \\ &= \frac{1}{c_n}\int_{[\mathbf{0},\mathbf{1}]} f(y) \prod_{i=1}^d \left[c_n^{-1}(1-(x_i-y_i)^2)^n\right] dy. \end{aligned}$$

Since the product in the above integrand is a polynomial if $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$, it follows easily that $p_n(x)$ is polynomial in x. ■

20.5 More Spherical Coordinates

In this section we will define spherical coordinates in all dimensions. Along the way we will develop an explicit method for computing surface integrals on spheres. As usual when $n = 2$ define spherical coordinates $(r, \theta) \in (0, \infty) \times [0, 2\pi)$ so that

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix} = T_2(\theta, r).$$

For $n = 3$ we let $x_3 = r\cos\phi_1$ and then

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = T_2(\theta, r\sin\phi_1),$$

as can be seen from Figure 20.6, so that

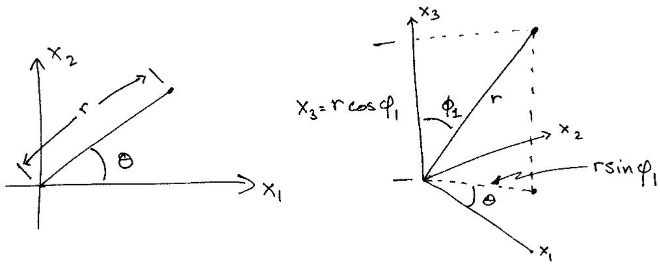


Fig. 20.6. Setting up polar coordinates in two and three dimensions.

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} T_2(\theta, r\sin\phi_1) \\ r\cos\phi_1 \end{pmatrix} = \begin{pmatrix} r\sin\phi_1\cos\theta \\ r\sin\phi_1\sin\theta \\ r\cos\phi_1 \end{pmatrix} =: T_3(\theta, \phi_1, r,).$$

We continue to work inductively this way to define

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \\ x_{n+1} \end{pmatrix} = \begin{pmatrix} T_n(\theta, \phi_1, \dots, \phi_{n-2}, r\sin\phi_{n-1},) \\ r\cos\phi_{n-1} \end{pmatrix} = T_{n+1}(\theta, \phi_1, \dots, \phi_{n-2}, \phi_{n-1}, r).$$

So for example,

$$\begin{aligned} x_1 &= r\sin\phi_2\sin\phi_1\cos\theta \\ x_2 &= r\sin\phi_2\sin\phi_1\sin\theta \\ x_3 &= r\sin\phi_2\cos\phi_1 \\ x_4 &= r\cos\phi_2 \end{aligned}$$

and more generally,

$$\begin{aligned} x_1 &= r\sin\phi_{n-2}\dots\sin\phi_2\sin\phi_1\cos\theta \\ x_2 &= r\sin\phi_{n-2}\dots\sin\phi_2\sin\phi_1\sin\theta \\ x_3 &= r\sin\phi_{n-2}\dots\sin\phi_2\cos\phi_1 \\ &\vdots \\ x_{n-2} &= r\sin\phi_{n-2}\sin\phi_{n-3}\cos\phi_{n-4} \\ x_{n-1} &= r\sin\phi_{n-2}\cos\phi_{n-3} \\ x_n &= r\cos\phi_{n-2}. \end{aligned} \tag{20.53}$$

By the change of variables formula,

$$\int_{\mathbb{R}^n} f(x)dm(x)$$
$$= \int_0^\infty dr \int_{0\le\phi_i\le\pi, 0\le\theta\le 2\pi} d\phi_1 \dots d\phi_{n-2} d\theta \Delta_n(\theta, \phi_1, \dots, \phi_{n-2}, r) f(T_n(\theta, \phi_1, \dots, \phi_{n-2}, r)) \tag{20.54}$$

where

$$\Delta_n(\theta, \phi_1, \dots, \phi_{n-2}, r) := \left|\det T_n'(\theta, \phi_1, \dots, \phi_{n-2}, r)\right|.$$

Proposition 20.32. *The Jacobian, Δ_n is given by*

$$\Delta_n(\theta, \phi_1, \dots, \phi_{n-2}, r) = r^{n-1} \sin^{n-2} \phi_{n-2} \dots \sin^2 \phi_2 \sin \phi_1. \tag{20.55}$$

If f is a function on rS^{n-1} – the sphere of radius r centered at 0 inside of $\mathbb{R}^n$, then

$$\int_{rS^{n-1}} f(x)d\sigma(x) = r^{n-1} \int_{S^{n-1}} f(r\omega)d\sigma(\omega)$$
$$= \int_{0\le\phi_i\le\pi, 0\le\theta\le 2\pi} f(T_n(\theta, \phi_1, \dots, \phi_{n-2}, r))\Delta_n(\theta, \phi_1, \dots, \phi_{n-2}, r)d\phi_1 \dots d\phi_{n-2} d\theta \tag{20.56}$$

Proof. We are going to compute Δ_n inductively. Letting $\rho := r \sin \phi_{n-1}$ and writing $\frac{\partial T_n}{\partial \xi}$ for $\frac{\partial T_n}{\partial \xi}(\theta, \phi_1, \dots, \phi_{n-2}, \rho)$ we have

$$\begin{aligned}
\Delta_{n+1}(\theta, &\phi_1, \dots, \phi_{n-2}, \phi_{n-1}, r) \\
&= \left| \begin{bmatrix} \frac{\partial T_n}{\partial \theta} & \frac{\partial T_n}{\partial \phi_1} & \dots & \frac{\partial T_n}{\partial \phi_{n-2}} & \frac{\partial T_n}{\partial \rho} r \cos \phi_{n-1} & \frac{\partial T_n}{\partial \rho} \sin \phi_{n-1} \\ 0 & 0 & \dots & 0 & -r \sin \phi_{n-1} & \cos \phi_{n-1} \end{bmatrix} \right| \\
&= r \left(\cos^2 \phi_{n-1} + \sin^2 \phi_{n-1}\right) \Delta_n(, \theta, \phi_1, \dots, \phi_{n-2}, \rho) \\
&= r \Delta_n(\theta, \phi_1, \dots, \phi_{n-2}, r \sin \phi_{n-1}),
\end{aligned}$$

i.e.

$$\Delta_{n+1}(\theta, \phi_1, \dots, \phi_{n-2}, \phi_{n-1}, r) = r\Delta_n(\theta, \phi_1, \dots, \phi_{n-2}, r \sin \phi_{n-1}). \tag{20.57}$$

To arrive at this result we have expanded the determinant along the bottom row. Staring with $\Delta_2(\theta, r) = r$ already derived in Example 20.23, Eq. (20.57) implies,

$$\begin{aligned}
\Delta_3(\theta, \phi_1, r) &= r\Delta_2(\theta, r \sin \phi_1) = r^2 \sin \phi_1 \\
\Delta_4(\theta, \phi_1, \phi_2, r) &= r\Delta_3(\theta, \phi_1, r \sin \phi_2) = r^3 \sin^2 \phi_2 \sin \phi_1 \\
&\vdots \\
\Delta_n(\theta, \phi_1, \dots, \phi_{n-2}, r) &= r^{n-1} \sin^{n-2} \phi_{n-2} \dots \sin^2 \phi_2 \sin \phi_1
\end{aligned}$$

which proves Eq. (20.55). Eq. (20.56) now follows from Eqs. (**??**), (20.54) and (20.55). ∎

As a simple application, Eq. (20.56) implies

$$\begin{aligned}\sigma(S^{n-1}) &= \int_{0\le\phi_i\le\pi,0\le\theta\le2\pi}\sin^{n-2}\phi_{n-2}\dots\sin^2\phi_2\sin\phi_1 d\phi_1\dots d\phi_{n-2}d\theta\\ &= 2\pi\prod_{k=1}^{n-2}\gamma_k = \sigma(S^{n-2})\gamma_{n-2} \qquad (20.58)\end{aligned}$$

where $\gamma_k := \int_0^\pi \sin^k \phi d\phi$. If $k \geq 1$, we have by integration by parts that,

$$\begin{aligned}\gamma_k &= \int_0^\pi \sin^k\phi d\phi = -\int_0^\pi \sin^{k-1}\phi\ d\cos\phi = 2\delta_{k,1} + (k-1)\int_0^\pi \sin^{k-2}\phi\ \cos^2\phi d\phi\\ &= 2\delta_{k,1} + (k-1)\int_0^\pi \sin^{k-2}\phi\ \left(1-\sin^2\phi\right)d\phi = 2\delta_{k,1} + (k-1)\left[\gamma_{k-2}-\gamma_k\right]\end{aligned}$$

and hence γ_k satisfies $\gamma_0 = \pi$, $\gamma_1 = 2$ and the recursion relation

$$\gamma_k = \frac{k-1}{k}\gamma_{k-2} \text{ for } k \geq 2.$$

Hence we may conclude

$$\gamma_0 = \pi,\ \gamma_1 = 2,\ \gamma_2 = \frac{1}{2}\pi,\ \gamma_3 = \frac{2}{3}2,\ \gamma_4 = \frac{3}{4}\frac{1}{2}\pi,\ \gamma_5 = \frac{4}{5}\frac{2}{3}2,\ \gamma_6 = \frac{5}{6}\frac{3}{4}\frac{1}{2}\pi$$

and more generally by induction that

$$\gamma_{2k} = \pi\frac{(2k-1)!!}{(2k)!!} \text{ and } \gamma_{2k+1} = 2\frac{(2k)!!}{(2k+1)!!}.$$

Indeed,

$$\gamma_{2(k+1)+1} = \frac{2k+2}{2k+3}\gamma_{2k+1} = \frac{2k+2}{2k+3}2\frac{(2k)!!}{(2k+1)!!} = 2\frac{[2(k+1)]!!}{(2(k+1)+1)!!}$$

and

$$\gamma_{2(k+1)} = \frac{2k+1}{2k+1}\gamma_{2k} = \frac{2k+1}{2k+2}\pi\frac{(2k-1)!!}{(2k)!!} = \pi\frac{(2k+1)!!}{(2k+2)!!}.$$

The recursion relation in Eq. (20.58) may be written as

$$\sigma(S^n) = \sigma\left(S^{n-1}\right)\gamma_{n-1} \qquad (20.59)$$

which combined with $\sigma\left(S^1\right) = 2\pi$ implies

$$
\begin{aligned}
\sigma\left(S^1\right) &= 2\pi, \\
\sigma(S^2) &= 2\pi \cdot \gamma_1 = 2\pi \cdot 2, \\
\sigma(S^3) &= 2\pi \cdot 2 \cdot \gamma_2 = 2\pi \cdot 2 \cdot \frac{1}{2}\pi = \frac{2^2\pi^2}{2!!}, \\
\sigma(S^4) &= \frac{2^2\pi^2}{2!!} \cdot \gamma_3 = \frac{2^2\pi^2}{2!!} \cdot 2\frac{2}{3} = \frac{2^3\pi^2}{3!!} \\
\sigma(S^5) &= 2\pi \cdot 2 \cdot \frac{1}{2}\pi \cdot \frac{2}{3}2 \cdot \frac{3}{4}\frac{1}{2}\pi = \frac{2^3\pi^3}{4!!}, \\
\sigma(S^6) &= 2\pi \cdot 2 \cdot \frac{1}{2}\pi \cdot \frac{2}{3}2 \cdot \frac{3}{4}\frac{1}{2}\pi \cdot \frac{4}{5}\frac{2}{3}2 = \frac{2^4\pi^3}{5!!}
\end{aligned}
$$

and more generally that

$$
\sigma(S^{2n}) = \frac{2\left(2\pi\right)^n}{(2n-1)!!} \text{ and } \sigma(S^{2n+1}) = \frac{\left(2\pi\right)^{n+1}}{(2n)!!} \tag{20.60}
$$

which is verified inductively using Eq. (20.59). Indeed,

$$
\sigma(S^{2n+1}) = \sigma(S^{2n})\gamma_{2n} = \frac{2\left(2\pi\right)^n}{(2n-1)!!}\pi\frac{(2n-1)!!}{(2n)!!} = \frac{\left(2\pi\right)^{n+1}}{(2n)!!}
$$

and

$$
\sigma(S^{(n+1)}) = \sigma(S^{2n+2}) = \sigma(S^{2n+1})\gamma_{2n+1} = \frac{\left(2\pi\right)^{n+1}}{(2n)!!}2\frac{(2n)!!}{(2n+1)!!} = \frac{2\left(2\pi\right)^{n+1}}{(2n+1)!!}.
$$

Using

$$
(2n)!! = 2n\left(2(n-1)\right)\dots(2\cdot 1) = 2^n n!
$$

we may write $\sigma(S^{2n+1}) = \frac{2\pi^{n+1}}{n!}$ which shows that Eqs. (**??**) and (20.60 in agreement. We may also write the formula in Eq. (20.60) as

$$
\sigma(S^n) = \begin{cases} \frac{2(2\pi)^{n/2}}{(n-1)!!} & \text{for } n \text{ even} \\ \frac{(2\pi)^{\frac{n+1}{2}}}{(n-1)!!} & \text{for } n \text{ odd.} \end{cases}
$$

20.6 Sard's Theorem

See p. 538 of Taylor and references. Also see Milnor's topology book. Add in the Brower's Fixed point theorem here as well. Also Spivak's calculus on manifolds.

Theorem 20.33. *Let $U \subset_o \mathbb{R}^m$, $f \in C^\infty(U, \mathbb{R}^d)$ and $C := \{x \in U : \operatorname{Ran}\mathrm{k}(f'(x)) < n\}$ be the set of critical points of f. Then the critical values, $f(C)$, is a Borel measurable subset of $\mathbb{R}^d$ of Lebesgue measure* 0.

Remark 20.34. This result clearly extends to manifolds.

For simplicity in the proof given below it will be convenient to use the norm, $|x| := \max_i |x_i|$. Recall that if $f \in C^1(U, \mathbb{R}^d)$ and $p \in U$, then

$$f(p+x) = f(p) + \int_0^1 f'(p+tx)x dt = f(p) + f'(p)x + \int_0^1 \left[f'(p+tx) - f'(p)\right] x dt$$

so that if

$$R(p,x) := f(p+x) - f(p) - f'(p)x = \int_0^1 \left[f'(p+tx) - f'(p)\right] x dt$$

we have

$$|R(p,x)| \le |x| \int_0^1 |f'(p+tx) - f'(p)| \, dt = |x| \, \varepsilon(p,x).$$

By uniform continuity, it follows for any compact subset $K \subset U$ that

$$\sup \{|\varepsilon(p,x)| : p \in K \text{ and } |x| \le \delta\} \to 0 \text{ as } \delta \downarrow 0.$$

Proof. Notice that if $x \in U \setminus C$, then $f'(x) : \mathbb{R}^m \to \mathbb{R}^n$ is surjective, which is an open condition, so that $U \setminus C$ is an open subset of U. This shows C is relatively closed in U, i.e. there exists $\tilde{C} \sqsubset \mathbb{R}^m$ such that $C = \tilde{C} \cap U$. Let $K_n \subset U$ be compact subsets of U such that $K_n \uparrow U$, then $K_n \cap C \uparrow C$ and $K_n \cap C = K_n \cap \tilde{C}$ is compact for each n. Therefore, $f(K_n \cap C) \uparrow f(C)$ i.e. $f(C) = \cup_n f(K_n \cap C)$ is a countable union of compact sets and therefore is Borel measurable. Moreover, since $m(f(C)) = \lim_{n\to\infty} m(f(K_n \cap C))$, it suffices to show $m(f(K)) = 0$ for all compact subsets $K \subset C$. Case 1. $(n \le m)$ Let $K = [a, a+\gamma]$ be a cube contained in U and by scaling the domain we may assume $\gamma = (1,1,1,\dots,1)$. For $N \in \mathbb{N}$ and $j \in S_N := \{0,1,\dots,N-1\}^n$ let $K_j = j/N + [a, a+\gamma/N]$ so that $K = \cup_{j \in S_N} K_j$ with $K_j^o \cap K_{j'}^o = \emptyset$ if $j \neq j'$. Let $\{Q_j : j = 1\dots,M\}$ be the collection of those $\{K_j : j \in S_N\}$ which intersect C. For each j, let $p_j \in Q_j \cap C$ and for $x \in Q_j - p_j$ we have

$$f(p_j + x) = f(p_j) + f'(p_j)x + R_j(x)$$

where $|R_j(x)| \le \varepsilon_j(N)/N$ and $\varepsilon(N) := \max_j \varepsilon_j(N) \to 0$ as $N \to \infty$. Now

$$\begin{aligned} m(f(Q_j)) &= m\left(f(p_j) + \left(f'(p_j) + R_j\right)(Q_j - p_j)\right) \\ &= m\left(\left(f'(p_j) + R_j\right)(Q_j - p_j)\right) \\ &= m\left(O_j\left(f'(p_j) + R_j\right)(Q_j - p_j)\right) \end{aligned} \tag{20.61}$$

where $O_j \in SO(n)$ is chosen so that $O_j f'(p_j)\mathbb{R}^n \subset \mathbb{R}^{m-1} \times \{0\}$. Now $O_j f'(p_j)(Q_j - p_j)$ is contained in $\Gamma \times \{0\}$ where $\Gamma \subset \mathbb{R}^{m-1}$ is a cube centered at $0 \in \mathbb{R}^{m-1}$ with side length at most $2|f'(p_j)|/N \le 2M/N$ where

$M = \max_{p\in K} |f'(p)|$. It now follows that $O_j(f'(p_j) + R_j)(Q_j - p_j)$ is contained the set of all points within $\varepsilon(N)/N$ of $\Gamma \times \{0\}$ and in particular

$$O_j(f'(p_j) + R_j)(Q_j - p_j) \subset (1 + \varepsilon(N)/N)\,\Gamma \times [\varepsilon(N)/N, \varepsilon(N)/N].$$

From this inclusion and Eq. (20.61) it follows that

$$\begin{aligned} m(f(Q_j)) &\le \left[2\frac{M}{N}(1 + \varepsilon(N)/N)\right]^{m-1} 2\varepsilon(N)/N \\ &= 2^m M^{m-1}\left[(1 + \varepsilon(N)/N)\right]^{m-1} \varepsilon(N)\frac{1}{N^m} \end{aligned}$$

and therefore,

$$\begin{aligned} m(f(C\cap K)) &\le \sum_j m(f(Q_j)) \le N^n 2^m M^{m-1}\left[(1 + \varepsilon(N)/N)\right]^{m-1} \varepsilon(N)\frac{1}{N^m} \\ &= 2^n M^{n-1}\left[(1 + \varepsilon(N)/N)\right]^{n-1} \varepsilon(N)\frac{1}{N^{m-n}} \to 0 \text{ as } N \to \infty \end{aligned}$$

since $m \ge n$. This proves the easy case since we may write U as a countable union of cubes K as above. **Remark**. The case $(m < n)$ also follows from the case $m = n$ as follows. When $m < n$, $C = U$ and we must show $m(f(U)) = 0$. Letting $F : U\times\mathbb{R}^{n-m} \to \mathbb{R}^n$ be the map $F(x, y) = f(x)$. Then $F'(x, y)(v, w) = f'(x)v$, and hence $C_F := U \times \mathbb{R}^{n-m}$. So if the assertion holds for $m = n$ we have

$$m(f(U)) = m(F(U \times \mathbb{R}^{n-m})) = 0.$$

Case 2. $(m > n)$ This is the hard case and the case we will need in the co-area formula to be proved later. Here I will follow the proof in Milnor. Let

$$C_i := \{x \in U : \partial^\alpha f(x) = 0 \text{ when } |\alpha| \le i\}$$

so that $C \supset C_1 \supset C_2 \supset C_3 \supset \ldots$. The proof is by induction on n and goes by the following steps:

1. $m(f(C \setminus C_1)) = 0$.
2. $m(f(C_i \setminus C_{i+1})) = 0$ for all $i \ge 1$.
3. $m(f(C_i)) = 0$ for all i sufficiently large.

Step 1. If $m = 1$, there is nothing to prove since $C = C_1$ so we may assume $m \ge 2$. Suppose that $x \in C \setminus C_1$, then $f'(p) \ne 0$ and so by reordering the components of x and $f(p)$ if necessary we may assume that $\partial_1 f_1(p) \ne 0$ where we are writing $\partial f(p)/\partial x_i$ as $\partial_i f(p)$. The map $h(x) := (f_1(x), x_2, \ldots, x_n)$ has differential

$$h'(p) = \begin{bmatrix} \partial_1 f_1(p) & \partial_2 f_1(p) & \ldots & \partial_n f_1(p) \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

which is not singular. So by the implicit function theorem, there exists there exists $V \in \tau_p$ such that $h : V \to h(V) \in \tau_{h(p)}$ is a diffeomorphism and in particular $\partial f_1(x)/\partial x_1 \neq 0$ for $x \in V$ and hence $V \subset U \setminus C_1$. Consider the map $g := f \circ h^{-1} : V' := h(V) \to \mathbb{R}^m$, which satisfies

$$(f_1(x), f_2(x), \ldots, f_m(x)) = f(x) = g(h(x)) = g((f_1(x), x_2, \ldots, x_n))$$

which implies $g(t, y) = (t, u(t, y))$ for $(t, y) \in V' := h(V) \in \tau_{h(p)}$, see Figure 20.7 below where $p = \bar{x}$ and $m = p$. Since

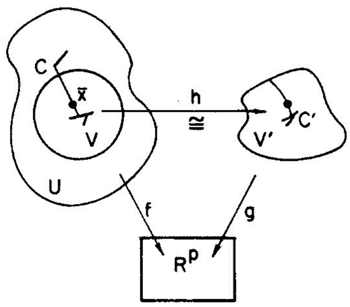


Figure . Construction of the map g

Fig. 20.7. Making a change of variable so as to apply induction.

$$g'(t, y) = \begin{bmatrix} 1 & 0 \\ \partial_t u(t, y) & \partial_y u(t, y) \end{bmatrix}$$

it follows that (t, y) is a critical point of g iff $y \in C'_t$ – the set of critical points of $y \to u(t, y)$. Since h is a diffeomorphism we have $C' := h(C \cap V)$ are the critical points of g in V' and

$$f(C \cap V) = g(C') = \cup_t \left[\{t\} \times u_t(C'_t)\right].$$

By the induction hypothesis, $m_{m-1}(u_t(C'_t)) = 0$ for all t, and therefore by Fubini's theorem,

$$m(f(C \cap V)) = \int_{\mathbb{R}} m_{m-1}(u_t(C'_t)) 1_{V'_t \neq \emptyset} dt = 0.$$

Since $C \setminus C_1$ may be covered by a countable collection of open sets V as above, it follows that $m(f(C \setminus C_1)) = 0$. **Step 2.** Suppose that $p \in C_k \setminus C_{k+1}$, then there is an α such that $|\alpha| = k + 1$ such that $\partial^\alpha f(p) = 0$ while $\partial^\beta f(p) = 0$ for all $|\beta| \leq k$. Again by permuting coordinates we may assume that $\alpha_1 \neq 0$ and $\partial^\alpha f_1(p) \neq 0$. Let $w(x) := \partial^{\alpha - e_1} f_1(x)$, then $w(p) = 0$ while $\partial_1 w(p) \neq 0$.

So again the implicit function theorem there exists $V \in \tau_p$ such that $h(x) := (w(x), x_2, \ldots, x_n)$ maps $V \to V' := h(V) \in \tau_{h(p)}$ in a diffeomorphic way and in particular $\partial_1 w(x) \neq 0$ on V so that $V \subset U \backslash C_{k+1}$. As before, let $g := f \circ h^{-1}$ and notice that $C_k' := h(C_k \cap V) \subset \{0\} \times \mathbb{R}^{n-1}$ and

$$f(C_k \cap V) = g(C_k') = \bar{g}(C_k')$$

where $\bar{g} := g|_{(\{0\} \times \mathbb{R}^{n-1}) \cap V'}$. Clearly C_k' is contained in the critical points of $\bar{g}$, and therefore, by induction

$$0 = m(\bar{g}(C_k')) = m(f(C_k \cap V)).$$

Since $C_k \backslash C_{k+1}$ is covered by a countable collection of such open sets, it follows that

$$m(f(C_k \setminus C_{k+1})) = 0 \text{ for all } k \geq 1.$$

Step 3. Suppose that Q is a closed cube with edge length δ contained in U and $k > n/m - 1$. We will show $m(f(Q \cap C_k)) = 0$ and since Q is arbitrary it will follows that $m(f(C_k)) = 0$ as desired. By Taylor's theorem with (integral) remainder, it follows for $x \in Q \cap C_k$ and h such that $x + h \in Q$ that

$$f(x+h) = f(x) + R(x,h)$$

where

$$|R(x,h)| \leq c \|h\|^{k+1}$$

where $c = c(Q,k)$. Now subdivide Q into r^n cubes of edge size δ/r and let Q' be one of the cubes in this subdivision such that $Q' \cap C_k \neq \emptyset$ and let $x \in Q' \cap C_k$. It then follows that $f(Q')$ is contained in a cube centered at $f(x) \in \mathbb{R}^m$ with side length at most $2c(\delta/r)^{k+1}$ and hence volume at most $(2c)^m (\delta/r)^{m(k+1)}$. Therefore, $f(Q \cap C_k)$ is contained in the union of at most r^n cubes of volume $(2c)^m (\delta/r)^{m(k+1)}$ and hence $m$each

$$m(f(Q \cap C_k)) \leq (2c)^m (\delta/r)^{m(k+1)} r^n = (2c)^m \delta^{m(k+1)} r^{n-m(k+1)} \to 0 \text{ as } r \uparrow \infty$$

provided that $n - m(k+1) < 0$, i.e. provided $k > n/m - 1$. ■

20.7 Exercises

Exercise 20.5. Prove Theorem 20.12. Suggestion, to get started define

$$\pi(A) := \int_{X_1} d\mu(x_1) \ldots \int_{X_n} d\mu(x_n) 1_A(x_1, \ldots, x_n)$$

and then show Eq. (20.18) holds. Use the case of two factors as the model of your proof.

Exercise 20.6. Let $(X_j, \mathcal{M}_j, \mu_j)$ for $j = 1, 2, 3$ be σ – finite measure spaces. Let $F : (X_1 \times X_2) \times X_3 \to X_1 \times X_2 \times X_3$ be defined by

$$F((x_1, x_2), x_3) = (x_1, x_2, x_3).$$

1. Show F is $((\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3, \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3)$ – measurable and F^{-1} is $(\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3, (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3)$ – measurable. That is
$$F : ((X_1 \times X_2) \times X_3, (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3) \to (X_1 \times X_2 \times X_3, \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3)$$
is a "measure theoretic isomorphism."
2. Let $\pi := F_* [(\mu_1 \otimes \mu_2) \otimes \mu_3]$, i.e. $\pi(A) = [(\mu_1 \otimes \mu_2) \otimes \mu_3] (F^{-1}(A))$ for all $A \in \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3$. Then π is the unique measure on $\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3$ such that
$$\pi(A_1 \times A_2 \times A_3) = \mu_1(A_1)\mu_2(A_2)\mu_3(A_3)$$
for all $A_i \in \mathcal{M}_i$. We will write $\pi := \mu_1 \otimes \mu_2 \otimes \mu_3$.
3. Let $f : X_1 \times X_2 \times X_3 \to [0, \infty]$ be a $(\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function. Verify the identity,
$$\int_{X_1 \times X_2 \times X_3} f d\pi = \int_{X_3} d\mu_3(x_3) \int_{X_2} d\mu_2(x_2) \int_{X_1} d\mu_1(x_1) f(x_1, x_2, x_3),$$
makes sense and is correct.
4. (Optional.) Also show the above identity holds for any one of the six possible orderings of the iterated integrals.

Exercise 20.7. Prove the second assertion of Theorem 20.18. That is show m^d is the unique translation invariant measure on $\mathcal{B}_{\mathbb{R}^d}$ such that $m^d((0, 1]^d) = 1$. **Hint:** Look at the proof of Theorem 19.10.

Exercise 20.8. (Part of Folland Problem 2.46 on p. 69.) Let $X = [0, 1]$, $\mathcal{M} = \mathcal{B}_{[0,1]}$ be the Borel σ – field on X, m be Lebesgue measure on $[0, 1]$ and ν be counting measure, $\nu(A) = \#(A)$. Finally let $D = \{(x, x) \in X^2 : x \in X\}$ be the diagonal in X^2. Show

$$\int_X \left[\int_X 1_D(x, y) d\nu(y) \right] dm(x) \neq \int_X \left[\int_X 1_D(x, y) dm(x) \right] d\nu(y)$$

by explicitly computing both sides of this equation.

Exercise 20.9. Folland Problem 2.48 on p. 69. (Counter example related to Fubini Theorem involving counting measures.)

Exercise 20.10. Folland Problem 2.50 on p. 69 pertaining to area under a curve. (Note the $\mathcal{M} \times \mathcal{B}_{\mathbb{R}}$ should be $\mathcal{M} \otimes \mathcal{B}_{\bar{\mathbb{R}}}$ in this problem.)

Exercise 20.11. Folland Problem 2.55 on p. 77. (Explicit integrations.)

Exercise 20.12. Folland Problem 2.56 on p. 77. Let $f \in L^1((0,a), dm)$, $g(x) = \int_x^a \frac{f(t)}{t} dt$ for $x \in (0,a)$, show $g \in L^1((0,a), dm)$ and

$$\int_0^a g(x)dx = \int_0^a f(t)dt.$$

Exercise 20.13. Show $\int_0^\infty \left|\frac{\sin x}{x}\right| dm(x) = \infty$. So $\frac{\sin x}{x} \notin L^1([0,\infty), m)$ and $\int_0^\infty \frac{\sin x}{x} dm(x)$ is not defined as a Lebesgue integral.

Exercise 20.14. Folland Problem 2.57 on p. 77.

Exercise 20.15. Folland Problem 2.58 on p. 77.

Exercise 20.16. Folland Problem 2.60 on p. 77. Properties of the Γ – function.

Exercise 20.17. Folland Problem 2.61 on p. 77. Fractional integration.

Exercise 20.18. Folland Problem 2.62 on p. 80. Rotation invariance of surface measure on S^{n-1}.

Exercise 20.19. Folland Problem 2.64 on p. 80. On the integrability of $|x|^a \left|\log |x|\right|^b$ for x near 0 and x near ∞ in $\mathbb{R}^n$.

Exercise 20.20. Show, using Problem 20.18 that

$$\int_{S^{d-1}} \omega_i \omega_j d\sigma(\omega) = \frac{1}{d}\delta_{ij}\sigma\left(S^{d-1}\right).$$

Hint: show $\int_{S^{d-1}} \omega_i^2 d\sigma(\omega)$ is independent of i and therefore

$$\int_{S^{d-1}} \omega_i^2 d\sigma(\omega) = \frac{1}{d}\sum_{j=1}^d \int_{S^{d-1}} \omega_j^2 d\sigma(\omega).$$

21

L^p-spaces

Let $(X, \mathcal{M}, \mu)$ be a measure space and for $0 < p < \infty$ and a measurable function $f : X \to \mathbb{C}$ let

$$\|f\|_p := \left(\int_X |f|^p \, d\mu \right)^{1/p}. \tag{21.1}$$

When $p = \infty$, let

$$\|f\|_\infty = \inf \{a \geq 0 : \mu(|f| > a) = 0\} \tag{21.2}$$

For $0 < p \leq \infty$, let

$$L^p(X, \mathcal{M}, \mu) = \{f : X \to \mathbb{C} : f \text{ is measurable and } \|f\|_p < \infty\} / \sim$$

where $f \sim g$ iff $f = g$ a.e. Notice that $\|f - g\|_p = 0$ iff $f \sim g$ and if $f \sim g$ then $\|f\|_p = \|g\|_p$. In general we will (by abuse of notation) use f to denote both the function f and the equivalence class containing f.

Remark 21.1. Suppose that $\|f\|_\infty \leq M$, then for all $a > M$, $\mu(|f| > a) = 0$ and therefore $\mu(|f| > M) = \lim_{n\to\infty} \mu(|f| > M + 1/n) = 0$, i.e. $|f(x)| \leq M$ for μ - a.e. x. Conversely, if $|f| \leq M$ a.e. and $a > M$ then $\mu(|f| > a) = 0$ and hence $\|f\|_\infty \leq M$. This leads to the identity:

$$\|f\|_\infty = \inf \{a \geq 0 : |f(x)| \leq a \text{ for } \mu - \text{a.e. } x\}.$$

The next theorem is a generalization Theorem 5.6 to general integrals and the proof is essentially identical to the proof of Theorem 5.6.

Theorem 21.2 (Hölder's inequality). *Suppose that $1 \leq p \leq \infty$ and $q := \frac{p}{p-1}$, or equivalently $p^{-1} + q^{-1} = 1$. If f and g are measurable functions then*

$$\|fg\|_1 \leq \|f\|_p \cdot \|g\|_q. \tag{21.3}$$

Assuming $p \in (1, \infty)$ and $\|f\|_p \cdot \|g\|_q < \infty$, equality holds in Eq. (21.3) iff $|f|^p$ and $|g|^q$ are linearly dependent as elements of L^1 which happens iff

$$|g|^q \|f\|_p^p = \|g\|_q^q \, |f|^p \quad a.e. \tag{21.4}$$

Proof. The cases where $\|f\|_q = 0$ or ∞ or $\|g\|_p = 0$ or ∞ are easy to deal with and are left to the reader. So we will now assume that $0 < \|f\|_q, \|g\|_p < \infty$. Let $s = |f| / \|f\|_p$ and $t = |g|/\|g\|_q$ then Lemma 5.5 implies

$$\frac{|fg|}{\|f\|_p \|g\|_q} \leq \frac{1}{p} \frac{|f|^p}{\|f\|_p} + \frac{1}{q} \frac{|g|^q}{\|g\|^q} \tag{21.5}$$

with equality iff $|g/\|g\|_q| = |f|^{p-1} / \|f\|_p^{(p-1)} = |f|^{p/q} / \|f\|_p^{p/q}$, i.e. $|g|^q \|f\|_p^p = \|g\|_q^q \, |f|^p$. Integrating Eq. (21.5) implies

$$\frac{\|fg\|_1}{\|f\|_p \|g\|_q} \leq \frac{1}{p} + \frac{1}{q} = 1$$

with equality iff Eq. (21.4) holds. The proof is finished since it is easily checked that equality holds in Eq. (21.3) when $|f|^p = c\,|g|^q$ of $|g|^q = c\,|f|^p$ for some constant c. ∎

The following corollary is an easy extension of Hölder's inequality.

Corollary 21.3. *Suppose that $f_i : X \to \mathbb{C}$ are measurable functions for $i = 1, \dots, n$ and $p_1, \dots, p_n$ and r are positive numbers such that $\sum_{i=1}^n p_i^{-1} = r^{-1}$, then*

$$\left\| \prod_{i=1}^n f_i \right\|_r \leq \prod_{i=1}^n \|f_i\|_{p_i} \text{ where } \sum_{i=1}^n p_i^{-1} = r^{-1}.$$

Proof. To prove this inequality, start with $n = 2$, then for any $p \in [1, \infty]$,

$$\|fg\|_r^r = \int_X |f|^r \, |g|^r \, d\mu \leq \|f^r\|_p \, \|g^r\|_{p^*}$$

where $p^* = \frac{p}{p-1}$ is the conjugate exponent. Let $p_1 = pr$ and $p_2 = p^* r$ so that $p_1^{-1} + p_2^{-1} = r^{-1}$ as desired. Then the previous equation states that

$$\|fg\|_r \leq \|f\|_{p_1} \|g\|_{p_2}$$

as desired. The general case is now proved by induction. Indeed,

$$\left\| \prod_{i=1}^{n+1} f_i \right\|_r = \left\| \prod_{i=1}^{n} f_i \cdot f_{n+1} \right\|_r \leq \left\| \prod_{i=1}^{n} f_i \right\|_q \|f_{n+1}\|_{p_{n+1}}$$

where $q^{-1} + p_{n+1}^{-1} = r^{-1}$. Since $\sum_{i=1}^n p_i^{-1} = q^{-1}$, we may now use the induction hypothesis to conclude

$$\left\| \prod_{i=1}^{n} f_i \right\|_q \leq \prod_{i=1}^{n} \|f_i\|_{p_i},$$

which combined with the previous displayed equation proves the generalized form of Holder's inequality. ∎

Theorem 21.4 (Minkowski's Inequality). *If $1 \le p \le \infty$ and $f, g \in L^p$ then*

$$\|f+g\|_p \le \|f\|_p + \|g\|_p. \tag{21.6}$$

Moreover, assuming f and g are not identically zero, equality holds in Eq. (21.6) iff $\operatorname{sgn}(f) \overset{\circ}{=} \operatorname{sgn}(g)$ a.e. (see the notation in Definition 5.7) when $p = 1$ and $f = cg$ a.e. for some $c > 0$ for $p \in (1, \infty)$.

Proof. When $p = \infty$, $|f| \le \|f\|_\infty$ a.e. and $|g| \le \|g\|_\infty$ a.e. so that $|f+g| \le |f| + |g| \le \|f\|_\infty + \|g\|_\infty$ a.e. and therefore

$$\|f+g\|_\infty \le \|f\|_\infty + \|g\|_\infty .$$

When $p < \infty$,

$$|f+g|^p \le (2\max(|f|, |g|))^p = 2^p \max(|f|^p, |g|^p) \le 2^p (|f|^p + |g|^p),$$

$$\|f+g\|_p^p \le 2^p \left(\|f\|_p^p + \|g\|_p^p\right) < \infty.$$

In case $p = 1$,

$$\|f+g\|_1 = \int_X |f+g| d\mu \le \int_X |f| \, d\mu + \int_X |g| d\mu$$

with equality iff $|f| + |g| = |f+g|$ a.e. which happens iff $\operatorname{sgn}(f) \overset{\circ}{=} \operatorname{sgn}(g)$ a.e. In case $p \in (1, \infty)$, we may assume $\|f+g\|_p$, $\|f\|_p$ and $\|g\|_p$ are all positive since otherwise the theorem is easily verified. Now

$$|f+g|^p = |f+g||f+g|^{p-1} \le (|f| + |g|)|f+g|^{p-1}$$

with equality iff $\operatorname{sgn}(f) \overset{\circ}{=} \operatorname{sgn}(g)$. Integrating this equation and applying Holder's inequality with $q = p/(p-1)$ gives

$$\begin{aligned}\int_X |f+g|^p d\mu &\le \int_X |f| \; |f+g|^{p-1} d\mu + \int_X |g| \; |f+g|^{p-1} d\mu \\ &\le (\|f\|_p + \|g\|_p) \, \| \, |f+g|^{p-1} \, \|_q \end{aligned} \tag{21.7}$$

with equality iff

$$\begin{aligned} &\operatorname{sgn}(f) \overset{\circ}{=} \operatorname{sgn}(g) \text{ and} \\ &\left(\frac{|f|}{\|f\|_p}\right)^p = \frac{|f+g|^p}{\|f+g\|_p^p} = \left(\frac{|g|}{\|g\|_p}\right)^p \text{ a.e.} \end{aligned} \tag{21.8}$$

Therefore

$$\||f+g|^{p-1}\|_q^q = \int_X (|f+g|^{p-1})^q d\mu = \int_X |f+g|^p d\mu. \tag{21.9}$$

Combining Eqs. (21.7) and (21.9) implies

$$\|f+g\|_p^p \le \|f\|_p \|f+g\|_p^{p/q} + \|g\|_p \|f+g\|_p^{p/q} \tag{21.10}$$

with equality iff Eq. (21.8) holds which happens iff $f = cg$ a.e. with $c > 0$. Solving for $\|f+g\|_p$ in Eq. (21.10) gives Eq. (21.6). ■

The next theorem gives another example of using Hölder's inequality

Theorem 21.5. *Suppose that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ – finite measure spaces, $p \in [1, \infty]$, $q = p/(p-1)$ and $k : X \times Y \to \mathbb{C}$ be a $\mathcal{M} \otimes \mathcal{N}$ – measurable function. Assume there exist finite constants C_1 and C_2 such that*

$$\int_X |k(x,y)|\, d\mu(x) \le C_1 \text{ for } \nu \text{ a.e. } y \text{ and}$$
$$\int_Y |k(x,y)|\, d\nu(y) \le C_2 \text{ for } \mu \text{ a.e. } x.$$

If $f \in L^p(\nu)$, then

$$\int_Y |k(x,y) f(y)|\, d\nu(y) < \infty \text{ for } \mu \text{ – a.e. } x,$$

$x \to Kf(x) := \int_Y k(x,y) f(y) d\nu(y) \in L^p(\mu)$ and

$$\|Kf\|_{L^p(\mu)} \le C_1^{1/p} C_2^{1/q} \|f\|_{L^p(\nu)} \tag{21.11}$$

Proof. Suppose $p \in (1, \infty)$ to begin with and let $q = p/(p-1)$, then by Hölder's inequality,

$$\begin{aligned}
\int_Y |k(x,y) f(y)|\, d\nu(y) &= \int_Y |k(x,y)|^{1/q} |k(x,y)|^{1/p} |f(y)|\, d\nu(y) \\
&\le \left[\int_Y |k(x,y)|\, d\nu(y)\right]^{1/q} \left[\int_Y |k(x,y)|\, |f(y)|^p\, d\nu(y)\right]^{1/p} \\
&\le C_2^{1/q} \left[\int_Y |k(x,y)|\, |f(y)|^p\, d\nu(y)\right]^{1/p}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left\| \int_Y |k(\cdot,y) f(y)|\, d\nu(y) \right\|_{L^p(\mu)}^p &= \int_X d\mu(x) \left[\int_Y |k(x,y) f(y)|\, d\nu(y)\right]^p \\
&\le C_2^{p/q} \int_X d\mu(x) \int_Y d\nu(y)\, |k(x,y)|\, |f(y)|^p \\
&= C_2^{p/q} \int_Y d\nu(y)\, |f(y)|^p \int_X d\mu(x)\, |k(x,y)| \\
&\le C_2^{p/q} C_1 \int_Y d\nu(y)\, |f(y)|^p = C_2^{p/q} C_1 \|f\|_{L^p(\nu)}^p,
\end{aligned}$$

wherein we used Tonelli's theorem in third line. From this it follows that $\int_Y |k(x,y)f(y)|\,d\nu(y) < \infty$ for μ - a.e. x,

$$x \to Kf(x) := \int_Y k(x,y)f(y)d\nu(y) \in L^p(\mu)$$

and that Eq. (21.11) holds.

Similarly if $p = \infty$,

$$\int_Y |k(x,y)f(y)|\,d\nu(y) \le \|f\|_{L^\infty(\nu)} \cdot \int_Y |k(x,y)|\,d\nu(y) \le C_2 \|f\|_{L^\infty(\nu)} \text{ for } \mu - \text{a.e. } x.$$

so that $\|Kf\|_{L^\infty(\mu)} \le C_2 \|f\|_{L^\infty(\nu)}$. If $p = 1$, then

$$\begin{aligned}\int_X d\mu(x) \int_Y d\nu(y)\,|k(x,y)f(y)| &= \int_Y d\nu(y)\,|f(y)| \int_X d\mu(x)\,|k(x,y)| \\ &\le C_1 \int_Y d\nu(y)\,|f(y)|\end{aligned}$$

which shows $\|Kf\|_{L^1(\mu)} \le C_1 \|f\|_{L^1(\nu)}$. ∎

21.1 Jensen's Inequality

Definition 21.6. *A function $\phi : (a,b) \to \mathbb{R}$ is convex if for all $a < x_0 < x_1 < b$ and $t \in [0,1]$ $\phi(x_t) \le t\phi(x_1) + (1-t)\phi(x_0)$ where $x_t = tx_1 + (1-t)x_0$.*

Example 21.7. The functions $\exp(x)$ and $-\log(x)$ are convex and x^p is convex iff $p \ge 1$ as follows from Corollary 21.9 below which in part states that any $\phi \in C^2((a,b),\mathbb{R})$ such that $\phi'' \ge 0$ is convex.

The following Proposition is clearly motivated by Figure 21.1.

Proposition 21.8. *Suppose $\phi : (a,b) \to \mathbb{R}$ is a convex function, then*

1. *For all $u,v,w,z \in (a,b)$ such that $u < z$, $w \in [u,z)$ and $v \in (u,z]$,*
$$\frac{\phi(v)-\phi(u)}{v-u} \le \frac{\phi(z)-\phi(w)}{z-w}. \tag{21.12}$$
2. *For each $c \in (a,b)$, the right and left sided derivatives $\phi'_\pm(c)$ exists in $\mathbb{R}$ and if $a < u < v < b$, then $\phi'_+(u) \le \phi'_-(v) \le \phi'_+(v)$.*
3. *The function ϕ is continuous.*
4. *For all $t \in (a,b)$ and $\beta \in [\phi'_-(t), \phi'_+(t)]$, $\phi(x) \ge \phi(t) + \beta(x-t)$ for all $x \in (a,b)$. In particular,*
$$\phi(x) \ge \phi(t) + \phi'_-(t)(x-t) \text{ for all } x,t \in (a,b).$$

Fig. 21.1. A convex function along with two cords corresponding to $x_0 = -2$ and $x_1 = 4$ and $x_0 = -5$ and $x_1 = -2$.

Proof. 1a) Suppose first that $u < v = w < z$, in which case Eq. (21.12) is equivalent to

$$(\phi(v) - \phi(u))\,(z - v) \le (\phi(z) - \phi(v))\,(v - u)$$

which after solving for $\phi(v)$ is equivalent to the following equations holding:

$$\phi(v) \le \phi(z)\frac{v - u}{z - u} + \phi(u)\frac{z - v}{z - u}.$$

But this last equation states that $\phi(v) \le \phi(z)t + \phi(u)\,(1 - t)$ where $t = \frac{v-u}{z-u}$ and $v = tz + (1 - t)u$ and hence is valid by the definition of ϕ being convex. 1b) Now assume $u = w < v < z$, in which case Eq. (21.12) is equivalent to

$$(\phi(v) - \phi(u))\,(z - u) \le (\phi(z) - \phi(u))\,(v - u)$$

which after solving for $\phi(v)$ is equivalent to

$$\phi(v)\,(z - u) \le \phi(z)\,(v - u) + \phi(u)\,(z - v)$$

which is equivalent to

$$\phi(v) \le \phi(z)\frac{v - u}{z - u} + \phi(u)\frac{z - v}{z - u}.$$

Again this equation is valid by the convexity of ϕ. 1c) $u < w < v = z$, in which case Eq. (21.12) is equivalent to

$$(\phi(z) - \phi(u))\,(z - w) \le (\phi(z) - \phi(w))\,(z - u)$$

and this is equivalent to the inequality,

$$\phi(w) \le \phi(z)\frac{w-u}{z-u} + \phi(u)\frac{z-w}{z-u}$$

which again is true by the convexity of ϕ. 1) General case. If $u < w < v < z$, then by 1a-1c)

$$\frac{\phi(z)-\phi(w)}{z-w} \ge \frac{\phi(v)-\phi(w)}{v-w} \ge \frac{\phi(v)-\phi(u)}{v-u}$$

and if $u < v < w < z$

$$\frac{\phi(z)-\phi(w)}{z-w} \ge \frac{\phi(w)-\phi(v)}{w-v} \ge \frac{\phi(w)-\phi(u)}{w-u}.$$

We have now taken care of all possible cases. 2) On the set $a < w < z < b$, Eq. (21.12) shows that $(\phi(z)-\phi(w))\,/\,(z-w)$ is a decreasing function in w and an increasing function in z and therefore $\phi'_{\pm}(x)$ exists for all $x \in (a,b)$. Also from Eq. (21.12) we learn that

$$\phi'_+(u) \le \frac{\phi(z)-\phi(w)}{z-w} \text{ for all } a < u < w < z < b, \tag{21.13}$$

$$\frac{\phi(v)-\phi(u)}{v-u} \le \phi'_-(z) \text{ for all } a < u < v < z < b, \tag{21.14}$$

and letting $w \uparrow z$ in the first equation also implies that

$$\phi'_+(u) \le \phi'_-(z) \text{ for all } a < u < z < b.$$

The inequality, $\phi'_-(z) \le \phi'_+(z)$, is also an easy consequence of Eq. (21.12). 3) Since $\phi(x)$ has both left and right finite derivatives, it follows that ϕ is continuous. (For an alternative proof, see Rudin.) 4) Given t, let $\beta \in [\phi'_-(t), \phi'_+(t)]$, then by Eqs. (21.13) and (21.14),

$$\frac{\phi(t)-\phi(u)}{t-u} \le \phi'_-(t) \le \beta \le \phi'_+(t) \le \frac{\phi(z)-\phi(t)}{z-t}$$

for all $a < u < t < z < b$. Item 4. now follows. ∎

Corollary 21.9. *Suppose $\phi : (a,b) \to \mathbb{R}$ is differential then ϕ is convex iff ϕ' is non decreasing. In particular if $\phi \in C^2(a,b)$ then ϕ is convex iff $\phi'' \ge 0$.*

Proof. By Proposition 21.8, if ϕ is convex then ϕ' is non-decreasing. Conversely if ϕ' is increasing then by the mean value theorem,

$$\frac{\phi(x_1)-\phi(c)}{x_1-c} = \phi'(\xi_1) \text{ for some } \xi_1 \in (c, x_1)$$

and

$$\frac{\phi(c)-\phi(x_0)}{c-x_0} = \phi'(\xi_2) \text{ for some } \xi_2 \in (x_0, c).$$

Hence

$$\frac{\phi(x_1) - \phi(c)}{x_1 - c} \geq \frac{\phi(c) - \phi(x_0)}{c - x_0}$$

for all $x_0 < c < x_1$. Solving this inequality for $\phi(c)$ gives

$$\phi(c) \leq \frac{c - x_0}{x_1 - x_0}\phi(x_1) + \frac{x_1 - c}{x_1 - x_0}\phi(x_0)$$

showing ϕ is convex. ■

Theorem 21.10 (Jensen's Inequality). *Suppose that $(X, \mathcal{M}, \mu)$ is a probability space, i.e. μ is a positive measure and $\mu(X) = 1$. Also suppose that $f \in L^1(\mu)$, $f : X \to (a, b)$, and $\phi : (a, b) \to \mathbb{R}$ is a convex function. Then*

$$\phi\left(\int_X f d\mu\right) \leq \int_X \phi(f) d\mu$$

where if $\phi \circ f \notin L^1(\mu)$, then $\phi \circ f$ is integrable in the extended sense and $\int_X \phi(f) d\mu = \infty$.

Proof. Let $t = \int_X f d\mu \in (a, b)$ and let $\beta \in \mathbb{R}$ be such that $\phi(s) - \phi(t) \geq \beta(s-t)$ for all $s \in (a, b)$. Then integrating the inequality, $\phi(f) - \phi(t) \geq \beta(f-t)$, implies that

$$0 \leq \int_X \phi(f) d\mu - \phi(t) = \int_X \phi(f) d\mu - \phi(\int_X f d\mu).$$

Moreover, if $\phi(f)$ is not integrable, then $\phi(f) \geq \phi(t) + \beta(f - t)$ which shows that negative part of $\phi(f)$ is integrable. Therefore, $\int_X \phi(f) d\mu = \infty$ in this case. ■

Example 21.11. The convex functions in Example 21.7 lead to the following inequalities,

$$\exp\left(\int_X f d\mu\right) \leq \int_X e^f d\mu, \tag{21.15}$$

$$\int_X \log(|f|) d\mu \leq \log\left(\int_X |f| \, d\mu\right)$$

and for $p \geq 1$,

$$\left|\int_X f d\mu\right|^p \leq \left(\int_X |f| \, d\mu\right)^p \leq \int_X |f|^p \, d\mu.$$

The last equation may also easily be derived using Hölder's inequality. As a special case of the first equation, we get another proof of Lemma 5.5. Indeed, more generally, suppose $p_i, s_i > 0$ for $i = 1, 2, \ldots, n$ and $\sum_{i=1}^n \frac{1}{p_i} = 1$, then

$$s_1 \dots s_n = e^{\sum_{i=1}^n \ln s_i} = e^{\sum_{i=1}^n \frac{1}{p_i} \ln s_i^{p_i}} \le \sum_{i=1}^n \frac{1}{p_i} e^{\ln s_i^{p_i}} = \sum_{i=1}^n \frac{s_i^{p_i}}{p_i} \tag{21.16}$$

where the inequality follows from Eq. (21.15) with $X = \{1, 2, \dots, n\}$, $\mu = \sum_{i=1}^n \frac{1}{p_i} \delta_i$ and $f(i) := \ln s_i^{p_i}$. Of course Eq. (21.16) may be proved directly using the convexity of the exponential function.

21.2 Modes of Convergence

As usual let $(X, \mathcal{M}, \mu)$ be a fixed measure space, assume $1 \le p \le \infty$ and let $\{f_n\}_{n=1}^\infty \cup \{f\}$ be a collection of complex valued measurable functions on X. We have the following notions of convergence and Cauchy sequences.

Definition 21.12. *1. $f_n \to f$ a.e. if there is a set $E \in \mathcal{M}$ such that $\mu(E) = 0$ and $\lim_{n\to\infty} 1_{E^c} f_n = 1_{E^c} f$.*
2. $f_n \to f$ in μ – measure if $\lim_{n\to\infty} \mu(|f_n - f| > \varepsilon) = 0$ for all $\varepsilon > 0$. We will abbreviate this by saying $f_n \to f$ in L^0 or by $f_n \xrightarrow{\mu} f$.
3. $f_n \to f$ in L^p iff $f \in L^p$ and $f_n \in L^p$ for all n, and $\lim_{n\to\infty} \|f_n - f\|_p = 0$.

Definition 21.13. *1. $\{f_n\}$ is a.e. Cauchy if there is a set $E \in \mathcal{M}$ such that $\mu(E) = 0$ and $\{1_{E^c} f_n\}$ is a pointwise Cauchy sequences.*
2. $\{f_n\}$ is Cauchy in μ – measure (or L^0 – Cauchy) if $\lim_{m,n\to\infty} \mu(|f_n - f_m| > \varepsilon) = 0$ for all $\varepsilon > 0$.
3. $\{f_n\}$ is Cauchy in L^p if $\lim_{m,n\to\infty} \|f_n - f_m\|_p = 0$.

Lemma 21.14 (Chebyshev's inequality again). *Let $p \in [1, \infty)$ and $f \in L^p$, then*

$$\mu(|f| \ge \varepsilon) \le \frac{1}{\varepsilon^p} \|f\|_p^p \text{ for all } \varepsilon > 0.$$

In particular if $\{f_n\} \subset L^p$ is L^p – convergent (Cauchy) then $\{f_n\}$ is also convergent (Cauchy) in measure.

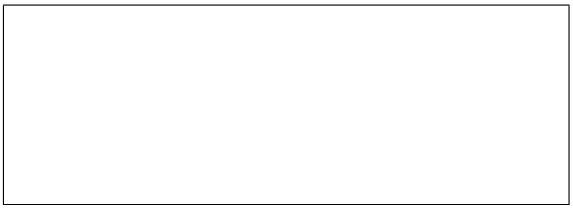
Proof. By Chebyshev's inequality (19.11),

$$\mu(|f| \ge \varepsilon) = \mu(|f|^p \ge \varepsilon^p) \le \frac{1}{\varepsilon^p} \int_X |f|^p \, d\mu = \frac{1}{\varepsilon^p} \|f\|_p^p$$

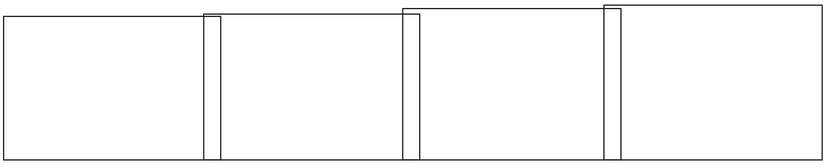
and therefore if $\{f_n\}$ is L^p – Cauchy, then

$$\mu(|f_n - f_m| \ge \varepsilon) \le \frac{1}{\varepsilon^p} \|f_n - f_m\|_p^p \to 0 \text{ as } m, n \to \infty$$

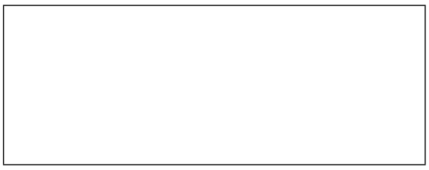
showing $\{f_n\}$ is L^0 – Cauchy. A similar argument holds for the L^p – convergent case. ■



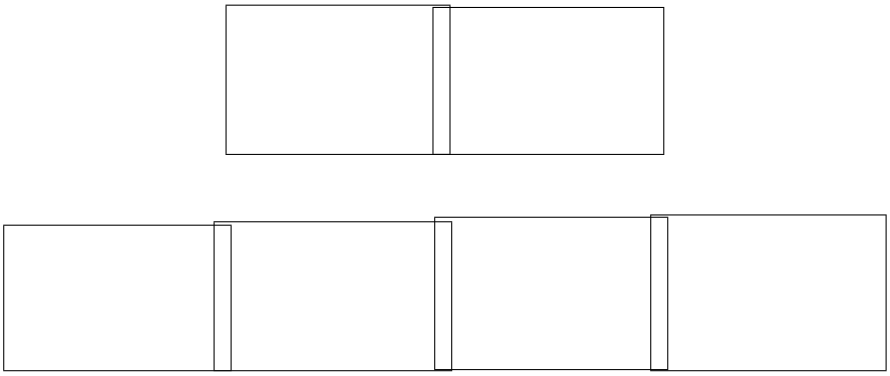
Here is a sequence of functions where $f_n \to 0$ a.e., $f_n \nrightarrow 0$ in L^1, $f_n \xrightarrow{m} 0$.



Above is a sequence of functions where $f_n \to 0$ a.e., yet $f_n \nrightarrow 0$ in L^1. or in measure.



Here is a sequence of functions where $f_n \to 0$ a.e., $f_n \xrightarrow{m} 0$ but $f_n \nrightarrow 0$ in L^1.



Above is a sequence of functions where $f_n \to 0$ in L^1, $f_n \nrightarrow 0$ a.e., and $f_n \xrightarrow{m} 0$.

Lemma 21.15. *Suppose $a_n \in \mathbb{C}$ and $|a_{n+1} - a_n| \le \varepsilon_n$ and $\sum_{n=1}^{\infty} \varepsilon_n < \infty$. Then $\lim_{n\to\infty} a_n = a \in \mathbb{C}$ exists and $|a - a_n| \le \delta_n := \sum_{k=n}^{\infty} \varepsilon_k$.*

Proof. (This is a special case of Exercise 6.9.) Let $m > n$ then

$$|a_m - a_n| = \left|\sum_{k=n}^{m-1} (a_{k+1} - a_k)\right| \le \sum_{k=n}^{m-1} |a_{k+1} - a_k| \le \sum_{k=n}^{\infty} \varepsilon_k := \delta_n . \quad (21.17)$$

So $|a_m - a_n| \le \delta_{\min(m,n)} \to 0$ as $, m, n \to \infty$, i.e. $\{a_n\}$ is Cauchy. Let $m \to \infty$ in (21.17) to find $|a - a_n| \le \delta_n$. ∎

Theorem 21.16. *Suppose $\{f_n\}$ is L^0-Cauchy. Then there exists a subsequence $g_j = f_{n_j}$ of $\{f_n\}$ such that $\lim g_j := f$ exists a.e. and $f_n \xrightarrow{\mu} f$ as $n \to \infty$. Moreover if g is a measurable function such that $f_n \xrightarrow{\mu} g$ as $n \to \infty$, then $f = g$ a.e.*

Proof. Let $\varepsilon_n > 0$ such that $\sum_{n=1}^{\infty} \varepsilon_n < \infty$ ($\varepsilon_n = 2^{-n}$ would do) and set $\delta_n = \sum_{k=n}^{\infty} \varepsilon_k$. Choose $g_j = f_{n_j}$ such that $\{n_j\}$ is a subsequence of $\mathbb{N}$ and

$$\mu(\{|g_{j+1} - g_j| > \varepsilon_j\}) \le \varepsilon_j .$$

Let $E_j = \{|g_{j+1} - g_j| > \varepsilon_j\}$,

$$F_N = \bigcup_{j=N}^{\infty} E_j = \bigcup_{j=N}^{\infty} \{|g_{j+1} - g_j| > \varepsilon_j\}$$

and

$$E := \bigcap_{N=1}^{\infty} F_N = \bigcap_{N=1}^{\infty} \bigcup_{j=N}^{\infty} E_j = \{|g_{j+1} - g_j| > \varepsilon_j \text{ i.o.}\}.$$

Then $\mu(E) = 0$ by Lemma 19.20 or the computation

$$\mu(E) \le \sum_{j=N}^{\infty} \mu(E_j) \le \sum_{j=N}^{\infty} \varepsilon_j = \delta_N \to 0 \text{ as } N \to \infty.$$

If $x \notin F_N$, i.e. $|g_{j+1}(x) - g_j(x)| \le \varepsilon_j$ for all $j \ge N$, then by Lemma 21.15, $f(x) = \lim_{j\to\infty} g_j(x)$ exists and $|f(x) - g_j(x)| \le \delta_j$ for all $j \ge N$. Therefore, since $E^c = \bigcup_{N=1}^{\infty} F_N^c$, $\lim_{j\to\infty} g_j(x) = f(x)$ exists for all $x \notin E$. Moreover, $\{x : |f(x) - g_j(x)| > \delta_j\} \subset F_j$ for all $j \ge N$ and hence

$$\mu(|f - g_j| > \delta_j) \le \mu(F_j) \le \delta_j \to 0 \text{ as } j \to \infty.$$

Therefore $g_j \xrightarrow{\mu} f$ as $j \to \infty$. Since

$$\begin{aligned}\{|f_n - f| > \varepsilon\} &= \{|f - g_j + g_j - f_n| > \varepsilon\} \\ &\subset \{|f - g_j| > \varepsilon/2\} \cup \{|g_j - f_n| > \varepsilon/2\},\end{aligned}$$

$$\mu(\{|f_n - f| > \varepsilon\}) \le \mu(\{|f - g_j| > \varepsilon/2\}) + \mu(|g_j - f_n| > \varepsilon/2)$$

and

$$\mu(\{|f_n - f| > \varepsilon\}) \le \lim_{j\to\infty} \sup \mu(|g_j - f_n| > \varepsilon/2) \to 0 \text{ as } n \to \infty.$$

If there is another function g such that $f_n \overset{\mu}{\to} g$ as $n \to \infty$, then arguing as above

$$\mu(|f - g| > \varepsilon) \le \mu(\{|f - f_n| > \varepsilon/2\}) + \mu(|g - f_n| > \varepsilon/2) \to 0 \text{ as } n \to \infty.$$

Hence

$$\mu(|f - g| > 0) = \mu(\cup_{n=1}^\infty \{|f - g| > \frac{1}{n}\}) \le \sum_{n=1}^\infty \mu(|f - g| > \frac{1}{n}) = 0,$$

i.e. $f = g$ a.e. ■

Corollary 21.17 (Dominated Convergence Theorem). *Suppose* $\{f_n\}$, $\{g_n\}$, *and* g *are in* L^1 *and* $f \in L^0$ *are functions such that*

$$|f_n| \le g_n \text{ a.e., } f_n \xrightarrow{\mu} f, \; g_n \xrightarrow{\mu} g, \text{ and } \int g_n \to \int g \text{ as } n \to \infty.$$

Then $f \in L^1$ *and* $\lim_{n\to\infty} \|f - f_n\|_1 = 0$, *i.e.* $f_n \to f$ *in* L^1. *In particular* $\lim_{n\to\infty} \int f_n = \int f$.

Proof. First notice that $|f| \le g$ a.e. and hence $f \in L^1$ since $g \in L^1$. To see that $|f| \le g$, use Theorem 21.16 to find subsequences $\{f_{n_k}\}$ and $\{g_{n_k}\}$ of $\{f_n\}$ and $\{g_n\}$ respectively which are almost everywhere convergent. Then

$$|f| = \lim_{k\to\infty} |f_{n_k}| \le \lim_{k\to\infty} g_{n_k} = g \text{ a.e.}$$

If (for sake of contradiction) $\lim_{n\to\infty} \|f - f_n\|_1 \ne 0$ there exists $\varepsilon > 0$ and a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ such that

$$\int |f - f_{n_k}| \ge \varepsilon \text{ for all } k. \tag{21.18}$$

Using Theorem 21.16 again, we may assume (by passing to a further subsequences if necessary) that $f_{n_k} \to f$ and $g_{n_k} \to g$ almost everywhere. Noting, $|f - f_{n_k}| \le g + g_{n_k} \to 2g$ and $\int (g + g_{n_k}) \to \int 2g$, an application of the dominated convergence Theorem 19.38 implies $\lim_{k\to\infty} \int |f - f_{n_k}| = 0$ which contradicts Eq. (21.18). ■

Exercise 21.1 (Fatou's Lemma). If $f_n \ge 0$ and $f_n \to f$ in measure, then $\int f \le \liminf_{n\to\infty} \int f_n$.

Theorem 21.18 (Egoroff's Theorem). *Suppose $\mu(X) < \infty$ and $f_n \to f$ a.e. Then for all $\varepsilon > 0$ there exists $E \in \mathcal{M}$ such that $\mu(E) < \varepsilon$ and $f_n \to f$ uniformly on E^c. In particular $f_n \xrightarrow{\mu} f$ as $n \to \infty$.*

Proof. Let $f_n \to f$ a.e. Then $\mu(\{|f_n - f| > \frac{1}{k} \text{ i.o. } n\}) = 0$ for all $k > 0$, i.e.

$$\lim_{N\to\infty} \mu\left(\bigcup_{n\geq N} \{|f_n - f| > \frac{1}{k}\}\right) = \mu\left(\bigcap_{N=1}^{\infty} \bigcup_{n\geq N} \{|f_n - f| > \frac{1}{k}\}\right) = 0.$$

Let $E_k := \bigcup_{n\geq N_k} \{|f_n - f| > \frac{1}{k}\}$ and choose an increasing sequence $\{N_k\}_{k=1}^{\infty}$ such that $\mu(E_k) < \varepsilon 2^{-k}$ for all k. Setting $E := \cup E_k$, $\mu(E) < \sum_k \varepsilon 2^{-k} = \varepsilon$ and if $x \notin E$, then $|f_n - f| \leq \frac{1}{k}$ for all $n \geq N_k$ and all k. That is $f_n \to f$ uniformly on E^c. ■

Exercise 21.2. Show that Egoroff's Theorem remains valid when the assumption $\mu(X) < \infty$ is replaced by the assumption that $|f_n| \leq g \in L^1$ for all n. **Hint:** make use of Theorem 21.18 applied to $f_n|_{X_k}$ where $X_k := \left\{|g| \geq k^{-1}\right\}$.

21.3 Completeness of L^p – spaces

Theorem 21.19. *Let $\|\cdot\|_\infty$ be as defined in Eq. (21.2), then $(L^\infty(X, \mathcal{M}, \mu), \|\cdot\|_\infty)$ is a Banach space. A sequence $\{f_n\}_{n=1}^{\infty} \subset L^\infty$ converges to $f \in L^\infty$ iff there exists $E \in \mathcal{M}$ such that $\mu(E) = 0$ and $f_n \to f$ uniformly on E^c. Moreover, bounded simple functions are dense in L^∞.*

Proof. By Minkowski's Theorem 21.4, $\|\cdot\|_\infty$ satisfies the triangle inequality. The reader may easily check the remaining conditions that ensure $\|\cdot\|_\infty$ is a norm. Suppose that $\{f_n\}_{n=1}^{\infty} \subset L^\infty$ is a sequence such $f_n \to f \in L^\infty$, i.e. $\|f - f_n\|_\infty \to 0$ as $n \to \infty$. Then for all $k \in \mathbb{N}$, there exists $N_k < \infty$ such that

$$\mu\left(|f - f_n| > k^{-1}\right) = 0 \text{ for all } n \geq N_k.$$

Let

$$E = \cup_{k=1}^{\infty} \cup_{n\geq N_k} \left\{|f - f_n| > k^{-1}\right\}.$$

Then $\mu(E) = 0$ and for $x \in E^c$, $|f(x) - f_n(x)| \leq k^{-1}$ for all $n \geq N_k$. This shows that $f_n \to f$ uniformly on E^c. Conversely, if there exists $E \in \mathcal{M}$ such that $\mu(E) = 0$ and $f_n \to f$ uniformly on E^c, then for any $\varepsilon > 0$,

$$\mu\left(|f - f_n| \geq \varepsilon\right) = \mu\left(\{|f - f_n| \geq \varepsilon\} \cap E^c\right) = 0$$

for all n sufficiently large. That is to say $\limsup_{n\to\infty} \|f - f_n\|_\infty \leq \varepsilon$ for all $\varepsilon > 0$. The density of simple functions follows from the approximation

Theorem 18.42. So the last item to prove is the completeness of L^∞ for which we will use Theorem 7.13.

Suppose that $\{f_n\}_{n=1}^\infty \subset L^\infty$ is a sequence such that $\sum_{n=1}^\infty \|f_n\|_\infty < \infty$. Let $M_n := \|f_n\|_\infty$, $E_n := \{|f_n| > M_n\}$, and $E := \cup_{n=1}^\infty E_n$ so that $\mu(E) = 0$. Then

$$\sum_{n=1}^\infty \sup_{x\in E^c} |f_n(x)| \le \sum_{n=1}^\infty M_n < \infty$$

which shows that $S_N(x) = \sum_{n=1}^N f_n(x)$ converges uniformly to $S(x) := \sum_{n=1}^\infty f_n(x)$ on E^c, i.e. $\lim_{n\to\infty} \|S - S_n\|_\infty = 0$.

Alternatively, suppose $\varepsilon_{m,n} := \|f_m - f_n\|_\infty \to 0$ as $m, n \to \infty$. Let $E_{m,n} = \{|f_n - f_m| > \varepsilon_{m,n}\}$ and $E := \cup E_{m,n}$, then $\mu(E) = 0$ and

$$\sup_{x\in E^c} |f_m(x) - f_n(x)| \le \varepsilon_{m,n} \to 0 \text{ as } m, n \to \infty.$$

Therefore, $f := \lim_{n\to\infty} f_n$ exists on E^c and the limit is uniform on E^c. Letting $f = \lim_{n\to\infty} 1_{E^c} f_n$, it then follows that $\lim_{n\to\infty} \|f_n - f\|_\infty = 0$. ■

Theorem 21.20 (Completeness of $L^p(\mu)$). *For $1 \le p \le \infty$, $L^p(\mu)$ equipped with the L^p – norm, $\|\cdot\|_p$ (see Eq. (21.1)), is a Banach space.*

Proof. By Minkowski's Theorem 21.4, $\|\cdot\|_p$ satisfies the triangle inequality. As above the reader may easily check the remaining conditions that ensure $\|\cdot\|_p$ is a norm. So we are left to prove the completeness of $L^p(\mu)$ for $1 \le p < \infty$, the case $p = \infty$ being done in Theorem 21.19.

Let $\{f_n\}_{n=1}^\infty \subset L^p(\mu)$ be a Cauchy sequence. By Chebyshev's inequality (Lemma 21.14), $\{f_n\}$ is L^0-Cauchy (i.e. Cauchy in measure) and by Theorem 21.16 there exists a subsequence $\{g_j\}$ of $\{f_n\}$ such that $g_j \to f$ a.e. By Fatou's Lemma,

$$\begin{aligned} \|g_j - f\|_p^p &= \int \lim_{k\to\infty} \inf |g_j - g_k|^p d\mu \le \lim_{k\to\infty} \inf \int |g_j - g_k|^p d\mu \\ &= \lim_{k\to\infty} \inf \|g_j - g_k\|_p^p \to 0 \text{ as } j \to \infty. \end{aligned}$$

In particular, $\|f\|_p \le \|g_j - f\|_p + \|g_j\|_p < \infty$ so the $f \in L^p$ and $g_j \xrightarrow{L^p} f$. The proof is finished because,

$$\|f_n - f\|_p \le \|f_n - g_j\|_p + \|g_j - f\|_p \to 0 \text{ as } j, n \to \infty.$$

■

The $L^p(\mu)$ – norm controls two types of behaviors of f, namely the "behavior at infinity" and the behavior of "local singularities." So in particular, if f is blows up at a point $x_0 \in X$, then locally near x_0 it is harder for f to be in $L^p(\mu)$ as p increases. On the other hand a function $f \in L^p(\mu)$ is allowed to decay at "infinity" slower and slower as p increases. With these insights in mind, we should not in general expect $L^p(\mu) \subset L^q(\mu)$ or $L^q(\mu) \subset L^p(\mu)$. However, there are two notable exceptions. (1) If $\mu(X) < \infty$, then there is no behavior at infinity to worry about and $L^q(\mu) \subset L^p(\mu)$ for all $q \le p$ as is shown in Corollary 21.21 below. (2) If μ is counting measure, i.e. $\mu(A) = \#(A)$, then all functions in $L^p(\mu)$ for any p can not blow up on a set of positive measure, so there are no local singularities. In this case $L^p(\mu) \subset L^q(\mu)$ for all $q \le p$, see Corollary 21.25 below.

Corollary 21.21. *If $\mu(X) < \infty$ and $0 < p < q \le \infty$, then $L^q(\mu) \subset L^p(\mu)$, the inclusion map is bounded and in fact*

$$\|f\|_p \le [\mu(X)]^{\left(\frac{1}{p}-\frac{1}{q}\right)} \|f\|_q .$$

Proof. Take $a \in [1, \infty]$ such that

$$\frac{1}{p} = \frac{1}{a} + \frac{1}{q}, \text{ i.e. } a = \frac{pq}{q-p}.$$

Then by Corollary 21.3,

$$\|f\|_p = \|f \cdot 1\|_p \le \|f\|_q \cdot \|1\|_a = \mu(X)^{1/a} \|f\|_q = \mu(X)^{\left(\frac{1}{p}-\frac{1}{q}\right)} \|f\|_q.$$

The reader may easily check this final formula is correct even when $q = \infty$ provided we interpret $1/p - 1/\infty$ to be $1/p$. ∎

Proposition 21.22. *Suppose that $0 < p_0 < p_1 \le \infty$, $\lambda \in (0, 1)$ and $p_\lambda \in (p_0, p_1)$ be defined by*

$$\frac{1}{p_\lambda} = \frac{1-\lambda}{p_0} + \frac{\lambda}{p_1} \tag{21.19}$$

with the interpretation that $\lambda/p_1 = 0$ if $p_1 = \infty$.[1] *Then $L^{p_\lambda} \subset L^{p_0} + L^{p_1}$, i.e. every function $f \in L^{p_\lambda}$ may be written as $f = g + h$ with $g \in L^{p_0}$ and $h \in L^{p_1}$. For $1 \le p_0 < p_1 \le \infty$ and $f \in L^{p_0} + L^{p_1}$ let*

$$\|f\| := \inf \left\{ \|g\|_{p_0} + \|h\|_{p_1} : f = g + h \right\}.$$

Then $(L^{p_0} + L^{p_1}, \|\cdot\|)$ is a Banach space and the inclusion map from L^{p_λ} to $L^{p_0} + L^{p_1}$ is bounded; in fact $\|f\| \le 2 \|f\|_{p_\lambda}$ for all $f \in L^{p_\lambda}$.

Proof. Let $M > 0$, then the local singularities of f are contained in the set $E := \{|f| > M\}$ and the behavior of f at “infinity” is solely determined by f on E^c. Hence let $g = f 1_E$ and $h = f 1_{E^c}$ so that $f = g + h$. By our earlier discussion we expect that $g \in L^{p_0}$ and $h \in L^{p_1}$ and this is the case since,

---

[1] A little algebra shows that λ may be computed in terms of p_0, p_λ and p_1 by

$$\lambda = \frac{p_0}{p_\lambda} \cdot \frac{p_1 - p_\lambda}{p_1 - p_0}.$$

$$\|g\|_{p_0}^{p_0} = \int |f|^{p_0} 1_{|f|>M} = M^{p_0} \int \left|\frac{f}{M}\right|^{p_0} 1_{|f|>M}$$
$$\leq M^{p_0} \int \left|\frac{f}{M}\right|^{p_\lambda} 1_{|f|>M} \leq M^{p_0-p_\lambda} \|f\|_{p_\lambda}^{p_\lambda} < \infty$$

and

$$\|h\|_{p_1}^{p_1} = \left\|f 1_{|f|\leq M}\right\|_{p_1}^{p_1} = \int |f|^{p_1} 1_{|f|\leq M} = M^{p_1} \int \left|\frac{f}{M}\right|^{p_1} 1_{|f|\leq M}$$
$$\leq M^{p_1} \int \left|\frac{f}{M}\right|^{p_\lambda} 1_{|f|\leq M} \leq M^{p_1-p_\lambda} \|f\|_{p_\lambda}^{p_\lambda} < \infty.$$

Moreover this shows

$$\|f\| \leq M^{1-p_\lambda/p_0} \|f\|_{p_\lambda}^{p_\lambda/p_0} + M^{1-p_\lambda/p_1} \|f\|_{p_\lambda}^{p_\lambda/p_1}.$$

Taking $M = \lambda \|f\|_{p_\lambda}$ then gives

$$\|f\| \leq \left(\lambda^{1-p_\lambda/p_0} + \lambda^{1-p_\lambda/p_1}\right) \|f\|_{p_\lambda}$$

and then taking $\lambda = 1$ shows $\|f\| \leq 2\|f\|_{p_\lambda}$. The the proof that $(L^{p_0} + L^{p_1}, \|\cdot\|)$ is a Banach space is left as Exercise 21.7 to the reader. ■

Corollary 21.23 (Interpolation of L^p – norms). *Suppose that $0 < p_0 < p_1 \leq \infty$, $\lambda \in (0,1)$ and $p_\lambda \in (p_0, p_1)$ be defined as in Eq. (21.19), then $L^{p_0} \cap L^{p_1} \subset L^{p_\lambda}$ and*

$$\|f\|_{p_\lambda} \leq \|f\|_{p_0}^{\lambda} \|f\|_{p_1}^{1-\lambda}. \tag{21.20}$$

Further assume $1 \leq p_0 < p_\lambda < p_1 \leq \infty$, and for $f \in L^{p_0} \cap L^{p_1}$ let

$$\|f\| := \|f\|_{p_0} + \|f\|_{p_1}.$$

Then $(L^{p_0} \cap L^{p_1}, \|\cdot\|)$ is a Banach space and the inclusion map of $L^{p_0} \cap L^{p_1}$ into L^{p_λ} is bounded, in fact

$$\|f\|_{p_\lambda} \leq \max\left(\lambda^{-1}, (1-\lambda)^{-1}\right)\left(\|f\|_{p_0} + \|f\|_{p_1}\right). \tag{21.21}$$

The heuristic explanation of this corollary is that if $f \in L^{p_0} \cap L^{p_1}$, then f has local singularities no worse than an L^{p_1} function and behavior at infinity no worse than an L^{p_0} function. Hence $f \in L^{p_\lambda}$ for any p_λ between p_0 and p_1.

Proof. Let λ be determined as above, $a = p_0/\lambda$ and $b = p_1/(1-\lambda)$, then by Corollary 21.3,

$$\|f\|_{p_\lambda} = \left\||f|^\lambda |f|^{1-\lambda}\right\|_{p_\lambda} \leq \left\||f|^\lambda\right\|_a \left\||f|^{1-\lambda}\right\|_b = \|f\|_{p_0}^{\lambda} \|f\|_{p_1}^{1-\lambda}.$$

It is easily checked that $\|\cdot\|$ is a norm on $L^{p_0} \cap L^{p_1}$. To show this space is complete, suppose that $\{f_n\} \subset L^{p_0} \cap L^{p_1}$ is a $\|\cdot\|$ – Cauchy sequence. Then

$\{f_n\}$ is both L^{p_0} and L^{p_1} – Cauchy. Hence there exist $f \in L^{p_0}$ and $g \in L^{p_1}$ such that $\lim_{n\to\infty} \|f - f_n\|_{p_0} = 0$ and $\lim_{n\to\infty} \|g - f_n\|_{p_\lambda} = 0$. By Chebyshev's inequality (Lemma 21.14) $f_n \to f$ and $f_n \to g$ in measure and therefore by Theorem 21.16, $f = g$ a.e. It now is clear that $\lim_{n\to\infty} \|f - f_n\| = 0$. The estimate in Eq. (21.21) is left as Exercise 21.6 to the reader. ∎

Remark 21.24. Combining Proposition 21.22 and Corollary 21.23 gives

$$L^{p_0} \cap L^{p_1} \subset L^{p_\lambda} \subset L^{p_0} + L^{p_1}$$

for $0 < p_0 < p_1 \le \infty$, $\lambda \in (0,1)$ and $p_\lambda \in (p_0, p_1)$ as in Eq. (21.19).

Corollary 21.25. *Suppose now that μ is counting measure on X. Then $L^p(\mu) \subset L^q(\mu)$ for all $0 < p < q \le \infty$ and $\|f\|_q \le \|f\|_p$.*

Proof. Suppose that $0 < p < q = \infty$, then

$$\|f\|_\infty^p = \sup\{|f(x)|^p : x \in X\} \le \sum_{x\in X} |f(x)|^p = \|f\|_p^p,$$

i.e. $\|f\|_\infty \le \|f\|_p$ for all $0 < p < \infty$. For $0 < p \le q \le \infty$, apply Corollary 21.23 with $p_0 = p$ and $p_1 = \infty$ to find

$$\|f\|_q \le \|f\|_p^{p/q} \|f\|_\infty^{1-p/q} \le \|f\|_p^{p/q} \|f\|_p^{1-p/q} = \|f\|_p.$$

∎

21.3.1 Summary:

1. Since $\mu(|f| > \varepsilon) \le \varepsilon^{-p} \|f\|_p^p$, L^p – convergence implies L^0 – convergence.
2. L^0 – convergence implies almost everywhere convergence for some subsequence.
3. If $\mu(X) < \infty$ then almost everywhere convergence implies uniform convergence off certain sets of small measure and in particular we have L^0 – convergence.
4. If $\mu(X) < \infty$, then $L^q \subset L^p$ for all $p \le q$ and L^q – convergence implies L^p – convergence.
5. $L^{p_0} \cap L^{p_1} \subset L^q \subset L^{p_0} + L^{p_1}$ for any $q \in (p_0, p_1)$.
6. If $p \le q$, then $\ell^p \subset \ell^q$ and$\|f\|_q \le \|f\|_p$.

21.4 Converse of Hölder's Inequality

Throughout this section we assume $(X, \mathcal{M}, \mu)$ is a σ – finite measure space, $q \in [1,\infty]$ and $p \in [1,\infty]$ are conjugate exponents, i.e. $p^{-1} + q^{-1} = 1$. For $g \in L^q$, let $\phi_g \in (L^p)^*$ be given by

$$\phi_g(f) = \int gf \, d\mu =: \langle g, f \rangle. \tag{21.22}$$

By Hölder's inequality

$$|\phi_g(f)| \le \int |gf| d\mu \le \|g\|_q \|f\|_p \tag{21.23}$$

which implies that

$$\|\phi_g\|_{(L^p)^*} := \sup\{|\phi_g(f)| : \|f\|_p = 1\} \le \|g\|_q. \tag{21.24}$$

Proposition 21.26 (Converse of Hölder's Inequality). *Let $(X, \mathcal{M}, \mu)$ be a σ – finite measure space and $1 \le p \le \infty$ as above. For all $g \in L^q$,*

$$\|g\|_q = \|\phi_g\|_{(L^p)^*} := \sup\left\{|\phi_g(f)| : \|f\|_p = 1\right\} \tag{21.25}$$

and for any measurable function $g : X \to \mathbb{C}$,

$$\|g\|_q = \sup\left\{\int_X |g|\, f d\mu : \|f\|_p = 1 \text{ and } f \ge 0\right\}. \tag{21.26}$$

Proof. We begin by proving Eq. (21.25). Assume first that $q < \infty$ so $p > 1$. Then

$$|\phi_g(f)| = \left|\int gf \, d\mu\right| \le \int |gf| \, d\mu \le \|g\|_q \|f\|_p$$

and equality occurs in the first inequality when $\operatorname{sgn}(gf)$ is constant a.e. while equality in the second occurs, by Theorem 21.2, when $|f|^p = c|g|^q$ for some constant $c > 0$. So let $f := \overline{\operatorname{sgn}(g)}|g|^{q/p}$ which for $p = \infty$ is to be interpreted as $f = \overline{\operatorname{sgn}(g)}$, i.e. $|g|^{q/\infty} \equiv 1$. When $p = \infty$,

$$|\phi_g(f)| = \int_X g \overline{\operatorname{sgn}(g)} d\mu = \|g\|_{L^1(\mu)} = \|g\|_1 \|f\|_\infty$$

which shows that $\|\phi_g\|_{(L^\infty)^*} \ge \|g\|_1$. If $p < \infty$, then

$$\|f\|_p^p = \int |f|^p = \int |g|^q = \|g\|_q^q$$

while

$$\phi_g(f) = \int gf d\mu = \int |g||g|^{q/p} d\mu = \int |g|^q d\mu = \|g\|_q^q.$$

Hence

$$\frac{|\phi_g(f)|}{\|f\|_p} = \frac{\|g\|_q^q}{\|g\|_q^{q/p}} = \|g\|_q^{q(1-\frac{1}{p})} = \|g\|_q.$$

This shows that $||\phi_g|| \ge \|g\|_q$ which combined with Eq. (21.24) implies Eq. (21.25).

The last case to consider is $p = 1$ and $q = \infty$. Let $M := \|g\|_\infty$ and choose $X_n \in \mathcal{M}$ such that $X_n \uparrow X$ as $n \to \infty$ and $\mu(X_n) < \infty$ for all n. For any $\varepsilon > 0$, $\mu(|g| \geq M - \varepsilon) > 0$ and $X_n \cap \{|g| \geq M - \varepsilon\} \uparrow \{|g| \geq M - \varepsilon\}$. Therefore, $\mu(X_n \cap \{|g| \geq M - \varepsilon\}) > 0$ for n sufficiently large. Let

$$f = \overline{\operatorname{sgn}(g)} 1_{X_n \cap \{|g| \geq M - \varepsilon\}},$$

then

$$\|f\|_1 = \mu(X_n \cap \{|g| \geq M - \varepsilon\}) \in (0, \infty)$$

and

$$\begin{aligned} |\phi_g(f)| &= \int_{X_n \cap \{|g| \geq M - \varepsilon\}} \overline{\operatorname{sgn}(g)} g d\mu = \int_{X_n \cap \{|g| \geq M - \varepsilon\}} |g| d\mu \\ &\geq (M - \varepsilon)\mu(X_n \cap \{|g| \geq M - \varepsilon\}) = (M - \varepsilon)\|f\|_1. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, it follows from this equation that $\|\phi_g\|_{(L^1)^*} \geq M = \|g\|_\infty$.

Now for the proof of Eq. (21.26). The key new point is that we no longer are assuming that $g \in L^q$. Let $M(g)$ denote the right member in Eq. (21.26) and set $g_n := 1_{X_n \cap \{|g| \leq n\}} g$. Then $|g_n| \uparrow |g|$ as $n \to \infty$ and it is clear that $M(g_n)$ is increasing in n. Therefore using Lemma 4.10 and the monotone convergence theorem,

$$\begin{aligned} \lim_{n\to\infty} M(g_n) = \sup_n M(g_n) &= \sup_n \sup \left\{ \int_X |g_n| f d\mu : \|f\|_p = 1 \text{ and } f \geq 0 \right\} \\ &= \sup \left\{ \sup_n \int_X |g_n| f d\mu : \|f\|_p = 1 \text{ and } f \geq 0 \right\} \\ &= \sup \left\{ \lim_{n\to\infty} \int_X |g_n| f d\mu : \|f\|_p = 1 \text{ and } f \geq 0 \right\} \\ &= \sup \left\{ \int_X |g| f d\mu : \|f\|_p = 1 \text{ and } f \geq 0 \right\} = M(g). \end{aligned}$$

Since $g_n \in L^q$ for all n and $M(g_n) = \|\phi_{g_n}\|_{(L^p)^*}$ (as you should verify), it follows from Eq. (21.25) that $M(g_n) = \|g_n\|_q$. When $q < \infty$ (by the monotone convergence theorem) and when $q = \infty$ (directly from the definitions) one learns that $\lim_{n\to\infty} \|g_n\|_q = \|g\|_q$. Combining this fact with $\lim_{n\to\infty} M(g_n) = M(g)$ just proved shows $M(g) = \|g\|_q$. ■

As an application we can derive a sweeping generalization of Minkowski's inequality. (See Reed and Simon, Vol II. Appendix IX.4 for a more thorough discussion of complex interpolation theory.)

Theorem 21.27 (Minkowski's Inequality for Integrals). *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ – finite measure spaces and $1 \leq p \leq \infty$. If f is a $\mathcal{M} \otimes \mathcal{N}$ measurable function, then $y \to \|f(\cdot, y)\|_{L^p(\mu)}$ is measurable and*

1. if f is a positive $\mathcal{M} \otimes \mathcal{N}$ measurable function, then

$$\| \int_Y f(\cdot, y) d\nu(y) \|_{L^p(\mu)} \leq \int_Y \|f(\cdot, y)\|_{L^p(\mu)} d\nu(y). \tag{21.27}$$

2. If $f : X \times Y \to \mathbb{C}$ is a $\mathcal{M} \otimes \mathcal{N}$ measurable function and $\int_Y \|f(\cdot, y)\|_{L^p(\mu)} d\nu(y) < \infty$ then

a) for μ – a.e. x, $f(x, \cdot) \in L^1(\nu)$,
b) the μ –a.e. defined function, $x \to \int_Y f(x, y) d\nu(y)$, is in $L^p(\mu)$ and
c) the bound in Eq. (21.27) holds.

Proof. For $p \in [1, \infty]$, let $F_p(y) := \|f(\cdot, y)\|_{L^p(\mu)}$. If $p \in [1, \infty)$

$$F_p(y) = \|f(\cdot, y)\|_{L^p(\mu)} = \left(\int_X |f(x, y)|^p \, d\mu(x) \right)^{1/p}$$

is a measurable function on Y by Fubini's theorem. To see that F_∞ is measurable, let $X_n \in \mathcal{M}$ such that $X_n \uparrow X$ and $\mu(X_n) < \infty$ for all n. Then by Exercise 21.5,

$$F_\infty(y) = \lim_{n \to \infty} \lim_{p \to \infty} \|f(\cdot, y) 1_{X_n}\|_{L^p(\mu)}$$

which shows that F_∞ is $(Y, \mathcal{N})$ – measurable as well. This shows that integral on the right side of Eq. (21.27) is well defined.

Now suppose that $f \geq 0$, $q = p/(p-1)$and $g \in L^q(\mu)$ such that $g \geq 0$ and $\|g\|_{L^q(\mu)} = 1$. Then by Tonelli's theorem and Hölder's inequality,

$$\begin{aligned}
\int_X \left[\int_Y f(x, y) d\nu(y) \right] g(x) d\mu(x) &= \int_Y d\nu(y) \int_X d\mu(x) f(x, y) g(x) \\
&\leq \|g\|_{L^q(\mu)} \int_Y \|f(\cdot, y)\|_{L^p(\mu)} d\nu(y) \\
&= \int_Y \|f(\cdot, y)\|_{L^p(\mu)} d\nu(y).
\end{aligned}$$

Therefore by the converse to Hölder's inequality (Proposition 21.26),

$$\begin{aligned}
&\| \int_Y f(\cdot, y) d\nu(y) \|_{L^p(\mu)} \\
&= \sup \left\{ \int_X \left[\int_Y f(x, y) d\nu(y) \right] g(x) d\mu(x) : \|g\|_{L^q(\mu)} = 1 \text{ and } g \geq 0 \right\} \\
&\leq \int_Y \|f(\cdot, y)\|_{L^p(\mu)} d\nu(y)
\end{aligned}$$

proving Eq. (21.27) in this case.

Now let $f : X \times Y \to \mathbb{C}$ be as in item 2) of the theorem. Applying the first part of the theorem to $|f|$ shows

$$\int_Y |f(x,y)|\, d\nu(y) < \infty \text{ for } \mu\text{– a.e. } x,$$

i.e. $f(x,\cdot) \in L^1(\nu)$ for the μ –a.e. x. Since $\left|\int_Y f(x,y)d\nu(y)\right| \le \int_Y |f(x,y)|\, d\nu(y)$ it follows by item 1) that

$$\left\|\int_Y f(\cdot,y)d\nu(y)\right\|_{L^p(\mu)} \le \left\|\int_Y |f(\cdot,y)|\, d\nu(y)\right\|_{L^p(\mu)} \le \int_Y \|f(\cdot,y)\|_{L^p(\mu)}\, d\nu(y).$$

Hence the function, $x \in X \to \int_Y f(x,y)d\nu(y)$, is in $L^p(\mu)$ and the bound in Eq. (21.27) holds. ■

Here is an application of Minkowski's inequality for integrals. In this theorem we will be using the convention that $x^{-1/\infty} := 1$.

Theorem 21.28 (Theorem 6.20 in Folland). *Suppose that $k : (0,\infty) \times (0,\infty) \to \mathbb{C}$ is a measurable function such that k is homogenous of degree -1, i.e. $k(\lambda x, \lambda y) = \lambda^{-1}k(x,y)$ for all $\lambda > 0$. If, for some $p \in [1,\infty]$,*

$$C_p := \int_0^\infty |k(x,1)|\, x^{-1/p}dx < \infty$$

then for $f \in L^p((0,\infty),m)$, $k(x,\cdot)f(\cdot) \in L^1((0,\infty),m)$ for m – a.e. x. Moreover, the m – a.e. defined function

$$(Kf)(x) = \int_0^\infty k(x,y)f(y)dy \tag{21.28}$$

is in $L^p((0,\infty),m)$ and

$$\|Kf\|_{L^p((0,\infty),m)} \le C_p\|f\|_{L^p((0,\infty),m)}.$$

Proof. By the homogeneity of k, $k(x,y) = x^{-1}k(1,\frac{y}{x})$. Using this relation and making the change of variables, $y = zx$, gives

$$\begin{aligned}\int_0^\infty |k(x,y)f(y)|\, dy &= \int_0^\infty x^{-1}\left|k(1,\frac{y}{x})f(y)\right| dy \\ &= \int_0^\infty x^{-1}\, |k(1,z)f(xz)|\, xdz = \int_0^\infty |k(1,z)f(xz)|\, dz.\end{aligned}$$

Since

$$\|f(\cdot\ z)\|^p_{L^p((0,\infty),m)} = \int_0^\infty |f(yz)|^p dy = \int_0^\infty |f(x)|^p\ \frac{dx}{z},$$

$$\|f(\cdot\ z)\|_{L^p((0,\infty),m)} = z^{-1/p}\|f\|_{L^p((0,\infty),m)}.$$

Using Minkowski's inequality for integrals then shows

$$\left\| \int_0^\infty |k(\cdot,y)f(y)|\,dy \right\|_{L^p((0,\infty),m)} \le \int_0^\infty |k(1,z)|\,\|f(\cdot z)\|_{L^p((0,\infty),m)}\,dz$$
$$= \|f\|_{L^p((0,\infty),m)} \int_0^\infty |k(1,z)|\,z^{-1/p}dz$$
$$= C_p\|f\|_{L^p((0,\infty),m)} < \infty.$$

This shows that Kf in Eq. (21.28) is well defined from m – a.e. x. The proof is finished by observing

$$\|Kf\|_{L^p((0,\infty),m)} \le \left\| \int_0^\infty |k(\cdot,y)f(y)|\,dy \right\|_{L^p((0,\infty),m)} \le C_p\|f\|_{L^p((0,\infty),m)}$$

for all $f \in L^p((0,\infty),m)$. ■

The following theorem is a strengthening of Proposition 21.26. It may be skipped on the first reading.

Theorem 21.29 (Converse of Hölder's Inequality II). *Assume that $(X,\mathcal{M},\mu)$ is a σ – finite measure space, $q,p \in [1,\infty]$ are conjugate exponents and let $\mathbb{S}_f$ denote the set of simple functions ϕ on X such that $\mu(\phi \ne 0) < \infty$. Let $g : X \to \mathbb{C}$ be a measurable function such that $\phi g \in L^1(\mu)$ for all $\phi \in \mathbb{S}_f$,*[2] *and define*

$$M_q(g) := \sup\left\{ \left| \int_X \phi g d\mu \right| : \phi \in \mathbb{S}_f \text{ with } \|\phi\|_p = 1 \right\}. \tag{21.29}$$

If $M_q(g) < \infty$ then $g \in L^q(\mu)$ and $M_q(g) = \|g\|_q$.

Proof. Let $X_n \in \mathcal{M}$ be sets such that $\mu(X_n) < \infty$ and $X_n \uparrow X$ as $n \uparrow \infty$. Suppose that $q = 1$ and hence $p = \infty$. Choose simple functions ϕ_n on X such that $|\phi_n| \le 1$ and $\overline{\operatorname{sgn}(g)} = \lim_{n\to\infty} \phi_n$ in the pointwise sense. Then $1_{X_m}\phi_n \in \mathbb{S}_f$ and therefore

$$\left| \int_X 1_{X_m}\phi_n g d\mu \right| \le M_q(g)$$

for all m,n. By assumption $1_{X_m}g \in L^1(\mu)$ and therefore by the dominated convergence theorem we may let $n \to \infty$ in this equation to find

$$\int_X 1_{X_m}|g|\,d\mu \le M_q(g)$$

for all m. The monotone convergence theorem then implies that

$$\int_X |g|\,d\mu = \lim_{m\to\infty} \int_X 1_{X_m}|g|\,d\mu \le M_q(g)$$

[2] This is equivalent to requiring $1_A g \in L^1(\mu)$ for all $A \in \mathcal{M}$ such that $\mu(A) < \infty$.

showing $g \in L^1(\mu)$ and $\|g\|_1 \leq M_q(g)$. Since Holder's inequality implies that $M_q(g) \leq \|g\|_1$, we have proved the theorem in case $q = 1$. For $q > 1$, we will begin by assuming that $g \in L^q(\mu)$. Since $p \in [1, \infty)$ we know that $\mathbb{S}_f$ is a dense subspace of $L^p(\mu)$ and therefore, using ϕ_g is continuous on $L^p(\mu)$,

$$M_q(g) = \sup \left\{ \left| \int_X \phi g d\mu \right| : \phi \in L^p(\mu) \text{ with } \|\phi\|_p = 1 \right\} = \|g\|_q$$

where the last equality follows by Proposition 21.26. So it remains to show that if $\phi g \in L^1$ for all $\phi \in \mathbb{S}_f$ and $M_q(g) < \infty$ then $g \in L^q(\mu)$. For $n \in \mathbb{N}$, let $g_n := 1_{X_n} 1_{|g| \leq n} g$. Then $g_n \in L^q(\mu)$, in fact $\|g_n\|_q \leq n\mu(X_n)^{1/q} < \infty$. So by the previous paragraph, $\|g_n\|_q = M_q(g_n)$ and hence

$$\begin{aligned} \|g_n\|_q &= \sup \left\{ \left| \int_X \phi 1_{X_n} 1_{|g| \leq n} g d\mu \right| : \phi \in L^p(\mu) \text{ with } \|\phi\|_p = 1 \right\} \\ &\leq M_q(g) \left\| \phi 1_{X_n} 1_{|g| \leq n} \right\|_p \leq M_q(g) \cdot 1 = M_q(g) \end{aligned}$$

wherein the second to last inequality we have made use of the definition of $M_q(g)$ and the fact that $\phi 1_{X_n} 1_{|g| \leq n} \in \mathbb{S}_f$. If $q \in (1, \infty)$, an application of the monotone convergence theorem (or Fatou's Lemma) along with the continuity of the norm, $\|\cdot\|_p$, implies

$$\|g\|_q = \lim_{n \to \infty} \|g_n\|_q \leq M_q(g) < \infty.$$

If $q = \infty$, then $\|g_n\|_\infty \leq M_q(g) < \infty$ for all n implies $|g_n| \leq M_q(g)$ a.e. which then implies that $|g| \leq M_q(g)$ a.e. since $|g| = \lim_{n \to \infty} |g_n|$. That is $g \in L^\infty(\mu)$ and $\|g\|_\infty \leq M_\infty(g)$. ∎

21.5 Uniform Integrability

This section will address the question as to what extra conditions are needed in order that an L^0 – convergent sequence is L^p – convergent.

Notation 21.30 *For $f \in L^1(\mu)$ and $E \in \mathcal{M}$, let*

$$\mu(f : E) := \int_E f d\mu.$$

and more generally if $A, B \in \mathcal{M}$ let

$$\mu(f : A, B) := \int_{A \cap B} f d\mu.$$

Lemma 21.31. *Suppose $g \in L^1(\mu)$, then for any $\varepsilon > 0$ there exist a $\delta > 0$ such that $\mu(|g| : E) < \varepsilon$ whenever $\mu(E) < \delta$.*

Proof. If the Lemma is false, there would exist $\varepsilon > 0$ and sets E_n such that $\mu(E_n) \to 0$ while $\mu(|g| : E_n) \geq \varepsilon$ for all n. Since $|1_{E_n} g| \leq |g| \in L^1$ and for any $\delta \in (0,1)$, $\mu(1_{E_n} |g| > \delta) \leq \mu(E_n) \to 0$ as $n \to \infty$, the dominated convergence theorem of Corollary 21.17 implies $\lim_{n\to\infty} \mu(|g| : E_n) = 0$. This contradicts $\mu(|g| : E_n) \geq \varepsilon$ for all n and the proof is complete. ∎

Suppose that $\{f_n\}_{n=1}^\infty$ is a sequence of measurable functions which converge in $L^1(\mu)$ to a function f. Then for $E \in \mathcal{M}$ and $n \in \mathbb{N}$,

$$|\mu(f_n : E)| \leq |\mu(f - f_n : E)| + |\mu(f : E)| \leq \|f - f_n\|_1 + |\mu(f : E)| .$$

Let $\varepsilon_N := \sup_{n>N} \|f - f_n\|_1$, then $\varepsilon_N \downarrow 0$ as $N \uparrow \infty$ and

$$\sup_n |\mu(f_n : E)| \leq \sup_{n\leq N} |\mu(f_n : E)| \vee (\varepsilon_N + |\mu(f : E)|) \leq \varepsilon_N + \mu(g_N : E), \tag{21.30}$$

where $g_N = |f| + \sum_{n=1}^N |f_n| \in L^1$. From Lemma 21.31 and Eq. (21.30) one easily concludes,

$$\forall\ \varepsilon > 0\ \exists\ \delta > 0\ \ni\ \sup_n |\mu(f_n : E)| < \varepsilon \text{ when } \mu(E) < \delta. \tag{21.31}$$

Definition 21.32. *Functions* $\{f_n\}_{n=1}^\infty \subset L^1(\mu)$ *satisfying Eq. (21.31) are said to be* uniformly integrable.

Remark 21.33. Let $\{f_n\}$ be real functions satisfying Eq. (21.31), E be a set where $\mu(E) < \delta$ and $E_n = E \cap \{f_n \geq 0\}$. Then $\mu(E_n) < \delta$ so that $\mu(f_n^+ : E) = \mu(f_n : E_n) < \varepsilon$ and similarly $\mu(f_n^- : E) < \varepsilon$. Therefore if Eq. (21.31) holds then

$$\sup_n \mu(|f_n| : E) < 2\varepsilon \text{ when } \mu(E) < \delta. \tag{21.32}$$

Similar arguments work for the complex case by looking at the real and imaginary parts of f_n. Therefore $\{f_n\}_{n=1}^\infty \subset L^1(\mu)$ is uniformly integrable iff

$$\forall\ \varepsilon > 0\ \exists\ \delta > 0\ \ni\ \sup_n \mu(|f_n| : E) < \varepsilon \text{ when } \mu(E) < \delta. \tag{21.33}$$

Lemma 21.34. *Assume that* $\mu(X) < \infty$, *then* $\{f_n\}$ *is uniformly bounded in* $L^1(\mu)$ *(i.e.* $K = \sup_n \|f_n\|_1 < \infty$*) and* $\{f_n\}$ *is uniformly integrable iff*

$$\lim_{M\to\infty} \sup_n \mu(|f_n| : |f_n| \geq M) = 0. \tag{21.34}$$

Proof. Since $\{f_n\}$ is uniformly bounded in $L^1(\mu)$, $\mu(|f_n| \geq M) \leq K/M$. So if (21.33) holds and $\varepsilon > 0$ is given, we may choose M sufficiently large so that $\mu(|f_n| \geq M) < \delta(\varepsilon)$ for all n and therefore,

$$\sup_n \mu(|f_n| : |f_n| \geq M) \leq \varepsilon.$$

Since ε is arbitrary, we concluded that Eq. (21.34) must hold. Conversely, suppose that Eq. (21.34) holds, then automatically $K = \sup_n \mu(|f_n|) < \infty$ because

$$\begin{aligned}\mu(|f_n|) &= \mu(|f_n| : |f_n| \geq M) + \mu(|f_n| : |f_n| < M) \\ &\leq \sup_n \mu(|f_n| : |f_n| \geq M) + M\mu(X) < \infty.\end{aligned}$$

Moreover,

$$\begin{aligned}\mu(|f_n| : E) &= \mu(|f_n| : |f_n| \geq M, E) + \mu(|f_n| : |f_n| < M, E) \\ &\leq \sup_n \mu(|f_n| : |f_n| \geq M) + M\mu(E).\end{aligned}$$

So given $\varepsilon > 0$ choose M so large that $\sup_n \mu(|f_n| : |f_n| \geq M) < \varepsilon/2$ and then take $\delta = \varepsilon/(2M)$. ■

Remark 21.35. It is not in general true that if $\{f_n\} \subset L^1(\mu)$ is uniformly integrable then $\sup_n \mu(|f_n|) < \infty$. For example take $X = \{*\}$ and $\mu(\{*\}) = 1$. Let $f_n(*) = n$. Since for $\delta < 1$ a set $E \subset X$ such that $\mu(E) < \delta$ is in fact the empty set, we see that Eq. (21.32) holds in this example. However, for finite measure spaces with out "atoms", for every $\delta > 0$ we may find a finite partition of X by sets $\{E_\ell\}_{\ell=1}^k$ with $\mu(E_\ell) < \delta$. Then if Eq. (21.32) holds with $2\varepsilon = 1$, then

$$\mu(|f_n|) = \sum_{\ell=1}^k \mu(|f_n| : E_\ell) \leq k$$

showing that $\mu(|f_n|) \leq k$ for all n.

The following Lemmas gives a concrete necessary and sufficient conditions for verifying a sequence of functions is uniformly bounded and uniformly integrable.

Lemma 21.36. *Suppose that $\mu(X) < \infty$, and $\Lambda \subset L^0(X)$ is a collection of functions.*

1. *If there exists a non decreasing function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{x\to\infty} \phi(x)/x = \infty$ and*
$$K := \sup_{f\in\Lambda} \mu(\phi(|f|)) < \infty \tag{21.35}$$
then
$$\lim_{M\to\infty} \sup_{f\in\Lambda} \mu\left(|f|\, 1_{|f|\geq M}\right) = 0. \tag{21.36}$$
2. *Conversely if Eq. (21.36) holds, there exists a non-decreasing continuous function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\phi(0) = 0$, $\lim_{x\to\infty} \phi(x)/x = \infty$ and Eq. (21.35) is valid.*

Proof. 1. Let ϕ be as in item 1. above and set $\varepsilon_M := \sup_{x\geq M} \frac{x}{\phi(x)} \to 0$ as $M \to \infty$ by assumption. Then for $f \in \Lambda$

$$\mu(|f| : |f| \geq M) = \mu(\frac{|f|}{\phi(|f|)}\phi(|f|) : |f| \geq M) \leq \varepsilon_M \mu(\phi(|f|) : |f| \geq M)$$
$$\leq \varepsilon_M \mu(\phi(|f|)) \leq K\varepsilon_M$$

and hence

$$\lim_{M\to\infty} \sup_{f\in\Lambda} \mu\left(|f| \, 1_{|f|\geq M}\right) \leq \lim_{M\to\infty} K\varepsilon_M = 0.$$

2. By assumption, $\varepsilon_M := \sup_{f\in\Lambda} \mu\left(|f| \, 1_{|f|\geq M}\right) \to 0$ as $M \to \infty$. Therefore we may choose $M_n \uparrow \infty$ such that

$$\sum_{n=0}^{\infty} (n+1) \, \varepsilon_{M_n} < \infty$$

where by convention $M_0 := 0$. Now define ϕ so that $\phi(0) = 0$ and

$$\phi'(x) = \sum_{n=0}^{\infty} (n+1) \, 1_{(M_n, M_{n+1}]}(x),$$

i.e.

$$\phi(x) = \int_0^x \phi'(y) dy = \sum_{n=0}^{\infty} (n+1) \left(x \wedge M_{n+1} - x \wedge M_n\right).$$

By construction ϕ is continuous, $\phi(0) = 0$, $\phi'(x)$ is increasing (so ϕ is convex) and $\phi'(x) \geq (n+1)$ for $x \geq M_n$. In particular

$$\frac{\phi(x)}{x} \geq \frac{\phi(M_n) + (n+1)x}{x} \geq n+1 \text{ for } x \geq M_n$$

from which we conclude $\lim_{x\to\infty} \phi(x)/x = \infty$. We also have $\phi'(x) \leq (n+1)$ on $[0, M_{n+1}]$ and therefore

$$\phi(x) \leq (n+1)x \text{ for } x \leq M_{n+1}.$$

So for $f \in \Lambda$,

$$\begin{aligned}
\mu\left(\phi(|f|)\right) &= \sum_{n=0}^{\infty} \mu\left(\phi(|f|) 1_{(M_n, M_{n+1}]}(|f|)\right) \\
&\leq \sum_{n=0}^{\infty} (n+1) \, \mu\left(|f| \, 1_{(M_n, M_{n+1}]}(|f|)\right) \\
&\leq \sum_{n=0}^{\infty} (n+1) \, \mu\left(|f| \, 1_{|f|\geq M_n}\right) \leq \sum_{n=0}^{\infty} (n+1) \, \varepsilon_{M_n}
\end{aligned}$$

and hence

$$\sup_{f\in\Lambda} \mu\left(\phi(|f|)\right) \leq \sum_{n=0}^{\infty} (n+1) \, \varepsilon_{M_n} < \infty.$$

■

Theorem 21.37 (Vitali Convergence Theorem). *(Folland 6.15) Suppose that $1 \le p < \infty$. A sequence $\{f_n\} \subset L^p$ is Cauchy iff*

1. *$\{f_n\}$ is L^0 – Cauchy,*
2. *$\{|f_n|^p\}$ – is uniformly integrable.*
3. *For all $\varepsilon > 0$, there exists a set $E \in \mathcal{M}$ such that $\mu(E) < \infty$ and $\int_{E^c} |f_n|^p \, d\mu < \varepsilon$ for all n. (This condition is vacuous when $\mu(X) < \infty$.)*

Proof. ($\Longrightarrow$) Suppose $\{f_n\} \subset L^p$ is Cauchy. Then (1) $\{f_n\}$ is L^0 – Cauchy by Lemma 21.14. (2) By completeness of L^p, there exists $f \in L^p$ such that $\|f_n - f\|_p \to 0$ as $n \to \infty$. By the mean value theorem,

$$||f|^p - |f_n|^p| \le p(\max(|f|, |f_n|))^{p-1} \, ||f| - |f_n|| \le p(|f| + |f_n|)^{p-1} \, ||f| - |f_n||$$

and therefore by Hölder's inequality,

$$\begin{aligned}\int ||f|^p - |f_n|^p| \, d\mu &\le p \int (|f| + |f_n|)^{p-1} \, ||f| - |f_n|| \, d\mu \le p \int (|f| + |f_n|)^{p-1} |f - f_n| d\mu \\ &\le p\|f - f_n\|_p \|(|f| + |f_n|)^{p-1}\|_q = p\| \, |f| + |f_n| \|_p^{p/q} \|f - f_n\|_p \\ &\le p(\|f\|_p + \|f_n\|_p)^{p/q} \|f - f_n\|_p\end{aligned}$$

where $q := p/(p-1)$. This shows that $\int ||f|^p - |f_n|^p| \, d\mu \to 0$ as $n \to \infty$.[3] By the remarks prior to Definition 21.32, $\{|f_n|^p\}$ is uniformly integrable. To verify (3), for $M > 0$ and $n \in \mathbb{N}$ let $E_M = \{|f| \ge M\}$ and $E_M(n) = \{|f_n| \ge M\}$. Then $\mu(E_M) \le \frac{1}{M^p} \|f\|_p^p < \infty$ and by the dominated convergence theorem,

$$\int_{E_M^c} |f|^p \, d\mu = \int |f|^p \, 1_{|f| < M} d\mu \to 0 \text{ as } M \to 0.$$

Moreover,

$$\left\|f_n 1_{E_M^c}\right\|_p \le \left\|f 1_{E_M^c}\right\|_p + \left\|(f_n - f) 1_{E_M^c}\right\|_p \le \left\|f 1_{E_M^c}\right\|_p + \|f_n - f\|_p \, . \quad (21.37)$$

So given $\varepsilon > 0$, choose N sufficiently large such that for all $n \ge N$, $\|f - f_n\|_p^p < \varepsilon$. Then choose M sufficiently small such that $\int_{E_M^c} |f|^p \, d\mu < \varepsilon$ and $\int_{E_M^c(n)} |f|^p \, d\mu < \varepsilon$ for all $n = 1, 2, \ldots, N-1$. Letting $E := E_M \cup E_M(1) \cup \cdots \cup E_M(N-1)$, we have

$$\mu(E) < \infty, \ \int_{E^c} |f_n|^p \, d\mu < \varepsilon \text{ for } n \le N - 1$$

and by Eq. (21.37)

[3] Here is an alternative proof. Let $h_n \equiv ||f_n|^p - |f|^p| \le |f_n|^p + |f|^p =: g_n \in L^1$ and $g \equiv 2|f|^p$. Then $g_n \xrightarrow{\mu} g$, $h_n \xrightarrow{\mu} 0$ and $\int g_n \to \int g$. Therefore by the dominated convergence theorem in Corollary 21.17, $\lim_{n\to\infty} \int h_n \, d\mu = 0$.

$$\int_{E^c} |f_n|^p \, d\mu < (\varepsilon^{1/p} + \varepsilon^{1/p})^p \le 2^p \varepsilon \text{ for } n \ge N.$$

Therefore we have found $E \in \mathcal{M}$ such that $\mu(E) < \infty$ and

$$\sup_n \int_{E^c} |f_n|^p \, d\mu \le 2^p \varepsilon$$

which verifies (3) since $\varepsilon > 0$ was arbitrary. ($\Longleftarrow$) Now suppose$\{f_n\} \subset L^p$ satisfies conditions (1) - (3). Let $\varepsilon > 0$, E be as in (3) and

$$A_{mn} := \{x \in E | f_m(x) - f_n(x)| \ge \varepsilon\}.$$

Then

$$\|(f_n - f_m)\, 1_{E^c}\|_p \le \|f_n 1_{E^c}\|_p + \|f_m\, 1_{E^c}\|_p < 2\varepsilon^{1/p}$$

and

$$\begin{aligned} \|f_n - f_m\|_p &= \|(f_n - f_m) 1_{E^c}\|_p + \|(f_n - f_m) 1_{E \setminus A_{mn}}\|_p \\ &\quad + \|(f_n - f_m) 1_{A_{mn}}\|_p \\ &\le \|(f_n - f_m) 1_{E \setminus A_{mn}}\|_p + \|(f_n - f_m) 1_{A_{mn}}\|_p + 2\varepsilon^{1/p}. \end{aligned} \tag{21.38}$$

Using properties (1) and (3) and $1_{E \cap \{|f_m - f_n| < \varepsilon\}} |f_m - f_n|^p \le \varepsilon^p 1_E \in L^1$, the dominated convergence theorem in Corollary 21.17 implies

$$\|(f_n - f_m)\, 1_{E \setminus A_{mn}}\|_p^p = \int 1_{E \cap \{|f_m - f_n| < \varepsilon\}}\, |f_m - f_n|^p \underset{m,n \to \infty}{\longrightarrow} 0.$$

which combined with Eq. (21.38) implies

$$\limsup_{m,n \to \infty} \|f_n - f_m\|_p \le \limsup_{m,n \to \infty} \|(f_n - f_m) 1_{A_{mn}}\|_p + 2\varepsilon^{1/p}.$$

Finally

$$\|(f_n - f_m) 1_{A_{mn}}\|_p \le \|f_n 1_{A_{mn}}\|_p + \|f_m\, 1_{A_{mn}}\|_p \le 2\delta(\varepsilon)$$

where

$$\delta(\varepsilon) := \sup_n \ \sup\{\ \|f_n\ 1_E\|_p : E \in \mathcal{M} \ni \mu(E) \le \varepsilon\}$$

By property (2), $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$. Therefore

$$\limsup_{m,n \to \infty} \|f_n - f_m\|_p \le 2\varepsilon^{1/p} + 0 + 2\delta(\varepsilon) \to 0 \text{ as } \varepsilon \downarrow 0$$

and therefore $\{f_n\}$ is L^p-Cauchy. ■

Here is another version of Vitali's Convergence Theorem.

Theorem 21.38 (Vitali Convergence Theorem). *(This is problem 9 on p. 133 in Rudin.) Assume that* $\mu(X) < \infty$, $\{f_n\}$ *is uniformly integrable,* $f_n \to f$ *a.e. and* $|f| < \infty$ *a.e., then* $f \in L^1(\mu)$ *and* $f_n \to f$ *in* $L^1(\mu)$.

Proof. Let $\varepsilon > 0$ be given and choose $\delta > 0$ as in the Eq. (21.32). Now use Egoroff's Theorem 21.18 to choose a set E^c where $\{f_n\}$ converges uniformly on E^c and $\mu(E) < \delta$. By uniform convergence on E^c, there is an integer $N < \infty$ such that $|f_n - f_m| \le 1$ on E^c for all $m, n \ge N$. Letting $m \to \infty$, we learn that

$$|f_N - f| \le 1 \text{ on } E^c.$$

Therefore $|f| \le |f_N| + 1$ on E^c and hence

$$\begin{aligned}\mu(|f|) &= \mu(|f| : E^c) + \mu(|f| : E) \\ &\le \mu(|f_N|) + \mu(X) + \mu(|f| : E).\end{aligned}$$

Now by Fatou's lemma,

$$\mu(|f| : E) \le \liminf_{n\to\infty} \mu(|f_n| : E) \le 2\varepsilon < \infty$$

by Eq. (21.32). This shows that $f \in L^1$. Finally

$$\begin{aligned}\mu(|f - f_n|) &= \mu(|f - f_n| : E^c) + \mu(|f - f_n| : E) \\ &\le \mu(|f - f_n| : E^c) + \mu(|f| + |f_n| : E) \\ &\le \mu(|f - f_n| : E^c) + 4\varepsilon\end{aligned}$$

and so by the Dominated convergence theorem we learn that

$$\limsup_{n\to\infty} \mu(|f - f_n|) \le 4\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary this completes the proof. ■

Theorem 21.39 (Vitali again). *Suppose that $f_n \to f$ in μ measure and Eq. (21.34) holds, then $f_n \to f$ in L^1.*

Proof. This could of course be proved using 21.38 after passing to subsequences to get $\{f_n\}$ to converge a.s. However I wish to give another proof. First off, by Fatou's lemma, $f \in L^1(\mu)$. Now let

$$\phi_K(x) = x1_{|x|\le K} + K1_{|x|>K}.$$

then $\phi_K(f_n) \xrightarrow{\mu} \phi_K(f)$ because $|\phi_K(f) - \phi_K(f_n)| \le |f - f_n|$ and since

$$|f - f_n| \le |f - \phi_K(f)| + |\phi_K(f) - \phi_K(f_n)| + |\phi_K(f_n) - f_n|$$

we have that

$$\begin{aligned}\mu|f - f_n| &\le \mu\,|f - \phi_K(f)| + \mu|\phi_K(f) - \phi_K(f_n)| + \mu\,|\phi_K(f_n) - f_n| \\ &= \mu(|f| : |f| \ge K) + \mu|\phi_K(f) - \phi_K(f_n)| + \mu(|f_n| : |f_n| \ge K).\end{aligned}$$

Therefore by the dominated convergence theorem

$$\limsup_{n\to\infty} \mu|f - f_n| \le \mu(|f| : |f| \ge K) + \limsup_{n\to\infty} \mu(|f_n| : |f_n| \ge K).$$

This last expression goes to zero as $K \to \infty$ by uniform integrability. ■

21.6 Exercises

Definition 21.40. *The* ***essential range*** *of* f, $\mathrm{essran}(f)$, *consists of those* $\lambda \in \mathbb{C}$ *such that* $\mu(|f-\lambda| < \varepsilon) > 0$ *for all* $\varepsilon > 0$.

Definition 21.41. *Let* (X, τ) *be a topological space and* ν *be a measure on* $\mathcal{B}_X = \sigma(\tau)$. *The* ***support*** *of* ν, $\mathrm{supp}(\nu)$, *consists of those* $x \in X$ *such that* $\nu(V) > 0$ *for all open neighborhoods,* V, *of* x.

Exercise 21.3. Let (X, τ) be a second countable topological space and ν be a measure on $\mathcal{B}_X$ – the Borel σ – algebra on X. Show

1. $\mathrm{supp}(\nu)$ is a closed set. (This is actually true on all topological spaces.)
2. $\nu(X \setminus \mathrm{supp}(\nu)) = 0$ and use this to conclude that $W := X \setminus \mathrm{supp}(\nu)$ is the largest open set in X such that $\nu(W) = 0$. **Hint:** let $\mathcal{U} \subset \tau$ be a countable base for the topology τ. Show that W may be written as a union of elements from $V \in \mathcal{V}$ with the property that $\mu(V) = 0$.

Exercise 21.4. Prove the following facts about $\mathrm{essran}(f)$.

1. Let $\nu = f_*\mu := \mu \circ f^{-1}$ – a Borel measure on $\mathbb{C}$. Show $\mathrm{essran}(f) = \mathrm{supp}(\nu)$.
2. $\mathrm{essran}(f)$ is a closed set and $f(x) \in \mathrm{essran}(f)$ for almost every x, i.e. $\mu(f \notin \mathrm{essran}(f)) = 0$.
3. If $F \subset \mathbb{C}$ is a closed set such that $f(x) \in F$ for almost every x then $\mathrm{essran}(f) \subset F$. So $\mathrm{essran}(f)$ is the smallest closed set F such that $f(x) \in F$ for almost every x.
4. $\|f\|_\infty = \sup\{|\lambda| : \lambda \in \mathrm{essran}(f)\}$.

Exercise 21.5. Let $f \in L^p \cap L^\infty$ for some $p < \infty$. Show $\|f\|_\infty = \lim_{q\to\infty} \|f\|_q$. If we further assume $\mu(X) < \infty$, show $\|f\|_\infty = \lim_{q\to\infty} \|f\|_q$ for all measurable functions $f : X \to \mathbb{C}$. In particular, $f \in L^\infty$ iff $\lim_{q\to\infty} \|f\|_q < \infty$. **Hints:** Use Corollary 21.23 to show $\limsup_{q\to\infty} \|f\|_q \leq \|f\|_\infty$ and to show $\liminf_{q\to\infty} \|f\|_q \geq \|f\|_\infty$, let $M < \|f\|_\infty$ and make use of Chebyshev's inequality.

Exercise 21.6. Prove Eq. (21.21) in Corollary 21.23. (Part of Folland 6.3 on p. 186.) **Hint:** Use the inequality, with $a, b \geq 1$ with $a^{-1} + b^{-1} = 1$ chosen appropriately,

$$st \leq \frac{s^a}{a} + \frac{t^b}{b},$$

(see Lemma 5.5 for Eq. (21.16)) applied to the right side of Eq. (21.20).

Exercise 21.7. Complete the proof of Proposition 21.22 by showing $(L^p + L^r, \|\cdot\|)$ is a Banach space. **Hint:** you may find using Theorem 7.13 is helpful here.

Exercise 21.8. Folland 6.5 on p. 186.

Exercise 21.9. By making the change of variables, $u = \ln x$, prove the following facts:

$$\int_0^{1/2} x^{-a} |\ln x|^b \, dx < \infty \iff a < 1 \text{ or } a = 1 \text{ and } b < -1$$
$$\int_2^{\infty} x^{-a} |\ln x|^b \, dx < \infty \iff a > 1 \text{ or } a = 1 \text{ and } b < -1$$
$$\int_0^{1} x^{-a} |\ln x|^b \, dx < \infty \iff a < 1 \text{ and } b > -1$$
$$\int_1^{\infty} x^{-a} |\ln x|^b \, dx < \infty \iff a > 1 \text{ and } b > -1.$$

Suppose $0 < p_0 < p_1 \le \infty$ and m is Lebesgue measure on $(0, \infty)$. Use the above results to manufacture a function f on $(0, \infty)$ such that $f \in L^p((0,\infty), m)$ iff (a) $p \in (p_0, p_1)$, (b) $p \in [p_0, p_1]$ and (c) $p = p_0$.

Exercise 21.10. Folland 6.9 on p. 186.

Exercise 21.11. Folland 6.10 on p. 186. Use the strong form of Theorem 19.38.

Exercise 21.12. Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ – finite measure spaces, $f \in L^2(\nu)$ and $k \in L^2(\mu \otimes \nu)$. Show

$$\int |k(x,y) f(y)| \, d\nu(y) < \infty \text{ for } \mu-\text{a.e. } x.$$

Let $Kf(x) := \int_Y k(x,y) f(y) d\nu(y)$ when the integral is defined. Show $Kf \in L^2(\mu)$ and $K : L^2(\nu) \to L^2(\mu)$ is a bounded operator with $\|K\|_{op} \le \|k\|_{L^2(\mu \otimes \nu)}$.

Exercise 21.13. Folland 6.27 on p. 196. **Hint**: Theorem 21.28.

Exercise 21.14. Folland 2.32 on p. 63.

Exercise 21.15. Folland 2.38 on p. 63.

22

Approximation Theorems and Convolutions

22.1 Density Theorems

In this section, $(X, \mathcal{M}, \mu)$ will be a measure space $\mathcal{A}$ will be a subalgebra of $\mathcal{M}$.

Notation 22.1 *Suppose $(X, \mathcal{M}, \mu)$ is a measure space and $\mathcal{A} \subset \mathcal{M}$ is a subalgebra of $\mathcal{M}$. Let $\mathbb{S}(\mathcal{A})$ denote those simple functions $\phi : X \to \mathbb{C}$ such that $\phi^{-1}(\{\lambda\}) \in \mathcal{A}$ for all $\lambda \in \mathbb{C}$ and let $\mathbb{S}_f(\mathcal{A}, \mu)$ denote those $\phi \in \mathbb{S}(\mathcal{A})$ such that $\mu(\phi \neq 0) < \infty$.*

Remark 22.2. For $\phi \in \mathbb{S}_f(\mathcal{A}, \mu)$ and $p \in [1, \infty)$, $|\phi|^p = \sum_{z \neq 0} |z|^p 1_{\{\phi = z\}}$ and hence

$$\int |\phi|^p \, d\mu = \sum_{z \neq 0} |z|^p \mu(\phi = z) < \infty \tag{22.1}$$

so that $\mathbb{S}_f(\mathcal{A}, \mu) \subset L^p(\mu)$. Conversely if $\phi \in \mathbb{S}(\mathcal{A}) \cap L^p(\mu)$, then from Eq. (22.1) it follows that $\mu(\phi = z) < \infty$ for all $z \neq 0$ and therefore $\mu(\phi \neq 0) < \infty$. Hence we have shown, for any $1 \leq p < \infty$,

$$\mathbb{S}_f(\mathcal{A}, \mu) = \mathbb{S}(\mathcal{A}) \cap L^p(\mu).$$

Lemma 22.3 (Simple Functions are Dense). *The simple functions, $\mathbb{S}_f(\mathcal{M}, \mu)$, form a dense subspace of $L^p(\mu)$ for all $1 \leq p < \infty$.*

Proof. Let $\{\phi_n\}_{n=1}^{\infty}$ be the simple functions in the approximation Theorem 18.42. Since $|\phi_n| \leq |f|$ for all n, $\phi_n \in \mathbb{S}_f(\mathcal{M}, \mu)$ and

$$|f - \phi_n|^p \leq (|f| + |\phi_n|)^p \leq 2^p |f|^p \in L^1(\mu).$$

Therefore, by the dominated convergence theorem,

$$\lim_{n\to\infty} \int |f - \phi_n|^p d\mu = \int \lim_{n\to\infty} |f - \phi_n|^p d\mu = 0.$$

■

The goal of this section is to find a number of other dense subspaces of $L^p(\mu)$ for $p \in [1, \infty)$. The next theorem is the key result of this section.

Theorem 22.4 (Density Theorem). *Let $p \in [1, \infty)$, $(X, \mathcal{M}, \mu)$ be a measure space and M be an algebra of bounded $\mathbb{F}$ – valued ($\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$) measurable functions such that*

1. *$M \subset L^p(\mu, \mathbb{F})$ and $\sigma(M) = \mathcal{M}$.*
2. *There exists $\psi_k \in M$ such that $\psi_k \to 1$ boundedly.*
3. *If $\mathbb{F} = \mathbb{C}$ we further assume that M is closed under complex conjugation.*

Then to every function $f \in L^p(\mu, \mathbb{F})$, there exists $\phi_n \in M$ such that $\lim_{n\to\infty} \|f - \phi_n\|_{L^p(\mu)} = 0$, i.e. M is dense in $L^p(\mu, \mathbb{F})$.

Proof. Fix $k \in \mathbb{N}$ for the moment and let $\mathcal{H}$ denote those bounded $\mathcal{M}$ – measurable functions, $f : X \to \mathbb{F}$, for which there exists $\{\phi_n\}_{n=1}^\infty \subset M$ such that $\lim_{n\to\infty} \|\psi_k f - \phi_n\|_{L^p(\mu)} = 0$. A routine check shows $\mathcal{H}$ is a subspace of $\ell^\infty(\mathcal{M}, \mathbb{F})$ such that $1 \in \mathcal{H}$, $M \subset \mathcal{H}$ and $\mathcal{H}$ is closed under complex conjugation if $\mathbb{F} = \mathbb{C}$. Moreover, $\mathcal{H}$ is closed under bounded convergence. To see this suppose $f_n \in \mathcal{H}$ and $f_n \to f$ boundedly. Then, by the dominated convergence theorem, $\lim_{n\to\infty} \|\psi_k (f - f_n)\|_{L^p(\mu)} = 0$.[1] (Take the dominating function to be $g = [2C |\psi_k|]^p$ where C is a constant bounding all of the $\{|f_n|\}_{n=1}^\infty$.) We may now choose $\phi_n \in M$ such that $\|\phi_n - \psi_k f_n\|_{L^p(\mu)} \le \frac{1}{n}$ then

$$\begin{aligned}\limsup_{n\to\infty} \|\psi_k f - \phi_n\|_{L^p(\mu)} &\le \limsup_{n\to\infty} \|\psi_k (f - f_n)\|_{L^p(\mu)} \\ &\quad + \limsup_{n\to\infty} \|\psi_k f_n - \phi_n\|_{L^p(\mu)} = 0 \end{aligned} \tag{22.2}$$

which implies $f \in \mathcal{H}$. An application of Dynkin's Multiplicative System Theorem 18.51 if $\mathbb{F} = \mathbb{R}$ or Theorem 18.52 if $\mathbb{F} = \mathbb{C}$ now shows $\mathcal{H}$ contains all bounded measurable functions on X.

Let $f \in L^p(\mu)$ be given. The dominated convergence theorem implies $\lim_{k\to\infty} \left\|\psi_k 1_{\{|f| \le k\}} f - f\right\|_{L^p(\mu)} = 0$. (Take the dominating function to be $g = [2C |f|]^p$ where C is a bound on all of the $|\psi_k|$.) Using this and what we have just proved, there exists $\phi_k \in M$ such that

$$\left\|\psi_k 1_{\{|f| \le k\}} f - \phi_k\right\|_{L^p(\mu)} \le \frac{1}{k}.$$

The same line of reasoning used in Eq. (22.2) now implies $\lim_{k\to\infty} \|f - \phi_k\|_{L^p(\mu)} = 0$. ■

[1] It is at this point that the proof would break down if $p = \infty$.

Definition 22.5. *Let (X,τ) be a topological space and μ be a measure on $\mathcal{B}_X = \sigma(\tau)$. A **locally integrable** function is a Borel measurable function $f : X \to \mathbb{C}$ such that $\int_K |f|\, d\mu < \infty$ for all compact subsets $K \subset X$. We will write $L^1_{loc}(\mu)$ for the space of locally integrable functions. More generally we say $f \in L^p_{loc}(\mu)$ iff $\|1_K f\|_{L^p(\mu)} < \infty$ for all compact subsets $K \subset X$.*

Definition 22.6. *Let (X,τ) be a topological space. A K**-finite measure** on X is Borel measure μ such that $\mu(K) < \infty$ for all compact subsets $K \subset X$.*

Lebesgue measure on $\mathbb{R}$ is an example of a K-finite measure while counting measure on $\mathbb{R}$ is not a K-finite measure.

Example 22.7. Suppose that μ is a K-finite measure on $\mathcal{B}_{\mathbb{R}^d}$. An application of Theorem 22.4 shows $C_c(\mathbb{R}, \mathbb{C})$ is dense in $L^p(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}, \mu; \mathbb{C})$. To apply Theorem 22.4, let $M := C_c(\mathbb{R}^d, \mathbb{C})$ and $\psi_k(x) := \psi(x/k)$ where $\psi \in C_c(\mathbb{R}^d, \mathbb{C})$ with $\psi(x) = 1$ in a neighborhood of 0. The proof is completed by showing $\sigma(M) = \sigma\left(C_c(\mathbb{R}^d, \mathbb{C})\right) = \mathcal{B}_{\mathbb{R}^d}$, which follows directly from Lemma 18.57.

We may also give a more down to earth proof as follows. Let $x_0 \in \mathbb{R}^d$, $R > 0$, $A := B(x_0, R)^c$ and $f_n(x) := d_A^{1/n}(x)$. Then $f_n \in M$ and $f_n \to 1_{B(x_0,R)}$ as $n \to \infty$ which shows $1_{B(x_0,R)}$ is $\sigma(M)$-measurable, i.e. $B(x_0, R) \in \sigma(M)$. Since $x_0 \in \mathbb{R}^d$ and $R > 0$ were arbitrary, $\sigma(M) = \mathcal{B}_{\mathbb{R}^d}$.

More generally we have the following result.

Theorem 22.8. *Let (X,τ) be a second countable locally compact Hausdorff space and $\mu : \mathcal{B}_X \to [0,\infty]$ be a K-finite measure. Then $C_c(X)$ (the space of continuous functions with compact support) is dense in $L^p(\mu)$ for all $p \in [1,\infty)$. (See also Proposition* **??** *below.)*

Proof. Let $M := C_c(X)$ and use Item 3. of Lemma 18.57 to find functions $\psi_k \in M$ such that $\psi_k \to 1$ to boundedly as $k \to \infty$. The result now follows from an application of Theorem 22.4 along with the aid of item 4. of Lemma 18.57. ∎

Exercise 22.1. Show that $BC(\mathbb{R}, \mathbb{C})$ is not dense in $L^\infty(\mathbb{R}, \mathcal{B}_\mathbb{R}, m; \mathbb{C})$. Hence the hypothesis that $p < \infty$ in Theorem 22.4 can not be removed.

Corollary 22.9. *Suppose $X \subset \mathbb{R}^n$ is an open set, $\mathcal{B}_X$ is the Borel σ – algebra on X and μ be a K-finite measure on $(X, \mathcal{B}_X)$. Then $C_c(X)$ is dense in $L^p(\mu)$ for all $p \in [1,\infty)$.*

Corollary 22.10. *Suppose that X is a compact subset of $\mathbb{R}^n$ and μ is a finite measure on $(X, \mathcal{B}_X)$, then polynomials are dense in $L^p(X,\mu)$ for all $1 \le p < \infty$.*

Proof. Consider X to be a metric space with usual metric induced from $\mathbb{R}^n$. Then X is a locally compact separable metric space and therefore

$C_c(X,\mathbb{C}) = C(X,\mathbb{C})$ is dense in $L^p(\mu)$ for all $p \in [1,\infty)$. Since, by the dominated convergence theorem, uniform convergence implies $L^p(\mu)$ – convergence, it follows from the Weierstrass approximation theorem (see Theorem 8.34 and Corollary 8.36 or Theorem 12.31 and Corollary 12.32) that polynomials are also dense in $L^p(\mu)$. ∎

Lemma 22.11. *Let (X,τ) be a second countable locally compact Hausdorff space and $\mu : \mathcal{B}_X \to [0,\infty]$ be a K-finite measure on X. If $h \in L^1_{loc}(\mu)$ is a function such that*

$$\int_X fhd\mu = 0 \text{ for all } f \in C_c(X) \tag{22.3}$$

then $h(x) = 0$ for μ – a.e. x. (See also Corollary **??** *below.)*

Proof. Let $d\nu(x) = |h(x)|\, dx$, then ν is a K-finite measure on X and hence $C_c(X)$ is dense in $L^1(\nu)$ by Theorem 22.8. Notice that

$$\int_X f \cdot \operatorname{sgn}(h) d\nu = \int_X fhd\mu = 0 \text{ for all } f \in C_c(X). \tag{22.4}$$

Let $\{K_k\}_{k=1}^\infty$ be a sequence of compact sets such that $K_k \uparrow X$ as in Lemma 11.23. Then $1_{K_k}\overline{\operatorname{sgn}(h)} \in L^1(\nu)$ and therefore there exists $f_m \in C_c(X)$ such that $f_m \to 1_{K_k}\overline{\operatorname{sgn}(h)}$ in $L^1(\nu)$. So by Eq. (22.4),

$$\nu(K_k) = \int_X 1_{K_k} d\nu = \lim_{m\to\infty} \int_X f_m \operatorname{sgn}(h) d\nu = 0.$$

Since $K_k \uparrow X$ as $k \to \infty$, $0 = \nu(X) = \int_X |h|\, d\mu$, i.e. $h(x) = 0$ for μ – a.e. x. ∎

As an application of Lemma 22.11 and Example 12.34, we will show that the Laplace transform is injective.

Theorem 22.12 (Injectivity of the Laplace Transform). *For $f \in L^1([0,\infty), dx)$, the Laplace transform of f is defined by*

$$\mathcal{L}f(\lambda) := \int_0^\infty e^{-\lambda x} f(x) dx \text{ for all } \lambda > 0.$$

If $\mathcal{L}f(\lambda) := 0$ then $f(x) = 0$ for m -a.e. x.

Proof. Suppose that $f \in L^1([0,\infty), dx)$ such that $\mathcal{L}f(\lambda) \equiv 0$. Let $g \in C_0([0,\infty), \mathbb{R})$ and $\varepsilon > 0$ be given. By Example 12.34 we may choose $\{a_\lambda\}_{\lambda>0}$ such that $\#(\{\lambda > 0 : a_\lambda \neq 0\}) < \infty$ and

$$|g(x) - \sum_{\lambda>0} a_\lambda e^{-\lambda x}| < \varepsilon \text{ for all } x \geq 0.$$

Then

$$\left|\int_0^\infty g(x)f(x)dx\right| = \left|\int_0^\infty \left(g(x) - \sum_{\lambda>0} a_\lambda e^{-\lambda x}\right) f(x)dx\right|$$
$$\leq \int_0^\infty \left|g(x) - \sum_{\lambda>0} a_\lambda e^{-\lambda x}\right| |f(x)|\, dx \leq \varepsilon \|f\|_1.$$

Since $\varepsilon > 0$ is arbitrary, it follows that $\int_0^\infty g(x)f(x)dx = 0$ for all $g \in C_0([0,\infty),\mathbb{R})$. The proof is finished by an application of Lemma 22.11. ■

Here is another variant of Theorem 22.8.

Theorem 22.13. *Let (X,d) be a metric space, τ_d be the topology on X generated by d and $\mathcal{B}_X = \sigma(\tau_d)$ be the Borel σ – algebra. Suppose $\mu : \mathcal{B}_X \to [0,\infty]$ is a measure which is σ – finite on τ_d and let $BC_f(X)$ denote the bounded continuous functions on X such that $\mu(f \neq 0) < \infty$. Then $BC_f(X)$ is a dense subspace of $L^p(\mu)$ for any $p \in [1,\infty)$.*

Proof. Let $X_k \in \tau_d$ be open sets such that $X_k \uparrow X$ and $\mu(X_k) < \infty$ and let

$$\psi_k(x) = \min(1, k \cdot d_{X_k^c}(x)) = \phi_k(d_{X_k^c}(x)),$$

see Figure 22.1 below. It is easily verified that $M := BC_f(X)$ is an algebra,

Fig. 22.1. The plot of ϕ_n for $n = 1$, 2, and 4. Notice that $\phi_n \to 1_{(0,\infty)}$.

$\psi_k \in M$ for all k and $\psi_k \to 1$ boundedly as $k \to \infty$. Given $V \in \tau$ and $k, n \in \mathbb{N}$,let

$$f_{k,n}(x) := \min(1, n \cdot d_{(V \cap X_k)^c}(x)).$$

Then $\{f_{k,n} \neq 0\} = V \cap X_k$ so $f_{k,n} \in BC_f(X)$. Moreover

$$\lim_{k\to\infty} \lim_{n\to\infty} f_{k,n} = \lim_{k\to\infty} 1_{V\cap X_k} = 1_V$$

which shows $V \in \sigma(M)$ and hence $\sigma(M) = \mathcal{B}_X$. The proof is now completed by an application of Theorem 22.4. ■

Exercise 22.2. (BRUCE: Should drop this exercise.) Suppose that (X,d) is a metric space, μ is a measure on $\mathcal{B}_X := \sigma(\tau_d)$ which is finite on bounded measurable subsets of X. Show $BC_b(X,\mathbb{R})$, defined in Eq. (19.26), is dense in $L^p(\mu)$. **Hints:** let ψ_k be as defined in Eq. (19.27) which incidentally may be used to show $\sigma(BC_b(X,\mathbb{R})) = \sigma(BC(X,\mathbb{R}))$. Then use the argument in the proof of Corollary 18.55 to show $\sigma(BC(X,\mathbb{R})) = \mathcal{B}_X$.

Theorem 22.14. *Suppose $p \in [1,\infty)$, $\mathcal{A} \subset \mathcal{M}$ is an algebra such that $\sigma(\mathcal{A}) = \mathcal{M}$ and μ is σ – finite on $\mathcal{A}$. Then $\mathbb{S}_f(\mathcal{A},\mu)$ is dense in $L^p(\mu)$. (See also Remark* **??** *below.)*

Proof. Let $M := \mathbb{S}_f(\mathcal{A},\mu)$. By assumption there exits $X_k \in \mathcal{A}$ such that $\mu(X_k) < \infty$ and $X_k \uparrow X$ as $k \to \infty$. If $A \in \mathcal{A}$, then $X_k \cap A \in \mathcal{A}$ and $\mu(X_k \cap A) < \infty$ so that $1_{X_k \cap A} \in M$. Therefore $1_A = \lim_{k\to\infty} 1_{X_k \cap A}$ is $\sigma(M)$ – measurable for every $A \in \mathcal{A}$. So we have shown that $\mathcal{A} \subset \sigma(M) \subset \mathcal{M}$ and therefore $\mathcal{M} = \sigma(\mathcal{A}) \subset \sigma(M) \subset \mathcal{M}$, i.e. $\sigma(M) = \mathcal{M}$. The theorem now follows from Theorem 22.4 after observing $\psi_k := 1_{X_k} \in M$ and $\psi_k \to 1$ boundedly. ∎

Theorem 22.15 (Separability of L^p – Spaces). *Suppose, $p \in [1,\infty)$, $\mathcal{A} \subset \mathcal{M}$ is a countable algebra such that $\sigma(\mathcal{A}) = \mathcal{M}$ and μ is σ – finite on $\mathcal{A}$. Then $L^p(\mu)$ is separable and*

$$\mathbb{D} = \{\sum a_j 1_{A_j} : a_j \in \mathbb{Q} + i\mathbb{Q},\ A_j \in \mathcal{A} \text{ with } \mu(A_j) < \infty\}$$

is a countable dense subset.

Proof. It is left to reader to check $\mathbb{D}$ is dense in $\mathbb{S}_f(\mathcal{A},\mu)$ relative to the $L^p(\mu)$ – norm. The proof is then complete since $\mathbb{S}_f(\mathcal{A},\mu)$ is a dense subspace of $L^p(\mu)$ by Theorem 22.14. ∎

Example 22.16. The collection of functions of the form $\phi = \sum_{k=1}^n c_k 1_{(a_k,b_k]}$ with $a_k, b_k \in \mathbb{Q}$ and $a_k < b_k$ are dense in $L^p(\mathbb{R},\mathcal{B}_\mathbb{R},m;\mathbb{C})$ and $L^p(\mathbb{R},\mathcal{B}_\mathbb{R},m;\mathbb{C})$ is separable for any $p \in [1,\infty)$. To prove this simply apply Theorem 22.14 with $\mathcal{A}$ being the algebra on $\mathbb{R}$ generated by the half open intervals $(a,b] \cap \mathbb{R}$ with $a < b$ and $a, b \in \mathbb{Q} \cup \{\pm\infty\}$, i.e. $\mathcal{A}$ consists of sets of the form $\coprod_{k=1}^n (a_k,b_k] \cap \mathbb{R}$, where $a_k, b_k \in \mathbb{Q} \cup \{\pm\infty\}$.

Exercise 22.3. Show $L^\infty([0,1],\mathcal{B}_\mathbb{R},m;\mathbb{C})$ is not separable. **Hint:** Suppose Γ is a dense subset of $L^\infty([0,1],\mathcal{B}_\mathbb{R},m;\mathbb{C})$ and for $\lambda \in (0,1)$, let $f_\lambda(x) := 1_{[0,\lambda]}(x)$. For each $\lambda \in (0,1)$, choose $g_\lambda \in \Gamma$ such that $\|f_\lambda - g_\lambda\|_\infty < 1/2$ and then show the map $\lambda \in (0,1) \to g_\lambda \in \Gamma$ is injective. Use this to conclude that Γ must be uncountable.

Corollary 22.17 (Riemann Lebesgue Lemma). *Suppose that $f \in L^1(\mathbb{R},m)$, then*

$$\lim_{\lambda\to\pm\infty} \int_\mathbb{R} f(x) e^{i\lambda x} dm(x) = 0.$$

Proof. By Example 22.16, given $\varepsilon > 0$ there exists $\phi = \sum_{k=1}^n c_k 1_{(a_k,b_k]}$ with $a_k, b_k \in \mathbb{R}$ such that

$$\int_{\mathbb{R}} |f - \phi| dm < \varepsilon.$$

Notice that

$$\begin{aligned}
\int_{\mathbb{R}} \phi(x) e^{i\lambda x} dm(x) &= \int_{\mathbb{R}} \sum_{k=1}^n c_k 1_{(a_k,b_k]}(x) e^{i\lambda x} dm(x) \\
&= \sum_{k=1}^n c_k \int_{a_k}^{b_k} e^{i\lambda x} dm(x) = \sum_{k=1}^n c_k \lambda^{-1} e^{i\lambda x}|_{a_k}^{b_k} \\
&= \lambda^{-1} \sum_{k=1}^n c_k \left(e^{i\lambda b_k} - e^{i\lambda a_k}\right) \to 0 \text{ as } |\lambda| \to \infty.
\end{aligned}$$

Combining these two equations with

$$\begin{aligned}
\left|\int_{\mathbb{R}} f(x) e^{i\lambda x} dm(x)\right| &\le \left|\int_{\mathbb{R}} (f(x) - \phi(x)) e^{i\lambda x} dm(x)\right| + \left|\int_{\mathbb{R}} \phi(x) e^{i\lambda x} dm(x)\right| \\
&\le \int_{\mathbb{R}} |f - \phi| dm + \left|\int_{\mathbb{R}} \phi(x) e^{i\lambda x} dm(x)\right| \\
&\le \varepsilon + \left|\int_{\mathbb{R}} \phi(x) e^{i\lambda x} dm(x)\right|
\end{aligned}$$

we learn that

$$\limsup_{|\lambda|\to\infty} \left|\int_{\mathbb{R}} f(x) e^{i\lambda x} dm(x)\right| \le \varepsilon + \limsup_{|\lambda|\to\infty} \left|\int_{\mathbb{R}} \phi(x) e^{i\lambda x} dm(x)\right| = \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this completes the proof of the Riemann Lebesgue lemma. ■

Corollary 22.18. *Suppose $\mathcal{A} \subset \mathcal{M}$ is an algebra such that $\sigma(\mathcal{A}) = \mathcal{M}$ and μ is σ – finite on $\mathcal{A}$. Then for every $B \in \mathcal{M}$ such that $\mu(B) < \infty$ and $\varepsilon > 0$ there exists $D \in \mathcal{A}$ such that $\mu(B \triangle D) < \varepsilon$. (See also Remark* **??** *below.)*

Proof. By Theorem 22.14, there exists a collection, $\{A_i\}_{i=1}^n$, of pairwise disjoint subsets of $\mathcal{A}$ and $\lambda_i \in \mathbb{R}$ such that $\int_X |1_B - f| \, d\mu < \varepsilon$ where $f = \sum_{i=1}^n \lambda_i 1_{A_i}$. Let $A_0 := X \setminus \cup_{i=1}^n A_i \in \mathcal{A}$ then

$$\begin{aligned}\int_X |1_B - f|\, d\mu &= \sum_{i=0}^{n} \int_{A_i} |1_B - f|\, d\mu \\ &= \mu(A_0 \cap B) + \sum_{i=1}^{n} \left[\int_{A_i \cap B} |1_B - \lambda_i|\, d\mu + \int_{A_i \setminus B} |1_B - \lambda_i|\, d\mu \right] \\ &= \mu(A_0 \cap B) + \sum_{i=1}^{n} \left[|1 - \lambda_i|\, \mu(B \cap A_i) + |\lambda_i|\, \mu(A_i \setminus B) \right] \qquad (22.5) \\ &\geq \mu(A_0 \cap B) + \sum_{i=1}^{n} \min\{\mu(B \cap A_i), \mu(A_i \setminus B)\} \qquad (22.6)\end{aligned}$$

where the last equality is a consequence of the fact that $1 \leq |\lambda_i| + |1 - \lambda_i|$. Let

$$\alpha_i = \begin{cases} 0 \text{ if } \mu(B \cap A_i) < \mu(A_i \setminus B) \\ 1 \text{ if } \mu(B \cap A_i) \geq \mu(A_i \setminus B) \end{cases}$$

and $g = \sum_{i=1}^{n} \alpha_i 1_{A_i} = 1_D$ where

$$D := \cup \{A_i : i > 0 \ \&\ \alpha_i = 1\} \in \mathcal{A}.$$

Equation (22.5) with λ_i replaced by α_i and f by g implies

$$\int_X |1_B - 1_D|\, d\mu = \mu(A_0 \cap B) + \sum_{i=1}^{n} \min\{\mu(B \cap A_i), \mu(A_i \setminus B)\}.$$

The latter expression, by Eq. (22.6), is bounded by $\int_X |1_B - f|\, d\mu < \varepsilon$ and therefore,

$$\mu(B \triangle D) = \int_X |1_B - 1_D|\, d\mu < \varepsilon.$$

∎

Remark 22.19. We have to assume that $\mu(B) < \infty$ as the following example shows. Let $X = \mathbb{R}$, $\mathcal{M} = \mathcal{B}$, $\mu = m$, $\mathcal{A}$ be the algebra generated by half open intervals of the form $(a, b]$, and $B = \cup_{n=1}^{\infty}(2n, 2n+1]$. It is easily checked that for every $D \in \mathcal{A}$, that $m(B \Delta D) = \infty$.

22.2 Convolution and Young's Inequalities

Throughout this section we will be solely concerned with d – dimensional Lebesgue measure, m, and we will simply write L^p for $L^p(\mathbb{R}^d, m)$.

Definition 22.20 (Convolution). *Let $f, g : \mathbb{R}^d \to \mathbb{C}$ be measurable functions. We define*

$$f * g(x) = \int_{\mathbb{R}^d} f(x - y) g(y) dy \qquad (22.7)$$

whenever the integral is defined, i.e. either $f(x-\cdot)\,g(\cdot) \in L^1(\mathbb{R}^d, m)$ *or* $f(x-\cdot)\,g(\cdot) \geq 0$. *Notice that the condition that* $f(x-\cdot)\,g(\cdot) \in L^1(\mathbb{R}^d, m)$ *is equivalent to writing* $|f| * |g|\,(x) < \infty$. *By convention, if the integral in Eq. (22.7) is not defined, let* $f * g(x) := 0$.

Notation 22.21 *Given a multi-index* $\alpha \in \mathbb{Z}_+^d$, *let* $|\alpha| = \alpha_1 + \cdots + \alpha_d$,

$$x^\alpha := \prod_{j=1}^d x_j^{\alpha_j}, \text{ and } \partial_x^\alpha = \left(\frac{\partial}{\partial x}\right)^\alpha := \prod_{j=1}^d \left(\frac{\partial}{\partial x_j}\right)^{\alpha_j}.$$

For $z \in \mathbb{R}^d$ *and* $f : \mathbb{R}^d \to \mathbb{C}$, *let* $\tau_z f : \mathbb{R}^d \to \mathbb{C}$ *be defined by* $\tau_z f(x) = f(x-z)$.

Remark 22.22 (The Significance of Convolution).

1. Suppose that $f, g \in L^1(m)$ are positive functions and let μ be the measure on $\left(\mathbb{R}^d\right)^2$ defined by
$$d\mu(x,y) := f(x)\,g(y)\,dm(x)\,dm(y).$$
Then if $h : \mathbb{R} \to [0,\infty]$ is a measurable function we have
$$\begin{aligned}\int_{(\mathbb{R}^d)^2} h(x+y)\,d\mu(x,y) &= \int_{(\mathbb{R}^d)^2} h(x+y)\,f(x)\,g(y)\,dm(x)\,dm(y)\\ &= \int_{(\mathbb{R}^d)^2} h(x)\,f(x-y)\,g(y)\,dm(x)\,dm(y)\\ &= \int_{\mathbb{R}^d} h(x)\,f*g(x)\,dm(x).\end{aligned}$$
In other words, this shows the measure $(f*g)\,m$ is the same as $S_*\mu$ where $S(x,y) := x+y$. In probability lingo, the distribution of a sum of two "independent" (i.e. product measure) random variables is the the convolution of the individual distributions.
2. Suppose that $L = \sum_{|\alpha|\leq k} a_\alpha \partial^\alpha$ is a constant coefficient differential operator and suppose that we can solve (uniquely) the equation $Lu = g$ in the form
$$u(x) = Kg(x) := \int_{\mathbb{R}^d} k(x,y)g(y)dy$$
where $k(x,y)$ is an "integral kernel." (This is a natural sort of assumption since, in view of the fundamental theorem of calculus, integration is the inverse operation to differentiation.) Since $\tau_z L = L\tau_z$ for all $z \in \mathbb{R}^d$, (this is another way to characterize constant coefficient differential operators) and $L^{-1} = K$ we should have $\tau_z K = K\tau_z$. Writing out this equation then says
$$\begin{aligned}\int_{\mathbb{R}^d} k(x-z,y)g(y)dy &= (Kg)(x-z) = \tau_z Kg(x) = (K\tau_z g)(x)\\ &= \int_{\mathbb{R}^d} k(x,y)g(y-z)dy = \int_{\mathbb{R}^d} k(x,y+z)g(y)dy.\end{aligned}$$

Since g is arbitrary we conclude that $k(x-z,y) = k(x,y+z)$. Taking $y=0$ then gives

$$k(x,z) = k(x-z,0) =: \rho(x-z).$$

We thus find that $Kg = \rho * g$. Hence we expect the convolution operation to appear naturally when solving constant coefficient partial differential equations. More about this point later.

Proposition 22.23. *Suppose $p \in [1,\infty]$, $f \in L^1$ and $g \in L^p$, then $f * g(x)$ exists for almost every x, $f * g \in L^p$ and*

$$\|f * g\|_p \le \|f\|_1 \|g\|_p .$$

Proof. This follows directly from Minkowski's inequality for integrals, Theorem 21.27. ■

Proposition 22.24. *Suppose that $p \in [1,\infty)$, then $\tau_z : L^p \to L^p$ is an isometric isomorphism and for $f \in L^p$, $z \in \mathbb{R}^d \to \tau_z f \in L^p$ is continuous.*

Proof. The assertion that $\tau_z : L^p \to L^p$ is an isometric isomorphism follows from translation invariance of Lebesgue measure and the fact that $\tau_{-z} \circ \tau_z = id$. For the continuity assertion, observe that

$$\|\tau_z f - \tau_y f\|_p = \|\tau_{-y}(\tau_z f - \tau_y f)\|_p = \|\tau_{z-y} f - f\|_p$$

from which it follows that it is enough to show $\tau_z f \to f$ in L^p as $z \to 0 \in \mathbb{R}^d$. When $f \in C_c(\mathbb{R}^d)$, $\tau_z f \to f$ uniformly and since the $K := \cup_{|z|\le 1}\mathrm{supp}(\tau_z f)$ is compact, it follows by the dominated convergence theorem that $\tau_z f \to f$ in L^p as $z \to 0 \in \mathbb{R}^d$. For general $g \in L^p$ and $f \in C_c(\mathbb{R}^d)$,

$$\begin{aligned}\|\tau_z g - g\|_p &\le \|\tau_z g - \tau_z f\|_p + \|\tau_z f - f\|_p + \|f-g\|_p \\ &= \|\tau_z f - f\|_p + 2\|f-g\|_p\end{aligned}$$

and thus

$$\limsup_{z\to 0} \|\tau_z g - g\|_p \le \limsup_{z\to 0} \|\tau_z f - f\|_p + 2\|f-g\|_p = 2\|f-g\|_p .$$

Because $C_c(\mathbb{R}^d)$ is dense in L^p, the term $\|f-g\|_p$ may be made as small as we please. ■

Exercise 22.4. Let $p \in [1,\infty]$ and $\|\tau_z - I\|_{L(L^p(m))}$ be the operator norm $\tau_z - I$. Show $\|\tau_z - I\|_{L(L^p(m))} = 2$ for all $z \in \mathbb{R}^d \setminus \{0\}$ and conclude from this that $z \in \mathbb{R}^d \to \tau_z \in L(L^p(m))$ is **not** continuous. **Hints:** 1) Show $\|\tau_z - I\|_{L(L^p(m))} = \left\|\tau_{|z|e_1} - I\right\|_{L(L^p(m))}$. 2) Let $z = te_1$ with $t > 0$ and look for $f \in L^p(m)$ such that $\tau_z f$ is approximately equal to $-f$. (In fact, if $p = \infty$, you can find $f \in L^\infty(m)$ such that $\tau_z f = -f$.) (BRUCE: add on a problem somewhere showing that $\sigma(\tau_z) = S^1 \subset \mathbb{C}$. This is very simple to prove if $p=2$ by using the Fourier transform.)

Definition 22.25. *Suppose that* (X, τ) *is a topological space and* μ *is a measure on* $\mathcal{B}_X = \sigma(\tau)$. *For a measurable function* $f : X \to \mathbb{C}$ *we define the essential support of* f *by*

$$\operatorname{supp}_\mu(f) = \{x \in X : \mu(\{y \in V : f(y) \neq 0\}) > 0 \ \forall \text{ neighborhoods } V \text{ of } x\}. \tag{22.8}$$

Equivalently, $x \notin \operatorname{supp}_\mu(f)$ *iff there exists an open neighborhood* V *of* x *such that* $1_V f = 0$ *a.e.*

It is not hard to show that if $\operatorname{supp}(\mu) = X$ (see Definition 21.41) and $f \in C(X)$ then $\operatorname{supp}_\mu(f) = \operatorname{supp}(f) := \overline{\{f \neq 0\}}$, see Exercise 22.7.

Lemma 22.26. *Suppose* (X, τ) *is second countable and* $f : X \to \mathbb{C}$ *is a measurable function and* μ *is a measure on* $\mathcal{B}_X$. *Then* $X := U \setminus \operatorname{supp}_\mu(f)$ *may be described as the largest open set* W *such that* $f 1_W(x) = 0$ *for* μ *– a.e.* x. *Equivalently put,* $C := \operatorname{supp}_\mu(f)$ *is the smallest closed subset of* X *such that* $f = f 1_C$ *a.e.*

Proof. To verify that the two descriptions of $\operatorname{supp}_\mu(f)$ are equivalent, suppose $\operatorname{supp}_\mu(f)$ is defined as in Eq. (22.8) and $W := X \setminus \operatorname{supp}_\mu(f)$. Then

$$\begin{aligned} W &= \{x \in X : \exists\ \tau \ni V \ni x \text{ such that } \mu(\{y \in V : f(y) \neq 0\}) = 0\} \\ &= \cup \{V \subset_o X : \mu(f 1_V \neq 0) = 0\} \\ &= \cup \{V \subset_o X : f 1_V = 0 \text{ for } \mu - \text{a.e.}\}. \end{aligned}$$

So to finish the argument it suffices to show $\mu(f 1_W \neq 0) = 0$. To to this let $\mathcal{U}$ be a countable base for τ and set

$$\mathcal{U}_f := \{V \in \mathcal{U} : f 1_V = 0 \text{ a.e.}\}.$$

Then it is easily seen that $W = \cup \mathcal{U}_f$ and since $\mathcal{U}_f$ is countable

$$\mu(f 1_W \neq 0) \leq \sum_{V \in \mathcal{U}_f} \mu(f 1_V \neq 0) = 0.$$

■

Lemma 22.27. *Suppose* $f, g, h : \mathbb{R}^d \to \mathbb{C}$ *are measurable functions and assume that* x *is a point in* $\mathbb{R}^d$ *such that* $|f| * |g|(x) < \infty$ *and* $|f| * (|g| * |h|)(x) < \infty$, *then*

1. $f * g(x) = g * f(x)$
2. $f * (g * h)(x) = (f * g) * h(x)$
3. *If* $z \in \mathbb{R}^d$ *and* $\tau_z(|f| * |g|)(x) = |f| * |g|(x - z) < \infty$, *then*

$$\tau_z(f * g)(x) = \tau_z f * g(x) = f * \tau_z g(x)$$

4. If $x \notin \operatorname{supp}_m(f) + \operatorname{supp}_m(g)$ *then* $f * g(x) = 0$ *and in particular,*

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

where in defining $\operatorname{supp}_m(f * g)$ *we will use the convention that "*$f * g(x) \neq 0$*" when* $|f| * |g|(x) = \infty$.

Proof. For item 1.,

$$|f| * |g|(x) = \int_{\mathbb{R}^d} |f|(x-y)|g|(y)dy = \int_{\mathbb{R}^d} |f|(y)|g|(y-x)dy = |g| * |f|(x)$$

where in the second equality we made use of the fact that Lebesgue measure invariant under the transformation $y \to x - y$. Similar computations prove all of the remaining assertions of the first three items of the lemma. Item 4. Since $f * g(x) = \tilde{f} * \tilde{g}(x)$ if $f = \tilde{f}$ and $g = \tilde{g}$ a.e. we may, by replacing f by $f 1_{\operatorname{supp}_m(f)}$ and g by $g 1_{\operatorname{supp}_m(g)}$ if necessary, assume that $\{f \neq 0\} \subset \operatorname{supp}_m(f)$ and $\{g \neq 0\} \subset \operatorname{supp}_m(g)$. So if $x \notin (\operatorname{supp}_m(f) + \operatorname{supp}_m(g))$ then $x \notin (\{f \neq 0\} + \{g \neq 0\})$ and for all $y \in \mathbb{R}^d$, either $x - y \notin \{f \neq 0\}$ or $y \notin \{g \neq 0\}$. That is to say either $x - y \in \{f = 0\}$ or $y \in \{g = 0\}$ and hence $f(x-y)g(y) = 0$ for all y and therefore $f * g(x) = 0$. This shows that $f * g = 0$ on $\mathbb{R}^d \setminus \left(\overline{\operatorname{supp}_m(f) + \operatorname{supp}_m(g)}\right)$ and therefore

$$\mathbb{R}^d \setminus \left(\overline{\operatorname{supp}_m(f) + \operatorname{supp}_m(g)}\right) \subset \mathbb{R}^d \setminus \operatorname{supp}_m(f * g),$$

i.e. $\operatorname{supp}_m(f * g) \subset \overline{\operatorname{supp}_m(f) + \operatorname{supp}_m(g)}$. ■

Remark 22.28. Let A, B be closed sets of $\mathbb{R}^d$, it is not necessarily true that $A + B$ is still closed. For example, take

$$A = \{(x, y) : x > 0 \text{ and } y \geq 1/x\} \text{ and } B = \{(x, y) : x < 0 \text{ and } y \geq 1/|x|\},$$

then every point of $A + B$ has a positive y - component and hence is not zero. On the other hand, for $x > 0$ we have $(x, 1/x) + (-x, 1/x) = (0, 2/x) \in A + B$ for all x and hence $0 \in \overline{A + B}$ showing $A + B$ is not closed. Nevertheless if one of the sets A or B is compact, then $A + B$ is closed again. Indeed, if A is compact and $x_n = a_n + b_n \in A + B$ and $x_n \to x \in \mathbb{R}^d$, then by passing to a subsequence if necessary we may assume $\lim_{n\to\infty} a_n = a \in A$ exists. In this case

$$\lim_{n\to\infty} b_n = \lim_{n\to\infty} (x_n - a_n) = x - a \in B$$

exists as well, showing $x = a + b \in A + B$.

Proposition 22.29. *Suppose that* $p, q \in [1, \infty]$ *and* p *and* q *are conjugate exponents,* $f \in L^p$ *and* $g \in L^q$*, then* $f * g \in BC(\mathbb{R}^d)$, $\|f * g\|_\infty \leq \|f\|_p \|g\|_q$ *and if* $p, q \in (1, \infty)$ *then* $f * g \in C_0(\mathbb{R}^d)$.

Proof. The existence of $f*g(x)$ and the estimate $|f*g|(x) \leq \|f\|_p \|g\|_q$ for all $x \in \mathbb{R}^d$ is a simple consequence of Holders inequality and the translation invariance of Lebesgue measure. In particular this shows $\|f*g\|_\infty \leq \|f\|_p \|g\|_q$. By relabeling p and q if necessary we may assume that $p \in [1,\infty)$. Since

$$\begin{aligned}\|\tau_z(f*g) - f*g\|_u &= \|\tau_z f * g - f*g\|_u \\ &\leq \|\tau_z f - f\|_p \|g\|_q \to 0 \text{ as } z \to 0\end{aligned}$$

it follows that $f*g$ is uniformly continuous. Finally if $p,q \in (1,\infty)$, we learn from Lemma 22.27 and what we have just proved that $f_m * g_m \in C_c(\mathbb{R}^d)$ where $f_m = f1_{|f|\leq m}$ and $g_m = g1_{|g|\leq m}$. Moreover,

$$\begin{aligned}\|f*g - f_m*g_m\|_\infty &\leq \|f*g - f_m*g\|_\infty + \|f_m*g - f_m*g_m\|_\infty \\ &\leq \|f-f_m\|_p \|g\|_q + \|f_m\|_p \|g-g_m\|_q \\ &\leq \|f-f_m\|_p \|g\|_q + \|f\|_p \|g-g_m\|_q \to 0 \text{ as } m \to \infty\end{aligned}$$

showing, with the aid of Proposition 12.23, $f*g \in C_0(\mathbb{R}^d)$. ■

Theorem 22.30 (Young's Inequality). *Let $p,q,r \in [1,\infty]$ satisfy*

$$\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}. \tag{22.9}$$

If $f \in L^p$ and $g \in L^q$ then $|f|*|g|(x) < \infty$ for m – a.e. x and

$$\|f*g\|_r \leq \|f\|_p \|g\|_q. \tag{22.10}$$

In particular L^1 is closed under convolution. (The space $(L^1,*)$ is an example of a "Banach algebra" without unit.)

Remark 22.31. Before going to the formal proof, let us first understand Eq. (22.9) by the following scaling argument. For $\lambda > 0$, let $f_\lambda(x) := f(\lambda x)$, then after a few simple change of variables we find

$$\|f_\lambda\|_p = \lambda^{-d/p}\|f\| \text{ and } (f*g)_\lambda = \lambda^d f_\lambda * g_\lambda.$$

Therefore if Eq. (22.10) holds for some $p,q,r \in [1,\infty]$, we would also have

$$\|f*g\|_r = \lambda^{d/r}\|(f*g)_\lambda\|_r \leq \lambda^{d/r}\lambda^d \|f_\lambda\|_p \|g_\lambda\|_q = \lambda^{(d+d/r-d/p-d/q)}\|f\|_p \|g\|_q$$

for all $\lambda > 0$. This is only possible if Eq. (22.9) holds.

Proof. By the usual sorts of arguments, we may assume f and g are positive functions. Let $\alpha, \beta \in [0,1]$ and $p_1, p_2 \in (0,\infty]$ satisfy $p_1^{-1} + p_2^{-1} + r^{-1} = 1$. Then by Hölder's inequality, Corollary 21.3,

$$\begin{aligned}
f * g(x) &= \int_{\mathbb{R}^d} \left[f(x-y)^{(1-\alpha)} g(y)^{(1-\beta)} \right] f(x-y)^{\alpha} g(y)^{\beta} dy \\
&\leq \left(\int_{\mathbb{R}^d} f(x-y)^{(1-\alpha)r} g(y)^{(1-\beta)r} dy \right)^{1/r} \left(\int_{\mathbb{R}^d} f(x-y)^{\alpha p_1} dy \right)^{1/p_1} \times \\
&\qquad \times \left(\int_{\mathbb{R}^d} g(y)^{\beta p_2} dy \right)^{1/p_2} \\
&= \left(\int_{\mathbb{R}^d} f(x-y)^{(1-\alpha)r} g(y)^{(1-\beta)r} dy \right)^{1/r} \|f\|_{\alpha p_1}^{\alpha} \|g\|_{\beta p_2}^{\beta}.
\end{aligned}$$

Taking the r^{th} power of this equation and integrating on x gives

$$\begin{aligned}
\|f * g\|_r^r &\leq \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(x-y)^{(1-\alpha)r} g(y)^{(1-\beta)r} dy \right) dx \cdot \|f\|_{\alpha p_1}^{\alpha} \|g\|_{\beta p_2}^{\beta} \\
&= \|f\|_{(1-\alpha)r}^{(1-\alpha)r} \|g\|_{(1-\beta)r}^{(1-\beta)r} \|f\|_{\alpha p_1}^{\alpha r} \|g\|_{\beta p_2}^{\beta r}. \qquad (22.11)
\end{aligned}$$

Let us now suppose, $(1-\alpha)r = \alpha p_1$ and $(1-\beta)r = \beta p_2$, in which case Eq. (22.11) becomes,

$$\|f * g\|_r^r \leq \|f\|_{\alpha p_1}^r \|g\|_{\beta p_2}^r$$

which is Eq. (22.10) with

$$p := (1-\alpha)r = \alpha p_1 \text{ and } q := (1-\beta)r = \beta p_2. \qquad (22.12)$$

So to finish the proof, it suffices to show p and q are arbitrary indices in $[1,\infty]$ satisfying $p^{-1} + q^{-1} = 1 + r^{-1}$. If α, β, p_1, p_2 satisfy the relations above, then

$$\alpha = \frac{r}{r+p_1} \text{ and } \beta = \frac{r}{r+p_2}$$

and

$$\begin{aligned}
\frac{1}{p} + \frac{1}{q} &= \frac{1}{\alpha p_1} + \frac{1}{\alpha p_2} = \frac{1}{p_1}\frac{r+p_1}{r} + \frac{1}{p_2}\frac{r+p_2}{r} \\
&= \frac{1}{p_1} + \frac{1}{p_2} + \frac{2}{r} = 1 + \frac{1}{r}.
\end{aligned}$$

Conversely, if p, q, r satisfy Eq. (22.9), then let α and β satisfy $p = (1-\alpha)r$ and $q = (1-\beta)r$, i.e.

$$\alpha := \frac{r-p}{r} = 1 - \frac{p}{r} \leq 1 \text{ and } \beta = \frac{r-q}{r} = 1 - \frac{q}{r} \leq 1.$$

Using Eq. (22.9) we may also express α and β as

$$\alpha = p(1 - \frac{1}{q}) \geq 0 \text{ and } \beta = q(1 - \frac{1}{p}) \geq 0$$

and in particular we have shown $\alpha, \beta \in [0,1]$. If we now define $p_1 := p/\alpha \in (0,\infty]$ and $p_2 := q/\beta \in (0,\infty]$, then

$$\begin{aligned}\frac{1}{p_1}+\frac{1}{p_2}+\frac{1}{r} &= \beta\frac{1}{q}+\alpha\frac{1}{p}+\frac{1}{r}\\ &= (1-\frac{1}{q})+(1-\frac{1}{p})+\frac{1}{r}\\ &= 2-\left(1+\frac{1}{r}\right)+\frac{1}{r}=1\end{aligned}$$

as desired. ∎

Theorem 22.32 (Approximate δ − functions). *Let $p \in [1,\infty]$, $\phi \in L^1(\mathbb{R}^d)$, $a := \int_{\mathbb{R}^d}\phi(x)dx$, and for $t>0$ let $\phi_t(x)=t^{-d}\phi(x/t)$. Then*

1. *If $f \in L^p$ with $p<\infty$ then $\phi_t * f \to af$ in L^p as $t \downarrow 0$.*
2. *If $f \in BC(\mathbb{R}^d)$ and f is uniformly continuous then $\|\phi_t * f - af\|_\infty \to 0$ as $t \downarrow 0$.*
3. *If $f \in L^\infty$ and f is continuous on $U \subset_o \mathbb{R}^d$ then $\phi_t * f \to af$ uniformly on compact subsets of U as $t \downarrow 0$.*

Proof. Making the change of variables $y = tz$ implies

$$\phi_t * f(x) = \int_{\mathbb{R}^d} f(x-y)\phi_t(y)dy = \int_{\mathbb{R}^d} f(x-tz)\phi(z)dz$$

so that

$$\begin{aligned}\phi_t * f(x) - af(x) &= \int_{\mathbb{R}^d}\left[f(x-tz)-f(x)\right]\phi(z)dz\\ &= \int_{\mathbb{R}^d}\left[\tau_{tz}f(x)-f(x)\right]\phi(z)dz. \qquad (22.13)\end{aligned}$$

Hence by Minkowski's inequality for integrals (Theorem 21.27), Proposition 22.24 and the dominated convergence theorem,

$$\|\phi_t * f - af\|_p \le \int_{\mathbb{R}^d}\|\tau_{tz}f-f\|_p\,|\phi(z)|\,dz \to 0 \text{ as } t\downarrow 0.$$

Item 2. is proved similarly. Indeed, form Eq. (22.13)

$$\|\phi_t * f - af\|_\infty \le \int_{\mathbb{R}^d}\|\tau_{tz}f-f\|_\infty\,|\phi(z)|\,dz$$

which again tends to zero by the dominated convergence theorem because $\lim_{t\downarrow 0}\|\tau_{tz}f-f\|_\infty = 0$ uniformly in z by the uniform continuity of f.

Item 3. Let $B_R = B(0,R)$ be a large ball in $\mathbb{R}^d$ and $K \sqsubset\sqsubset U$, then

$$\begin{aligned}
&\sup_{x\in K} |\phi_t * f(x) - af(x)| \\
&\quad \le \left| \int_{B_R} [f(x-tz) - f(x)]\, \phi(z) dz \right| + \left| \int_{B_R^c} [f(x-tz) - f(x)]\, \phi(z) dz \right| \\
&\quad \le \int_{B_R} |\phi(z)|\, dz \cdot \sup_{x\in K, z\in B_R} |f(x-tz) - f(x)| + 2\, \|f\|_\infty \int_{B_R^c} |\phi(z)|\, dz \\
&\quad \le \|\phi\|_1 \cdot \sup_{x\in K, z\in B_R} |f(x-tz) - f(x)| + 2\, \|f\|_\infty \int_{|z|>R} |\phi(z)|\, dz
\end{aligned}$$

so that using the uniform continuity of f on compact subsets of U,

$$\limsup_{t\downarrow 0} \sup_{x\in K} |\phi_t * f(x) - af(x)| \le 2\, \|f\|_\infty \int_{|z|>R} |\phi(z)|\, dz \to 0 \text{ as } R\to\infty.$$

■

See Theorem 8.15 of Folland for a statement about almost everywhere convergence.

Exercise 22.5. Let

$$f(t) = \begin{cases} e^{-1/t} & \text{if } t > 0 \\ 0 & \text{if } t \le 0. \end{cases}$$

Show $f \in C^\infty(\mathbb{R}, [0,1])$.

Lemma 22.33. *There exists $\phi \in C_c^\infty(\mathbb{R}^d, [0,\infty))$ such that $\phi(0) > 0$, $\operatorname{supp}(\phi) \subset \bar{B}(0,1)$ and $\int_{\mathbb{R}^d} \phi(x)dx = 1$.*

Proof. Define $h(t) = f(1-t)f(t+1)$ where f is as in Exercise 22.5. Then $h \in C_c^\infty(\mathbb{R}, [0,1])$, $\operatorname{supp}(h) \subset [-1,1]$ and $h(0) = e^{-2} > 0$. Define $c = \int_{\mathbb{R}^d} h(|x|^2)dx$. Then $\phi(x) = c^{-1}h(|x|^2)$ is the desired function. ■

The reader asked to prove the following proposition in Exercise 22.9 below.

Proposition 22.34. *Suppose that $f \in L^1_{loc}(\mathbb{R}^d, m)$ and $\phi \in C_c^1(\mathbb{R}^d)$, then $f * \phi \in C^1(\mathbb{R}^d)$ and $\partial_i(f * \phi) = f * \partial_i\phi$. Moreover if $\phi \in C_c^\infty(\mathbb{R}^d)$ then $f * \phi \in C^\infty(\mathbb{R}^d)$.*

Corollary 22.35 (C^∞ – Uryhson's Lemma). *Given $K \sqsubset\sqsubset U \subset_o \mathbb{R}^d$, there exists $f \in C_c^\infty(\mathbb{R}^d, [0,1])$ such that $\operatorname{supp}(f) \subset U$ and $f = 1$ on K.*

Proof. Let ϕ be as in Lemma 22.33, $\phi_t(x) = t^{-d}\phi(x/t)$ be as in Theorem 22.32, d be the standard metric on $\mathbb{R}^d$ and $\varepsilon = d(K, U^c)$. Since K is compact and U^c is closed, $\varepsilon > 0$. Let $V_\delta = \left\{x \in \mathbb{R}^d : d(x,K) < \delta\right\}$ and $f = \phi_{\varepsilon/3} * 1_{V_{\varepsilon/3}}$, then

$$\operatorname{supp}(f) \subset \overline{\operatorname{supp}(\phi_{\varepsilon/3}) + V_{\varepsilon/3}} \subset \bar{V}_{2\varepsilon/3} \subset U.$$

Since $\bar{V}_{2\varepsilon/3}$ is closed and bounded, $f \in C_c^\infty(U)$ and for $x \in K$,

$$f(x) = \int_{\mathbb{R}^d} 1_{d(y,K)<\varepsilon/3} \cdot \phi_{\varepsilon/3}(x-y)dy = \int_{\mathbb{R}^d} \phi_{\varepsilon/3}(x-y)dy = 1.$$

The proof will be finished after the reader (easily) verifies $0 \leq f \leq 1$. ■

Here is an application of this corollary whose proof is left to the reader, Exercise 22.10.

Lemma 22.36 (Integration by Parts). *Suppose f and g are measurable functions on $\mathbb{R}^d$ such that $t \to f(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_d)$ and $t \to g(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_d)$ are continuously differentiable functions on $\mathbb{R}$ for each fixed $x = (x_1, \dots, x_d) \in \mathbb{R}^d$. Moreover assume $f \cdot g$, $\frac{\partial f}{\partial x_i} \cdot g$ and $f \cdot \frac{\partial g}{\partial x_i}$ are in $L^1(\mathbb{R}^d, m)$. Then*

$$\int_{\mathbb{R}^d} \frac{\partial f}{\partial x_i} \cdot g dm = -\int_{\mathbb{R}^d} f \cdot \frac{\partial g}{\partial x_i} dm.$$

With this result we may give another proof of the Riemann Lebesgue Lemma.

Lemma 22.37 (Riemann Lebesgue Lemma). *For $f \in L^1(\mathbb{R}^d, m)$ let*

$$\hat{f}(\xi) := (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x) e^{-i\xi \cdot x} dm(x)$$

be the Fourier transform of f. Then $\hat{f} \in C_0(\mathbb{R}^d)$ and $\left\|\hat{f}\right\|_\infty \leq (2\pi)^{-d/2} \|f\|_1$. (The choice of the normalization factor, $(2\pi)^{-d/2}$, in $\hat{f}$ is for later convenience.)

Proof. The fact that $\hat{f}$ is continuous is a simple application of the dominated convergence theorem. Moreover,

$$\left|\hat{f}(\xi)\right| \leq \int_{\mathbb{R}^d} |f(x)| \, dm(x) \leq (2\pi)^{-d/2} \|f\|_1$$

so it only remains to see that $\hat{f}(\xi) \to 0$ as $|\xi| \to \infty$. First suppose that $f \in C_c^\infty(\mathbb{R}^d)$ and let $\Delta = \sum_{j=1}^d \frac{\partial^2}{\partial x_j^2}$ be the Laplacian on $\mathbb{R}^d$. Notice that $\frac{\partial}{\partial x_j} e^{-i\xi \cdot x} = -i\xi_j e^{-i\xi \cdot x}$ and $\Delta e^{-i\xi \cdot x} = -|\xi|^2 e^{-i\xi \cdot x}$. Using Lemma 22.36 repeatedly,

$$\begin{aligned}\int_{\mathbb{R}^d} \Delta^k f(x) e^{-i\xi \cdot x} dm(x) &= \int_{\mathbb{R}^d} f(x) \Delta_x^k e^{-i\xi \cdot x} dm(x) = -|\xi|^{2k} \int_{\mathbb{R}^d} f(x) e^{-i\xi \cdot x} dm(x) \\ &= -(2\pi)^{d/2} |\xi|^{2k} \hat{f}(\xi)\end{aligned}$$

for any $k \in \mathbb{N}$. Hence

$$(2\pi)^{d/2} \left|\hat{f}(\xi)\right| \leq |\xi|^{-2k} \left\|\Delta^k f\right\|_1 \to 0$$

as $|\xi| \to \infty$ and $\hat{f} \in C_0(\mathbb{R}^d)$. Suppose that $f \in L^1(m)$ and $f_k \in C_c^\infty(\mathbb{R}^d)$ is a sequence such that $\lim_{k\to\infty} \|f - f_k\|_1 = 0$, then $\lim_{k\to\infty} \left\|\hat{f} - \hat{f}_k\right\|_\infty = 0$. Hence $\hat{f} \in C_0(\mathbb{R}^d)$ by an application of Proposition 12.23. ■

Corollary 22.38. *Let $X \subset \mathbb{R}^d$ be an open set and μ be a K-finite measure on $\mathcal{B}_X$.*

1. *Then $C_c^\infty(X)$ is dense in $L^p(\mu)$ for all $1 \le p < \infty$.*
2. *If $h \in L^1_{loc}(\mu)$ satisfies*
$$\int_X fhd\mu = 0 \text{ for all } f \in C_c^\infty(X) \tag{22.14}$$
then $h(x) = 0$ for μ – a.e. x.

Proof. Let $f \in C_c(X)$, ϕ be as in Lemma 22.33, ϕ_t be as in Theorem 22.32 and set $\psi_t := \phi_t * (f1_X)$. Then by Proposition 22.34 $\psi_t \in C^\infty(X)$ and by Lemma 22.27 there exists a compact set $K \subset X$ such that $\text{supp}(\psi_t) \subset K$ for all t sufficiently small. By Theorem 22.32, $\psi_t \to f$ uniformly on X as $t \downarrow 0$

1. The dominated convergence theorem (with dominating function being $\|f\|_\infty 1_K$), shows $\psi_t \to f$ in $L^p(\mu)$ as $t \downarrow 0$. This proves Item 1., since Theorem 22.8 guarantees that $C_c(X)$ is dense in $L^p(\mu)$.
2. Keeping the same notation as above, the dominated convergence theorem (with dominating function being $\|f\|_\infty |h| 1_K$) implies
$$0 = \lim_{t\downarrow 0} \int_X \psi_t h d\mu = \int_X \lim_{t\downarrow 0} \psi_t h d\mu = \int_X fhd\mu.$$
The proof is now finished by an application of Lemma 22.11.

■

22.2.1 Smooth Partitions of Unity

We have the following smooth variants of Proposition 12.16, Theorem 12.18 and Corollary 12.20. The proofs of these results are the same as their continuous counterparts. One simply uses the smooth version of Urysohn's Lemma of Corollary 22.35 in place of Lemma 12.8.

Proposition 22.39 (Smooth Partitions of Unity for Compacts). *Suppose that X is an open subset of $\mathbb{R}^d$, $K \subset X$ is a compact set and $\mathcal{U} = \{U_j\}_{j=1}^n$ is an open cover of K. Then there exists a smooth (i.e. $h_j \in C^\infty(X, [0,1])$) partition of unity $\{h_j\}_{j=1}^n$ of K such that $h_j \prec U_j$ for all $j = 1, 2, \ldots, n$.*

Theorem 22.40 (Locally Compact Partitions of Unity). *Suppose that X is an open subset of $\mathbb{R}^d$ and $\mathcal{U}$ is an open cover of X. Then there exists a smooth partition of unity of $\{h_i\}_{i=1}^N$ ($N = \infty$ is allowed here) subordinate to the cover $\mathcal{U}$ such that $\text{supp}(h_i)$ is compact for all i.*

Corollary 22.41. *Suppose that X is an open subset of $\mathbb{R}^d$ and $\mathcal{U} = \{U_\alpha\}_{\alpha\in A} \subset \tau$ is an open cover of X. Then there exists a smooth partition of unity of $\{h_\alpha\}_{\alpha\in A}$ subordinate to the cover $\mathcal{U}$ such that $\text{supp}(h_\alpha) \subset U_\alpha$ for all $\alpha \in A$. Moreover if $\bar{U}_\alpha$ is compact for each $\alpha \in A$ we may choose h_α so that $h_\alpha \prec U_\alpha$.*

22.3 Exercises

Exercise 22.6. Let (X, τ) be a topological space, μ a measure on $\mathcal{B}_X = \sigma(\tau)$ and $f : X \to \mathbb{C}$ be a measurable function. Letting ν be the measure, $d\nu = |f|\, d\mu$, show $\mathrm{supp}(\nu) = \mathrm{supp}_\mu(f)$, where $\mathrm{supp}(\nu)$ is defined in Definition 21.41).

Exercise 22.7. Let (X, τ) be a topological space, μ a measure on $\mathcal{B}_X = \sigma(\tau)$ such that $\mathrm{supp}(\mu) = X$ (see Definition 21.41). Show $\mathrm{supp}_\mu(f) = \mathrm{supp}(f) = \overline{\{f \neq 0\}}$ for all $f \in C(X)$.

Exercise 22.8. Prove the following strong version of item 3. of Proposition 10.52, namely to every pair of points, $x_0,\ x_1$, in a connected open subset V of $\mathbb{R}^d$ there exists $\sigma \in C^\infty(\mathbb{R}, V)$ such that $\sigma(0) = x_0$ and $\sigma(1) = x_1$. **Hint**: First choose a continuous path $\gamma : [0,1] \to V$ such that $\gamma(t) = x_0$ for t near 0 and $\gamma(t) = x_1$ for t near 1 and then use a convolution argument to smooth γ.

Exercise 22.9. Prove Proposition 22.34 by appealing to Corollary 19.43.

Exercise 22.10 (Integration by Parts). Suppose that $(x, y) \in \mathbb{R} \times \mathbb{R}^{d-1} \to f(x, y) \in \mathbb{C}$ and $(x, y) \in \mathbb{R} \times \mathbb{R}^{d-1} \to g(x, y) \in \mathbb{C}$ are measurable functions such that for each fixed $y \in \mathbb{R}^d$, $x \to f(x, y)$ and $x \to g(x, y)$ are continuously differentiable. Also assume $f \cdot g$, $\partial_x f \cdot g$ and $f \cdot \partial_x g$ are integrable relative to Lebesgue measure on $\mathbb{R} \times \mathbb{R}^{d-1}$, where $\partial_x f(x, y) := \frac{d}{dt} f(x + t, y)|_{t=0}$. Show

$$\int_{\mathbb{R}\times\mathbb{R}^{d-1}} \partial_x f(x, y) \cdot g(x, y) dx dy = -\int_{\mathbb{R}\times\mathbb{R}^{d-1}} f(x, y) \cdot \partial_x g(x, y) dx dy. \quad (22.15)$$

(Note: this result and Fubini's theorem proves Lemma 22.36.)

Hints: Let $\psi \in C_c^\infty(\mathbb{R})$ be a function which is 1 in a neighborhood of $0 \in \mathbb{R}$ and set $\psi_\varepsilon(x) = \psi(\varepsilon x)$. First verify Eq. (22.15) with $f(x, y)$ replaced by $\psi_\varepsilon(x) f(x, y)$ by doing the x – integral first. Then use the dominated convergence theorem to prove Eq. (22.15) by passing to the limit, $\varepsilon \downarrow 0$.

Exercise 22.11. Let μ be a finite measure on $\mathcal{B}_{\mathbb{R}^d}$, then $\mathbb{D} := \mathrm{span}\{e^{i\lambda\cdot x} : \lambda \in \mathbb{R}^d\}$ is a dense subspace of $L^p(\mu)$ for all $1 \le p < \infty$. **Hints:** By Theorem 22.8, $C_c(\mathbb{R}^d)$ is a dense subspace of $L^p(\mu)$. For $f \in C_c(\mathbb{R}^d)$ and $N \in \mathbb{N}$, let

$$f_N(x) := \sum_{n\in\mathbb{Z}^d} f(x + 2\pi N n).$$

Show $f_N \in BC(\mathbb{R}^d)$ and $x \to f_N(Nx)$ is 2π – periodic, so by Exercise 12.13, $x \to f_N(Nx)$ can be approximated uniformly by trigonometric polynomials. Use this fact to conclude that $f_N \in \bar{\mathbb{D}}^{L^p(\mu)}$. After this show $f_N \to f$ in $L^p(\mu)$.

Exercise 22.12. Suppose that μ and ν are two finite measures on $\mathbb{R}^d$ such that

$$\int_{\mathbb{R}^d} e^{i\lambda\cdot x} d\mu(x) = \int_{\mathbb{R}^d} e^{i\lambda\cdot x} d\nu(x) \tag{22.16}$$

for all $\lambda \in \mathbb{R}^d$. Show $\mu = \nu$.

Hint: Perhaps the easiest way to do this is to use Exercise 22.11 with the measure μ being replaced by $\mu + \nu$. Alternatively, use the method of proof of Exercise 22.11 to show Eq. (22.16) implies $\int_{\mathbb{R}^d} f d\mu(x) = \int_{\mathbb{R}^d} f d\nu(x)$ for all $f \in C_c(\mathbb{R}^d)$ and then apply Corollary 18.58.

Exercise 22.13. Again let μ be a finite measure on $\mathcal{B}_{\mathbb{R}^d}$. Further assume that $C_M := \int_{\mathbb{R}^d} e^{M|x|} d\mu(x) < \infty$ for all $M \in (0, \infty)$. Let $\mathcal{P}(\mathbb{R}^d)$ be the space of polynomials, $\rho(x) = \sum_{|\alpha| \le N} \rho_\alpha x^\alpha$ with $\rho_\alpha \in \mathbb{C}$, on $\mathbb{R}^d$. (Notice that $|\rho(x)|^p \le Ce^{M|x|}$ for some constant $C = C(\rho, p, M)$, so that $\mathcal{P}(\mathbb{R}^d) \subset L^p(\mu)$ for all $1 \le p < \infty$.) Show $\mathcal{P}(\mathbb{R}^d)$ is dense in $L^p(\mu)$ for all $1 \le p < \infty$. Here is a possible outline.

Outline: Fix a $\lambda \in \mathbb{R}^d$ and let $f_n(x) = (\lambda \cdot x)^n / n!$ for all $n \in \mathbb{N}$.

1. Use calculus to verify $\sup_{t \ge 0} t^\alpha e^{-Mt} = (\alpha/M)^\alpha e^{-\alpha}$ for all $\alpha \ge 0$ where $(0/M)^0 := 1$. Use this estimate along with the identity
$$|\lambda \cdot x|^{pn} \le |\lambda|^{pn} |x|^{pn} = \left(|x|^{pn} e^{-M|x|}\right) |\lambda|^{pn} e^{M|x|}$$
to find an estimate on $\|f_n\|_p$.
2. Use your estimate on $\|f_n\|_p$ to show $\sum_{n=0}^\infty \|f_n\|_p < \infty$ and conclude
$$\lim_{N\to\infty} \left\| e^{i\lambda\cdot(\cdot)} - \sum_{n=0}^N i^n f_n \right\|_p = 0.$$
3. Now finish by appealing to Exercise 22.11.

Exercise 22.14. Again let μ be a finite measure on $\mathcal{B}_{\mathbb{R}^d}$ but now assume there exists an $\varepsilon > 0$ such that $C := \int_{\mathbb{R}^d} e^{\varepsilon|x|} d\mu(x) < \infty$. Also let $q > 1$ and $h \in L^q(\mu)$ be a function such that $\int_{\mathbb{R}^d} h(x) x^\alpha d\mu(x) = 0$ for all $\alpha \in \mathbb{N}_0^d$. (As mentioned in Exercise 22.14, $\mathcal{P}(\mathbb{R}^d) \subset L^p(\mu)$ for all $1 \le p < \infty$, so $x \to h(x)x^\alpha$ is in $L^1(\mu)$.) Show $h(x) = 0$ for μ– a.e. x using the following outline.

Outline: Fix a $\lambda \in \mathbb{R}^d$, let $f_n(x) = (\lambda \cdot x)^n / n!$ for all $n \in \mathbb{N}$, and let $p = q/(q-1)$ be the conjugate exponent to q.

1. Use calculus to verify $\sup_{t \ge 0} t^\alpha e^{-\varepsilon t} = (\alpha/\varepsilon)^\alpha e^{-\alpha}$ for all $\alpha \ge 0$ where $(0/\varepsilon)^0 := 1$. Use this estimate along with the identity
$$|\lambda \cdot x|^{pn} \le |\lambda|^{pn} |x|^{pn} = \left(|x|^{pn} e^{-\varepsilon|x|}\right) |\lambda|^{pn} e^{\varepsilon|x|}$$
to find an estimate on $\|f_n\|_p$.

2. Use your estimate on $\|f_n\|_p$ to show there exists $\delta > 0$ such that $\sum_{n=0}^{\infty} \|f_n\|_p < \infty$ when $|\lambda| \le \delta$ and conclude for $|\lambda| \le \delta$ that $e^{i\lambda \cdot x} = L^p(\mu)\text{-}\sum_{n=0}^{\infty} i^n f_n(x)$. Conclude from this that
$$\int_{\mathbb{R}^d} h(x) e^{i\lambda \cdot x} d\mu(x) = 0 \text{ when } |\lambda| \le \delta.$$
3. Let $\lambda \in \mathbb{R}^d$ ($|\lambda|$ not necessarily small) and set $g(t) := \int_{\mathbb{R}^d} e^{it\lambda \cdot x} h(x) d\mu(x)$ for $t \in \mathbb{R}$. Show $g \in C^\infty(\mathbb{R})$ and
$$g^{(n)}(t) = \int_{\mathbb{R}^d} (i\lambda \cdot x)^n e^{it\lambda \cdot x} h(x) d\mu(x) \text{ for all } n \in \mathbb{N}.$$
4. Let $T = \sup\{\tau \ge 0 : g|_{[0,\tau]} \equiv 0\}$. By Step 2., $T \ge \delta$. If $T < \infty$, then
$$0 = g^{(n)}(T) = \int_{\mathbb{R}^d} (i\lambda \cdot x)^n e^{iT\lambda \cdot x} h(x) d\mu(x) \text{ for all } n \in \mathbb{N}.$$
Use Step 3. with h replaced by $e^{iT\lambda \cdot x} h(x)$ to conclude
$$g(T+t) = \int_{\mathbb{R}^d} e^{i(T+t)\lambda \cdot x} h(x) d\mu(x) = 0 \text{ for all } t \le \delta / |\lambda|.$$
This violates the definition of T and therefore $T = \infty$ and in particular we may take $T = 1$ to learn
$$\int_{\mathbb{R}^d} h(x) e^{i\lambda \cdot x} d\mu(x) = 0 \text{ for all } \lambda \in \mathbb{R}^d.$$
5. Use Exercise 22.11 to conclude that
$$\int_{\mathbb{R}^d} h(x) g(x) d\mu(x) = 0$$
for all $g \in L^p(\mu)$. Now choose g judiciously to finish the proof.

23

$\mathbf{L^2}$ - Hilbert Spaces Techniques and Fourier Series

This section is concerned with Hilbert spaces presented as in the following example.

Example 23.1. Let $(X, \mathcal{M}, \mu)$ be a measure space. Then $H := L^2(X, \mathcal{M}, \mu)$ with inner product

$$\langle f|g\rangle = \int_X f \cdot \bar{g} d\mu$$

is a Hilbert space.

It will be convenient to define

$$\langle f, g\rangle := \int_X f\,(x)\, g\,(x)\, d\mu\,(x) \tag{23.1}$$

for all measurable functions f, g on X such that $fg \in L^1(\mu)$. So with this notation we have $\langle f|g\rangle = \langle f, \bar{g}\rangle$ for all $f, g \in H$.

Exercise 23.1. Let $K : L^2(\nu) \to L^2(\mu)$ be the operator defined in Exercise 21.12. Show $K^* : L^2(\mu) \to L^2(\nu)$ is the operator given by

$$K^* g(y) = \int_X \bar{k}(x, y) g(x) d\mu(x).$$

23.1 $\mathbf{L^2}$-Orthonoramal Basis

Example 23.2. 1. Let $H = L^2([-1, 1], dm)$, $A := \{1, x, x^2, x^3 \dots\}$ and $\beta \subset H$ be the result of doing the Gram-Schmidt procedure on A. By the Stone-Weierstrass theorem or by Exercise 22.13 directly, A is total in H. Hence by Remark 14.26, β is an orthonormal basis for H. The basis, β, consists of polynomials which up to normalization are the so called "**Legendre polynomials.**"

2. Let $H = L^2(\mathbb{R}, e^{-\frac{1}{2}x^2}dx)$ and $A := \{1, x, x^2, x^3 \dots\}$. Again by Exercise 22.13, A is total in H and hence the Gram-Schmidt procedure applied to A produces an orthonormal basis, β, of polynomial functions for H. This basis consists, up to normalizations, of the so called "**Hermite polynomials**" on $\mathbb{R}$.

Remark 23.3 (An Interesting Phenomena). Let $H = L^2([-1,1], dm)$ and $B := \{1, x^3, x^6, x^9, \dots\}$. Then again A is total in H by the same argument as in item 2. Example 23.2. This is true even though B is a proper subset of A. Notice that A is an algebraic basis for the polynomials on $[-1,1]$ while B is not! The following computations may help relieve some of the reader's anxiety. Let $f \in L^2([-1,1], dm)$, then, making the change of variables $x = y^{1/3}$, shows that

$$\int_{-1}^{1} |f(x)|^2 \, dx = \int_{-1}^{1} \left|f(y^{1/3})\right|^2 \frac{1}{3} y^{-2/3} dy = \int_{-1}^{1} \left|f(y^{1/3})\right|^2 d\mu(y) \tag{23.2}$$

where $d\mu(y) = \frac{1}{3}y^{-2/3}dy$. Since $\mu([-1,1]) = m([-1,1]) = 2$, μ is a finite measure on $[-1,1]$ and hence by Exercise 22.13 $A := \{1, x, x^2, x^3 \dots\}$ is a total (see Definition 14.25) in $L^2([-1,1], d\mu)$. In particular for any $\varepsilon > 0$ there exists a polynomial $p(y)$ such that

$$\int_{-1}^{1} \left|f(y^{1/3}) - p(y)\right|^2 d\mu(y) < \varepsilon^2.$$

However, by Eq. (23.2) we have

$$\varepsilon^2 > \int_{-1}^{1} \left|f(y^{1/3}) - p(y)\right|^2 d\mu(y) = \int_{-1}^{1} \left|f(x) - p(x^3)\right|^2 dx.$$

Alternatively, if $f \in C([-1,1])$, then $g(y) = f(y^{1/3})$ is back in $C([-1,1])$. Therefore for any $\varepsilon > 0$, there exists a polynomial $p(y)$ such that

$$\begin{aligned} \varepsilon > \|g - p\|_\infty &= \sup\{|g(y) - p(y)| : y \in [-1,1]\} \\ &= \sup\left\{\left|g(x^3) - p(x^3)\right| : x \in [-1,1]\right\} \\ &= \sup\left\{\left|f(x) - p(x^3)\right| : x \in [-1,1]\right\}. \end{aligned}$$

This gives another proof the polynomials in x^3 are dense in $C([-1,1])$ and hence in $L^2([-1,1])$.

Exercise 23.2. Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces such that $L^2(\mu)$ and $L^2(\nu)$ are separable. If $\{f_n\}_{n=1}^\infty$ and $\{g_m\}_{m=1}^\infty$ are orthonormal bases for $L^2(\mu)$ and $L^2(\nu)$ respectively, then $\beta := \{f_n \otimes g_m : m, n \in \mathbb{N}\}$ is an orthonormal basis for $L^2(\mu \otimes \nu)$. (Recall that $f \otimes g(x,y) := f(x) g(y)$, see Notation 20.4.) **Hint:** model your proof of the proof of Proposition 14.28.

Exercise 23.3. Suppose H is a Hilbert space and $\{H_n : n \in \mathbb{N}\}$ are closed subspaces of H such that $H_n \perp H_m$ for all $m \neq n$ and if $f \in H$ with $f \perp H_n$ for all $n \in \mathbb{N}$, then $f = 0$. Show:

1. If $f_n \in H_n$ for all $n \in \mathbb{N}$ satisfy $\sum_{n=1}^{\infty} \|f_n\|^2 < \infty$ then $\sum_{n=1}^{\infty} f_n$ exists in H.
2. Every element $f \in H$ may be uniquely written as $f = \sum_{n=1}^{\infty} f_n$ with $f_n \in H$ as in item 1.
   (For this reason we will write $H = \oplus_{n=1}^{\infty} H_n$ under the hypothesis of this exercise.)

Exercise 23.4. Suppose $(X, \mathcal{M}, \mu)$ is a measure space and $X = \coprod_{n=1}^{\infty} X_n$ with $X_n \in \mathcal{M}$ and $\mu(X_n) > 0$ for all n. Then $U : L^2(X, \mu) \to \oplus_{n=1}^{\infty} L^2(X_n, \mu)$ defined by $(Uf)_n := f 1_{X_n}$ is unitary.

23.2 Hilbert Schmidt Operators

In this section H and B will be Hilbert spaces.

Proposition 23.4. *Let H and B be a separable Hilbert spaces, $K : H \to B$ be a bounded linear operator, $\{e_n\}_{n=1}^{\infty}$ and $\{u_m\}_{m=1}^{\infty}$ be orthonormal basis for H and B respectively. Then:*

1. *$\sum_{n=1}^{\infty} \|Ke_n\|^2 = \sum_{m=1}^{\infty} \|K^* u_m\|^2$ allowing for the possibility that the sums are infinite. In particular the **Hilbert Schmidt norm** of K,*
$$\|K\|_{HS}^2 := \sum_{n=1}^{\infty} \|Ke_n\|^2,$$
*is well defined independent of the choice of orthonormal basis $\{e_n\}_{n=1}^{\infty}$. We say $K : H \to B$ is a **Hilbert Schmidt operator** if $\|K\|_{HS} < \infty$ and let $HS(H, B)$ denote the space of Hilbert Schmidt operators from H to B.*
2. *For all $K \in L(H, B)$, $\|K\|_{HS} = \|K^*\|_{HS}$ and*
$$\|K\|_{HS} \geq \|K\|_{op} := \sup \{\|Kh\| : h \in H \ni \|h\| = 1\}.$$
3. *The set $HS(H, B)$ is a subspace of $L(H, B)$ (the bounded operators from $H \to B$), $\|\cdot\|_{HS}$ is a norm on $HS(H, B)$ for which $(HS(H, B), \|\cdot\|_{HS})$ is a Hilbert space, and the corresponding inner product is given by*
$$\langle K_1 | K_2 \rangle_{HS} = \sum_{n=1}^{\infty} \langle K_1 e_n | K_2 e_n \rangle. \tag{23.3}$$
4. *If $K : H \to B$ is a bounded finite rank operator, then K is Hilbert Schmidt.*

5. *Let* $P_N x := \sum_{n=1}^{N} \langle x|e_n\rangle e_n$ *be orthogonal projection onto* $\mathrm{span}\{e_n : n \leq N\} \subset H$ *and for* $K \in HS(H,B)$, *let* $K_N := KP_N$. *Then*

$$\|K - K_N\|_{op}^2 \leq \|K - K_N\|_{HS}^2 \to 0 \text{ as } N \to \infty,$$

which shows that finite rank operators are dense in $(HS(H,B), \|\cdot\|_{HS})$. *In particular of* $HS(H,B) \subset \mathcal{K}(H,B)$ *– the space of compact operators from* $H \to B$.

6. *If* Y *is another Hilbert space and* $A : Y \to H$ *and* $C : B \to Y$ *are bounded operators, then*

$$\|KA\|_{HS} \leq \|K\|_{HS} \|A\|_{op} \text{ and } \|CK\|_{HS} \leq \|K\|_{HS} \|C\|_{op},$$

in particular $HS(H,H)$ *is an ideal in* $L(H)$.

Proof. Items 1. and 2. By Parseval's equality and Fubini's theorem for sums,

$$\begin{aligned}\sum_{n=1}^{\infty} \|Ke_n\|^2 &= \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} |\langle Ke_n|u_m\rangle|^2 \\ &= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} |\langle e|K^* u_m\rangle|^2 = \sum_{m=1}^{\infty} \|K^* u_m\|^2 .\end{aligned}$$

This proves $\|K\|_{HS}$ is well defined independent of basis and that $\|K\|_{HS} = \|K^*\|_{HS}$. For $x \in H \setminus \{0\}$, $x/\|x\|$ may be taken to be the first element in an orthonormal basis for H and hence

$$\left\| K \frac{x}{\|x\|} \right\| \leq \|K\|_{HS} .$$

Multiplying this inequality by $\|x\|$ shows $\|Kx\| \leq \|K\|_{HS} \|x\|$ and hence $\|K\|_{op} \leq \|K\|_{HS}$.

Item 3. For $K_1, K_2 \in L(H,B)$,

$$\begin{aligned}\|K_1 + K_2\|_{HS} &= \sqrt{\sum_{n=1}^{\infty} \|K_1 e_n + K_2 e_n\|^2} \\ &\leq \sqrt{\sum_{n=1}^{\infty} [\|K_1 e_n\| + \|K_2 e_n\|]^2} \\ &= \|\{\|K_1 e_n\| + \|K_2 e_n\|\}_{n=1}^{\infty}\|_{\ell_2} \\ &\leq \|\{\|K_1 e_n\|\}_{n=1}^{\infty}\|_{\ell_2} + \|\{\|K_2 e_n\|\}_{n=1}^{\infty}\|_{\ell_2} \\ &= \|K_1\|_{HS} + \|K_2\|_{HS} .\end{aligned}$$

From this triangle inequality and the homogeneity properties of $\|\cdot\|_{HS}$, we now easily see that $HS(H,B)$ is a subspace of $L(H,B)$ and $\|\cdot\|_{HS}$ is a norm on $HS(H,B)$. Since

$$\begin{aligned}\sum_{n=1}^{\infty}|\langle K_1e_n|K_2e_n\rangle| &\le \sum_{n=1}^{\infty}\|K_1e_n\|\,\|K_2e_n\| \\ &\le \sqrt{\sum_{n=1}^{\infty}\|K_1e_n\|^2}\sqrt{\sum_{n=1}^{\infty}\|K_2e_n\|^2}=\|K_1\|_{HS}\,\|K_2\|_{HS},\end{aligned}$$

the sum in Eq. (23.3) is well defined and is easily checked to define an inner product on $HS(H,B)$ such that $\|K\|_{HS}^2=\langle K|K\rangle_{HS}$.

To see that $HS(H,B)$ is complete in this inner product suppose $\{K_m\}_{m=1}^{\infty}$ is a $\|\cdot\|_{HS}$ – Cauchy sequence in $HS(H,B)$. Because $L(H,B)$ is complete, there exists $K\in L(H,B)$ such that $\|K-K_m\|_{op}\to 0$ as $m\to\infty$. Thus, making use of Fatou's Lemma 4.12,

$$\begin{aligned}\|K-K_m\|_{HS}^2 &= \sum_{n=1}^{\infty}\|(K-K_m)\,e_n\|^2 \\ &= \sum_{n=1}^{\infty}\liminf_{l\to\infty}\|(K_l-K_m)\,e_n\|^2 \\ &\le \liminf_{l\to\infty}\sum_{n=1}^{\infty}\|(K_l-K_m)\,e_n\|^2 \\ &= \liminf_{l\to\infty}\|K_l-K_m\|_{HS}^2\to 0 \text{ as } m\to\infty.\end{aligned}$$

Hence $K\in HS(H,B)$ and $\lim_{m\to\infty}\|K-K_m\|_{HS}^2=0$.

Item 4. Let $N:=\dim K(H)$ and $\{v_n\}_{n=1}^{N}$ be an orthonormal basis for $\mathrm{Ran}(K)=K(H)$. Then, for all $h\in H$,

$$\|Kh\|_B^2=\sum_{n=1}^{N}|\langle Kh|v_n\rangle|^2=\sum_{n=1}^{N}|\langle h|K^*v_n\rangle|^2.$$

Summing this equation on $h\in\beta$ where an β is an orthonormal basis for H shows

$$\|K\|_{HS}^2=\sum_{h\in\beta}\|Kh\|_B^2=\sum_{n=1}^{N}\|K^*v_n\|_H^2<\infty.$$

Item 5. Simply observe,

$$\|K-K_N\|_{op}^2\le\|K-K_N\|_{HS}^2=\sum_{n>N}\|Ke_n\|^2\to 0 \text{ as } N\to\infty.$$

Item 6. For $C\in L(B,Y)$ and $K\in L(H,B)$ then

$$\|CK\|_{HS}^2 = \sum_{n=1}^{\infty} \|CKe_n\|^2 \le \|C\|_{op}^2 \sum_{n=1}^{\infty} \|Ke_n\|^2 = \|C\|_{op}^2 \|K\|_{HS}^2$$

and for $A \in L(Y, H)$,

$$\|KA\|_{HS} = \|A^* K^*\|_{HS} \le \|A^*\|_{op} \|K^*\|_{HS} = \|A\|_{op} \|K\|_{HS}.$$

■

Remark 23.5. The separability assumptions made in Proposition 23.4 are unnecessary. In general, we define

$$\|K\|_{HS}^2 = \sum_{e \in \beta} \|Ke\|^2$$

where $\beta \subset H$ is an orthonormal basis. The same proof of Item 1. of Proposition 23.4 shows $\|K\|_{HS}$ is well defined and $\|K\|_{HS} = \|K^*\|_{HS}$. If $\|K\|_{HS}^2 < \infty$, then there exists a countable subset $\beta_0 \subset \beta$ such that $Ke = 0$ if $e \in \beta \setminus \beta_0$. Let $H_0 := \overline{\mathrm{span}(\beta_0)}$ and $B_0 := \overline{K(H_0)}$. Then $K(H) \subset B_0$, $K|_{H_0^\perp} = 0$ and hence by applying the results of Proposition 23.4 to $K|_{H_0} : H_0 \to B_0$ one easily sees that the separability of H and B are unnecessary in Proposition 23.4.

Example 23.6. Let (X, μ) be a measure space, $H = L^2(X, \mu)$ and

$$k(x, y) := \sum_{i=1}^{n} f_i(x) g_i(y)$$

where

$$f_i, g_i \in L^2(X, \mu) \text{ for } i = 1, \dots, n.$$

Define $(Kf)(x) = \int_X k(x, y) f(y) d\mu(y)$, then $K : L^2(X, \mu) \to L^2(X, \mu)$ is a finite rank operator and hence Hilbert Schmidt.

Exercise 23.5. Suppose that (X, μ) is a σ–finite measure space such that $H = L^2(X, \mu)$ is separable and $k : X \times X \to \mathbb{R}$ is a measurable function, such that

$$\|k\|_{L^2(X \times X, \mu \otimes \mu)}^2 := \int_{X \times X} |k(x, y)|^2 d\mu(x) d\mu(y) < \infty.$$

Define, for $f \in H$,

$$Kf(x) = \int_X k(x, y) f(y) d\mu(y),$$

when the integral makes sense. Show:

1. $Kf(x)$ is defined for μ–a.e. x in X.
2. The resulting function Kf is in H and $K : H \to H$ is linear.
3. $\|K\|_{HS} = \|k\|_{L^2(X \times X, \mu \otimes \mu)} < \infty$. (This implies $K \in HS(H, H)$.)

Example 23.7. Suppose that $\Omega \subset \mathbb{R}^n$ is a bounded set, $\alpha < n$, then the operator $K : L^2(\Omega, m) \to L^2(\Omega, m)$ defined by

$$Kf(x) := \int_\Omega \frac{1}{|x-y|^\alpha} f(y)dy$$

is compact.

Proof. For $\varepsilon \geq 0$, let

$$K_\varepsilon f(x) := \int_\Omega \frac{1}{|x-y|^\alpha + \varepsilon} f(y)dy = [g_\varepsilon * (1_\Omega f)](x)$$

where $g_\varepsilon(x) = \frac{1}{|x|^\alpha + \varepsilon} 1_C(x)$ with $C \subset \mathbb{R}^n$ a sufficiently large ball such that $\Omega - \Omega \subset C$. Since $\alpha < n$, it follows that

$$g_\varepsilon \leq g_0 = |\cdot|^{-\alpha} 1_C \in L^1(\mathbb{R}^n, m).$$

Hence it follows by Proposition 22.23 that

$$\begin{aligned} \|(K - K_\varepsilon) f\|_{L^2(\Omega)} &\leq \|(g_0 - g_\varepsilon) * (1_\Omega f)\|_{L^2(\mathbb{R}^n)} \\ &\leq \|(g_0 - g_\varepsilon)\|_{L^1(\mathbb{R}^n)} \|1_\Omega f\|_{L^2(\mathbb{R}^n)} \\ &= \|(g_0 - g_\varepsilon)\|_{L^1(\mathbb{R}^n)} \|f\|_{L^2(\Omega)} \end{aligned}$$

which implies

$$\begin{aligned} \|K - K_\varepsilon\|_{B(L^2(\Omega))} &\leq \|g_0 - g_\varepsilon\|_{L^1(\mathbb{R}^n)} \\ &= \int_C \left| \frac{1}{|x|^\alpha + \varepsilon} - \frac{1}{|x|^\alpha} \right| dx \to 0 \text{ as } \varepsilon \downarrow 0 \end{aligned} \tag{23.4}$$

by the dominated convergence theorem. For any $\varepsilon > 0$,

$$\int_{\Omega \times \Omega} \left[\frac{1}{|x-y|^\alpha + \varepsilon} \right]^2 dxdy < \infty,$$

and hence K_ε is Hilbert Schmidt and hence compact. By Eq. (23.4), $K_\varepsilon \to K$ as $\varepsilon \downarrow 0$ and hence it follows that K is compact as well. ∎

Exercise 23.6. Let $H := L^2([0,1], m)$, $k(x,y) := \min(x,y)$ for $x, y \in [0,1]$ and define $K : H \to H$ by

$$Kf(x) = \int_0^1 k(x,y) f(y)\, dy.$$

By Exercise 23.5, K is a Hilbert Schmidt operator and it is easily seen that K is self-adjoint. Show:

1. Show $\langle Kf|g''\rangle = -\langle f|g\rangle$ for all $g \in C_c^\infty((0,1))$ and use this to conclude that $\mathrm{Nul}(K) = \{0\}$.
2. Now suppose that $f \in H$ is an eigenvector of K with eigenvalue $\lambda \neq 0$. Show that there is a version of f in $C([0,1]) \cap C^2((0,1))$ and this version, still denoted by f, solves

$$\lambda f'' = -f \text{ with } f(0) = f'(1) = 0. \tag{23.5}$$

where $f'(1) := \lim_{x\uparrow 1} f'(x)$.
3. Use Eq. (23.5) to find all the eigenvalues and eigenfunctions of K.
4. Use the results above along with the spectral Theorem 14.45, to show

$$\left\{\sqrt{2}\sin\left(n\frac{\pi}{2}x\right) : n \in \mathbb{N}\right\}$$

is an orthonormal basis for $L^2([0,1], m)$.

Exercise 23.7. Let $(X, \mathcal{M}, \mu)$ be a σ – finite measure space, $a \in L^\infty(\mu)$ and let A be the bounded operator on $H := L^2(\mu)$ defined by $Af(x) = a(x) f(x)$ for all $f \in H$. (We will denote A by M_a in the future.) Show:

1. $\|A\|_{op} = \|a\|_{L^\infty(\mu)}$.
2. $A^* = M_{\bar{a}}$.
3. $\sigma(A) = \mathrm{essran}(a)$ where $\sigma(A)$ is the spectrum of A and $\mathrm{essran}(a)$ is the essential range of a, see Definitions 14.30 and 21.40 respectively.
4. Show λ is an eigenvalue for $A = M_a$ iff $\mu(\{a = \lambda\}) > 0$, i.e. iff a has a "flat spot of height λ."

23.3 Fourier Series Considerations

Throughout this section we will let $d\theta$, dx, $d\alpha$, etc. denote Lebesgue measure on $\mathbb{R}^d$ normalized so that the cube, $Q := (-\pi, \pi]^d$, has measure one, i.e. $d\theta = (2\pi)^{-d} dm(\theta)$ where m is standard Lebesgue measure on $\mathbb{R}^d$. As usual, for $\alpha \in \mathbb{N}_0^d$, let

$$D_\theta^\alpha = \left(\frac{1}{i}\right)^{|\alpha|} \frac{\partial^{|\alpha|}}{\partial\theta_1^{\alpha_1} \dots \partial\theta_d^{\alpha_d}}.$$

Notation 23.8 *Let $C_{per}^k(\mathbb{R}^d)$ denote the 2π – periodic functions in $C^k(\mathbb{R}^d)$, that is $f \in C_{per}^k(\mathbb{R}^d)$ iff $f \in C^k(\mathbb{R}^d)$ and $f(\theta + 2\pi e_i) = f(\theta)$ for all $\theta \in \mathbb{R}^d$ and $i = 1, 2, \dots, d$. Further let $\langle\cdot|\cdot\rangle$ denote the inner product on the Hilbert space, $H := L^2([-\pi, \pi]^d)$, given by*

$$\langle f|g\rangle := \int_Q f(\theta)\bar{g}(\theta) d\theta = \left(\frac{1}{2\pi}\right)^d \int_Q f(\theta)\bar{g}(\theta) dm(\theta)$$

and define $e_k(\theta) := e^{ik\cdot\theta}$ for all $k \in \mathbb{Z}^d$. For $f \in L^1(Q)$, we will write $\tilde{f}(k)$ for the ***Fourier coefficient****,*

$$\tilde{f}(k) := \langle f|e_k\rangle = \int_Q f(\theta)e^{-ik\cdot\theta}d\theta. \tag{23.6}$$

Since any 2π – periodic functions on $\mathbb{R}^d$ may be identified with function on the d - dimensional torus, $\mathbb{T}^d \cong \mathbb{R}^d/(2\pi\mathbb{Z})^d \cong \left(S^1\right)^d$, I may also write $C^k(\mathbb{T}^d)$ for $C^k_{per}(\mathbb{R}^d)$ and $L^p\left(\mathbb{T}^d\right)$ for $L^p(Q)$ where elements in $f \in L^p(Q)$ are to be thought of as there extensions to 2π – periodic functions on $\mathbb{R}^d$.

Theorem 23.9 (Fourier Series). *The functions* $\beta := \left\{e_k : k \in \mathbb{Z}^d\right\}$ *form an orthonormal basis for* H, *i.e. if* $f \in H$ *then*

$$f = \sum_{k\in\mathbb{Z}^d} \langle f|e_k\rangle e_k = \sum_{k\in\mathbb{Z}^d} \tilde{f}(k)e_k \tag{23.7}$$

where the convergence takes place in $L^2([-\pi,\pi]^d)$.

Proof. Simple computations show $\beta := \left\{e_k : k \in \mathbb{Z}^d\right\}$ is an orthonormal set. We now claim that β is an orthonormal basis. To see this recall that $C_c((-\pi,\pi)^d)$ is dense in $L^2((-\pi,\pi)^d, dm)$. Any $f \in C_c((-\pi,\pi))$ may be extended to be a continuous 2π – periodic function on $\mathbb{R}$ and hence by Exercise 12.13 and Remark 12.44, f may uniformly (and hence in L^2) be approximated by a trigonometric polynomial. Therefore β is a total orthonormal set, i.e. β is an orthonormal basis.

This may also be proved by first proving the case $d = 1$ as above and then using Exercise 23.2 inductively to get the result for any d. ■

Exercise 23.8. Let A be the operator defined in Lemma 14.36 and for $g \in L^2(\mathbb{T})$, let $Ug(k) := \tilde{g}(k)$ so that $U : L^2(\mathbb{T}) \to \ell^2(\mathbb{Z})$ is unitary. Show $U^{-1}AU = M_a$ where $a \in C^\infty_{per}(\mathbb{R})$ is a function to be found. Use this representation and the results in Exercise 23.7 to give a simple proof of the results in Lemma 14.36.

23.3.1 Dirichlet, Fejér and Kernels

Although the sum in Eq. (23.7) is guaranteed to converge relative to the Hilbertian norm on H it certainly need not converge pointwise even if $f \in C_{per}\left(\mathbb{R}^d\right)$ as will be proved in Section **??** below. Nevertheless, if f is sufficiently regular, then the sum in Eq. (23.7) will converge pointwise as we will now show. In the process we will give a direct and constructive proof of the result in Exercise 12.13, see Theorem 23.11 below.

Let us restrict our attention to $d = 1$ here. Consider

$$\begin{aligned} f_n(\theta) &= \sum_{|k|\le n} \tilde{f}(k)e_k(\theta) = \sum_{|k|\le n} \frac{1}{2\pi}\left[\int_{[-\pi,\pi]} f(x)e^{-ik\cdot x}dx\right] e_k(\theta) \\ &= \frac{1}{2\pi}\int_{[-\pi,\pi]} f(x) \sum_{|k|\le n} e^{ik\cdot(\theta-x)}dx = \frac{1}{2\pi}\int_{[-\pi,\pi]} f(x)D_n(\theta-x)dx \end{aligned} \tag{23.8}$$

where

$$D_n(\theta) := \sum_{k=-n}^{n} e^{ik\theta}$$

is called the **Dirichlet kernel**. Letting $\alpha = e^{i\theta/2}$, we have

$$\begin{aligned} D_n(\theta) &= \sum_{k=-n}^{n} \alpha^{2k} = \frac{\alpha^{2(n+1)} - \alpha^{-2n}}{\alpha^2 - 1} = \frac{\alpha^{2n+1} - \alpha^{-(2n+1)}}{\alpha - \alpha^{-1}} \\ &= \frac{2i\sin(n+\frac{1}{2})\theta}{2i\sin\frac{1}{2}\theta} = \frac{\sin(n+\frac{1}{2})\theta}{\sin\frac{1}{2}\theta}. \end{aligned}$$

and therefore

$$D_n(\theta) := \sum_{k=-n}^{n} e^{ik\theta} = \frac{\sin(n+\frac{1}{2})\theta}{\sin\frac{1}{2}\theta}, \tag{23.9}$$

see Figure 23.3.1.

This is a plot D_1 and D_{10}.

with the understanding that the right side of this equation is $2n+1$ whenever $\theta \in 2\pi\mathbb{Z}$.

Theorem 23.10. *Suppose $f \in L^1([-\pi,\pi], dm)$ and f is differentiable at some $\theta \in [-\pi,\pi]$, then $\lim_{n\to\infty} f_n(\theta) = f(\theta)$ where f_n is as in Eq. (23.8).*

Proof. Observe that

$$\frac{1}{2\pi}\int_{[-\pi,\pi]} D_n(\theta - x)dx = \frac{1}{2\pi}\int_{[-\pi,\pi]} \sum_{|k|\le n} e^{ik\cdot(\theta-x)}dx = 1$$

and therefore,

$$\begin{aligned} f_n(\theta) - f(\theta) &= \frac{1}{2\pi}\int_{[-\pi,\pi]} [f(x) - f(\theta)]\, D_n(\theta - x)dx \\ &= \frac{1}{2\pi}\int_{[-\pi,\pi]} [f(x) - f(\theta - x)]\, D_n(x)dx \\ &= \frac{1}{2\pi}\int_{[-\pi,\pi]} \left[\frac{f(\theta - x) - f(\theta)}{\sin\frac{1}{2}x}\right] \sin(n + \frac{1}{2})x\, dx. \end{aligned} \tag{23.10}$$

If f is differentiable at θ, the last expression in Eq. (23.10) tends to 0 as $n \to \infty$ by the Riemann Lebesgue Lemma (Corollary 22.17 or Lemma 22.37) and the fact that $1_{[-\pi,\pi]}(x)\frac{f(\theta-x)-f(\theta)}{\sin\frac{1}{2}x} \in L^1(dx)$. ∎

Despite the Dirichlet kernel not being positive, it still satisfies the approximate δ – sequence property, $\frac{1}{2\pi}D_n \to \delta_0$ as $n \to \infty$, when acting on C^1 – periodic functions in θ. In order to improve the convergence properties it is reasonable to try to replace $\{f_n : n \in \mathbb{N}_0\}$ by the sequence of averages (see Exercise 7.13),

$$\begin{aligned} F_N(\theta) &= \frac{1}{N+1}\sum_{n=0}^{N} f_n(\theta) = \frac{1}{N+1}\sum_{n=0}^{N}\frac{1}{2\pi}\int_{[-\pi,\pi]} f(x) \sum_{|k|\le n} e^{ik\cdot(\theta - x)}dx \\ &= \frac{1}{2\pi}\int_{[-\pi,\pi]} K_N(\theta - x) f(x)dx \end{aligned}$$

where

$$K_N(\theta) := \frac{1}{N+1}\sum_{n=0}^{N}\sum_{|k|\le n} e^{ik\cdot\theta} \tag{23.11}$$

is the **Fejér kernel.**

Theorem 23.11. *The Fejér kernel K_N in Eq. (23.11) satisfies:*

1.
$$K_N(\theta) = \sum_{n=-N}^{N}\left[1 - \frac{|n|}{N+1}\right] e^{in\theta} \tag{23.12}$$
$$= \frac{1}{N+1}\frac{\sin^2\left(\frac{N+1}{2}\theta\right)}{\sin^2\left(\frac{\theta}{2}\right)}. \tag{23.13}$$
2. $K_N(\theta) \ge 0$.
3. $\frac{1}{2\pi}\int_{-\pi}^{\pi} K_N(\theta)d\theta = 1$
4. $\sup_{\varepsilon \le |\theta| \le \pi} K_N(\theta) \to 0$ *as* $N \to \infty$ *for all* $\varepsilon > 0$, *see Figure 23.1.*
5. *For any continuous* 2π *– periodic function* f *on* $\mathbb{R}$, $K_N * f(\theta) \to f(\theta)$ *uniformly in* θ *as* $N \to \infty$, *where*

$$\begin{aligned} K_N * f(\theta) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} K_N(\theta-\alpha) f(\alpha) d\alpha \\ &= \sum_{n=-N}^{N} \left[1 - \frac{|n|}{N+1}\right] \tilde{f}(n)\, e^{in\theta}. \end{aligned} \tag{23.14}$$

Fig. 23.1. Plots of $K_N(\theta)$ for $N = 2$, 7 and 13.

Proof. 1. Equation (23.12) is a consequence of the identity,

$$\sum_{n=0}^{N} \sum_{|k|\le n} e^{ik\cdot\theta} = \sum_{|k|\le n\le N} e^{ik\cdot\theta} = \sum_{|k|\le N} (N+1-|k|)\, e^{ik\cdot\theta}.$$

Moreover, letting $\alpha = e^{i\theta/2}$ and using Eq. (3.3) shows

$$
\begin{aligned}
K_N(\theta) &= \frac{1}{N+1}\sum_{n=0}^{N}\sum_{|k|\le n}\alpha^{2k} = \frac{1}{N+1}\sum_{n=0}^{N}\frac{\alpha^{2n+2}-\alpha^{-2n}}{\alpha^2-1} \\
&= \frac{1}{(N+1)\left(\alpha-\alpha^{-1}\right)}\sum_{n=0}^{N}\left[\alpha^{2n+1}-\alpha^{-2n-1}\right] \\
&= \frac{1}{(N+1)\left(\alpha-\alpha^{-1}\right)}\sum_{n=0}^{N}\left[\alpha\alpha^{2n}-\alpha^{-1}\alpha^{-2n}\right] \\
&= \frac{1}{(N+1)\left(\alpha-\alpha^{-1}\right)}\left[\alpha\frac{\alpha^{2N+2}-1}{\alpha^2-1}-\alpha^{-1}\frac{\alpha^{-2N-2}-1}{\alpha^{-2}-1}\right] \\
&= \frac{1}{(N+1)\left(\alpha-\alpha^{-1}\right)^2}\left[\alpha^{2(N+1)}-1+\alpha^{-2(N+1)}-1\right] \\
&= \frac{1}{(N+1)\left(\alpha-\alpha^{-1}\right)^2}\left[\alpha^{(N+1)}-\alpha^{-(N+1)}\right]^2 \\
&= \frac{1}{N+1}\frac{\sin^2\left((N+1)\,\theta/2\right)}{\sin^2\left(\theta/2\right)}.
\end{aligned}
$$

Items 2. and 3. follow easily from Eqs. (23.13) and (23.12) respectively. Item 4. is a consequence of the elementary estimate;

$$
\sup_{\varepsilon\le|\theta|\le\pi} K_N(\theta) \le \frac{1}{N+1}\frac{1}{\sin^2\left(\frac{\varepsilon}{2}\right)}
$$

and is clearly indicated in Figure 23.1. Item 5. now follows by the standard approximate δ – function arguments, namely,

$$
\begin{aligned}
|K_N * f(\theta) - f(\theta)| &= \frac{1}{2\pi}\left|\int_{-\pi}^{\pi} K_N(\theta-\alpha)\left[f(\alpha)-f(\theta)\right]d\alpha\right| \\
&\le \frac{1}{2\pi}\int_{-\pi}^{\pi} K_N(\alpha)\left|f(\theta-\alpha)-f(\theta)\right|d\alpha \\
&\le \frac{1}{\pi}\frac{1}{N+1}\frac{1}{\sin^2\left(\frac{\varepsilon}{2}\right)}\|f\|_\infty + \frac{1}{2\pi}\int_{|\alpha|\le\varepsilon} K_N(\alpha)\left|f(\theta-\alpha)-f(\theta)\right|d\alpha \\
&\le \frac{1}{\pi}\frac{1}{N+1}\frac{1}{\sin^2\left(\frac{\varepsilon}{2}\right)}\|f\|_\infty + \sup_{|\alpha|\le\varepsilon}\left|f(\theta-\alpha)-f(\theta)\right|.
\end{aligned}
$$

Therefore,

$$
\limsup_{N\to\infty}\|K_N * f - f\|_\infty \le \sup_\theta \sup_{|\alpha|\le\varepsilon}\left|f(\theta-\alpha)-f(\theta)\right| \to 0 \text{ as } \varepsilon\downarrow 0.
$$

■

23.3.2 The Dirichlet Problems on D and the Poisson Kernel

Let $D := \{z \in \mathbb{C} : |z| < 1\}$ be the open unit disk in $\mathbb{C} \cong \mathbb{R}^2$, write $z \in \mathbb{C}$ as $z = x + iy$ or $z = re^{i\theta}$, and let $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ be the **Laplacian** acting on $C^2(D)$.

Theorem 23.12 (Dirichlet problem for D). *To every continuous function $g \in C(\mathrm{bd}(D))$ there exists a unique function $u \in C(\bar{D}) \cap C^2(D)$ solving*

$$\Delta u(z) = 0 \text{ for } z \in D \text{ and } u|_{\partial D} = g. \tag{23.15}$$

Moreover for $r < 1$, u is given by,

$$u(re^{i\theta}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_r(\theta - \alpha) u(e^{i\alpha}) d\alpha =: P_r * u(e^{i\theta}) \tag{23.16}$$

$$= \frac{1}{2\pi} \operatorname{Re} \int_{-\pi}^{\pi} \frac{1 + re^{i(\theta-\alpha)}}{1 - re^{i(\theta-\alpha)}} u(e^{i\alpha}) d\alpha \tag{23.17}$$

where P_r is the ***Poisson kernel*** *defined by*

$$P_r(\delta) := \frac{1 - r^2}{1 - 2r\cos\delta + r^2}.$$

(The problem posed in Eq. (23.15) is called the ***Dirichlet problem for*** *D.)*

Proof. In this proof, we are going to be identifying $S^1 = \mathrm{bd}(D) := \{z \in \bar{D} : |z| = 1\}$ with $[-\pi, \pi]/(\pi \sim -\pi)$ by the map $\theta \in [-\pi, \pi] \to e^{i\theta} \in S^1$. Also recall that the Laplacian Δ may be expressed in polar coordinates as,

$$\Delta u = r^{-1}\partial_r\left(r^{-1}\partial_r u\right) + \frac{1}{r^2}\partial_\theta^2 u,$$

where

$$(\partial_r u)\left(re^{i\theta}\right) = \frac{\partial}{\partial r} u\left(re^{i\theta}\right) \text{ and } (\partial_\theta u)\left(re^{i\theta}\right) = \frac{\partial}{\partial \theta} u\left(re^{i\theta}\right).$$

Uniqueness. Suppose u is a solution to Eq. (23.15) and let

$$\tilde{g}(k) := \frac{1}{2\pi} \int_{-\pi}^{\pi} g(e^{ik\theta}) e^{-ik\theta} d\theta$$

and

$$\tilde{u}(r,k) := \frac{1}{2\pi} \int_{-\pi}^{\pi} u(re^{i\theta}) e^{-ik\theta} d\theta \tag{23.18}$$

be the Fourier coefficients of $g(\theta)$ and $\theta \to u\left(re^{i\theta}\right)$ respectively. Then for $r \in (0, 1)$,

$$\begin{aligned}
r^{-1}\partial_r\left(r\partial_r\tilde{u}(r,k)\right) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} r^{-1}\partial_r\left(r^{-1}\partial_r u\right)(re^{i\theta})e^{-ik\theta}d\theta \\
&= -\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{r^2}\partial_\theta^2 u(re^{i\theta})e^{-ik\theta}d\theta \\
&= -\frac{1}{r^2}\frac{1}{2\pi}\int_{-\pi}^{\pi} u(re^{i\theta})\partial_\theta^2 e^{-ik\theta}d\theta \\
&= \frac{1}{r^2}k^2\tilde{u}(r,k)
\end{aligned}$$

or equivalently

$$r\partial_r\left(r\partial_r\tilde{u}(r,k)\right) = k^2\tilde{u}(r,k). \tag{23.19}$$

Recall the general solution to

$$r\partial_r\left(r\partial_r y(r)\right) = k^2 y(r) \tag{23.20}$$

may be found by trying solutions of the form $y(r) = r^\alpha$ which then implies $\alpha^2 = k^2$ or $\alpha = \pm k$. From this one sees that $\tilde{u}(r,k)$ solving Eq. (23.19) may be written as $\tilde{u}(r,k) = A_k r^{|k|} + B_k r^{-|k|}$ for some constants A_k and B_k when $k \neq 0$. If $k = 0$, the solution to Eq. (23.20) is gotten by simple integration and the result is $\tilde{u}(r,0) = A_0 + B_0 \ln r$. Since $\tilde{u}(r,k)$ is bounded near the origin for each k it must be that $B_k = 0$ for all $k \in \mathbb{Z}$. Hence we have shown there exists $A_k \in \mathbb{C}$ such that, for all $r \in (0,1)$,

$$A_k r^{|k|} = \tilde{u}(r,k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} u(re^{i\theta})e^{-ik\theta}d\theta. \tag{23.21}$$

Since all terms of this equation are continuous for $r \in [0,1]$, Eq. (23.21) remains valid for all $r \in [0,1]$ and in particular we have, at $r = 1$, that

$$A_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} u(e^{i\theta})e^{-ik\theta}d\theta = \tilde{g}(k).$$

Hence if u is a solution to Eq. (23.15) then u must be given by

$$u(re^{i\theta}) = \sum_{k\in\mathbb{Z}}\tilde{g}(k)r^{|k|}e^{ik\theta} \text{ for } r < 1. \tag{23.22}$$

or equivalently,

$$u(z) = \sum_{k\in\mathbb{N}_0}\tilde{g}(k)z^k + \sum_{k\in\mathbb{N}}\tilde{g}(-k)\bar{z}^k.$$

Notice that the theory of the Fourier series implies Eq. (23.22) is valid in the $L^2(d\theta)$ - sense. However more is true, since for $r < 1$, the series in Eq. (23.22) is absolutely convergent and in fact defines a C^∞ – function (see Exercise 4.11 or Corollary 19.43) which must agree with the continuous function, $\theta \to u\left(re^{i\theta}\right)$, for almost every θ and hence for all θ. This completes the proof of uniqueness.

Existence. Given $g \in C\left(\mathrm{bd}(D)\right)$, let u be defined as in Eq. (23.22). Then, again by Exercise 4.11 or Corollary 19.43, $u \in C^{\infty}\left(D\right)$. So to finish the proof it suffices to show $\lim_{x\to y} u\left(x\right) = g\left(y\right)$ for all $y \in \mathrm{bd}(D)$. Inserting the formula for $\tilde{g}(k)$ into Eq. (23.22) gives

$$u(re^{i\theta}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_r\left(\theta-\alpha\right) u(e^{i\alpha})d\alpha \text{ for all } r<1$$

where

$$\begin{aligned}
P_r\left(\delta\right) &= \sum_{k\in\mathbb{Z}} r^{|k|}e^{ik\delta} = \sum_{k=0}^{\infty} r^k e^{ik\delta} + \sum_{k=0}^{\infty} r^k e^{-ik\delta} - 1 = \\
&= \mathrm{Re}\left[2\frac{1}{1-re^{i\delta}} - 1\right] = \mathrm{Re}\left[\frac{1+re^{i\delta}}{1-re^{i\delta}}\right] \\
&= \mathrm{Re}\left[\frac{\left(1+re^{i\delta}\right)\left(1-re^{-i\delta}\right)}{|1-re^{i\delta}|^2}\right] = \mathrm{Re}\left[\frac{1-r^2+2ir\sin\delta}{1-2r\cos\delta+r^2}\right] \qquad (23.23) \\
&= \frac{1-r^2}{1-2r\cos\delta+r^2}.
\end{aligned}$$

The Poisson kernel again solves the usual approximate δ – function properties (see Figure 2), namely:

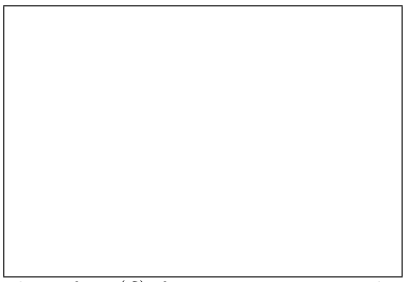
1. $P_r\left(\delta\right) > 0$ and

$$\begin{aligned}
\frac{1}{2\pi}\int_{-\pi}^{\pi} P_r\left(\theta-\alpha\right)d\alpha &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{k\in\mathbb{Z}} r^{|k|}e^{ik(\theta-\alpha)}d\alpha \\
&= \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} r^{|k|}\int_{-\pi}^{\pi} e^{ik(\theta-\alpha)}d\alpha = 1
\end{aligned}$$

and

2.

$$\sup_{\varepsilon\le|\theta|\le\pi} P_r(\theta) \le \frac{1-r^2}{1-2r\cos\varepsilon+r^2} \to 0 \text{ as } r\uparrow 1.$$



A plot of $P_r(\delta)$ for $r = 0.2$, 0.5 and 0.7.

Therefore by the same argument used in the proof of Theorem 23.11,

$$\lim_{r\uparrow 1}\sup_{\theta}\left|u\left(re^{i\theta}\right)-g\left(e^{i\theta}\right)\right|=\lim_{r\uparrow 1}\sup_{\theta}\left|\left(P_r*g\right)\left(e^{i\theta}\right)-g\left(e^{i\theta}\right)\right|=0$$

which certainly implies $\lim_{x\to y}u(x)=g(y)$ for all $y\in\mathrm{bd}(D)$. ∎

Remark 23.13 (Harmonic Conjugate). Writing $z=re^{i\theta}$, Eq. (23.17) may be rewritten as

$$u(z)=\frac{1}{2\pi}\operatorname{Re}\int_{-\pi}^{\pi}\frac{1+ze^{-i\alpha}}{1-ze^{-i\alpha}}u(e^{i\alpha})d\alpha$$

which shows $u=\operatorname{Re}F$ where

$$F(z):=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1+ze^{-i\alpha}}{1-ze^{-i\alpha}}u(e^{i\alpha})d\alpha.$$

Moreover it follows from Eq. (23.23) that

$$\begin{aligned}\operatorname{Im}F(re^{i\theta})&=\frac{1}{\pi}\operatorname{Im}\int_{-\pi}^{\pi}\frac{r\sin(\theta-\alpha)}{1-2r\cos(\theta-\alpha)+r^2}g(e^{i\alpha})d\alpha\\&=:\left(Q_r*u\right)\left(e^{i\theta}\right)\end{aligned}$$

where

$$Q_r(\delta):=\frac{r\sin(\delta)}{1-2r\cos(\delta)+r^2}.$$

From these remarks it follows that $v=:\left(Q_r*g\right)\left(e^{i\theta}\right)$ is the harmonic conjugate of u and $\tilde{P}_r=Q_r$. For more on this point see Section **??** below.

23.4 Weak L^2-Derivatives

Theorem 23.14 (Weak and Strong Differentiability). *Suppose that $f\in L^2(\mathbb{R}^n)$ and $v\in\mathbb{R}^n\setminus\{0\}$. Then the following are equivalent:*

1. *There exists $\{t_n\}_{n=1}^\infty\subset\mathbb{R}\setminus\{0\}$ such that $\lim_{n\to\infty}t_n=0$ and*

$$\sup_n\left\|\frac{f(\cdot+t_nv)-f(\cdot)}{t_n}\right\|_2<\infty.$$

2. *There exists $g\in L^2(\mathbb{R}^n)$ such that $\langle f,\partial_v\phi\rangle=-\langle g,\phi\rangle$ for all $\phi\in C_c^\infty(\mathbb{R}^n)$.*
3. *There exists $g\in L^2(\mathbb{R}^n)$ and $f_n\in C_c^\infty(\mathbb{R}^n)$ such that $f_n\overset{L^2}{\to}f$ and $\partial_vf_n\overset{L^2}{\to}g$ as $n\to\infty$.*
4. *There exists $g\in L^2$ such that*

$$\frac{f(\cdot+tv)-f(\cdot)}{t}\overset{L^2}{\to}g\text{ as }t\to 0.$$

(See Theorem **??** *for the L^p generalization of this theorem.)*

Proof. 1. $\implies$ 2. We may assume, using Theorem 14.52 and passing to a subsequence if necessary, that $\frac{f(\cdot+t_nv)-f(\cdot)}{t_n} \xrightarrow{w} g$ for some $g \in L^2(\mathbb{R}^n)$. Now for $\phi \in C_c^\infty(\mathbb{R}^n)$,

$$\begin{aligned}\langle g|\phi\rangle &= \lim_{n\to\infty}\left\langle \frac{f(\cdot+t_nv)-f(\cdot)}{t_n},\phi\right\rangle = \lim_{n\to\infty}\left\langle f,\frac{\phi(\cdot-t_nv)-\phi(\cdot)}{t_n}\right\rangle \\ &= \left\langle f,\lim_{n\to\infty}\frac{\phi(\cdot-t_nv)-\phi(\cdot)}{t_n}\right\rangle = -\langle f,\partial_v\phi\rangle,\end{aligned}$$

wherein we have used the translation invariance of Lebesgue measure and the dominated convergence theorem. 2. $\implies$ 3. Let $\phi \in C_c^\infty(\mathbb{R}^n,\mathbb{R})$ such that $\int_{\mathbb{R}^n}\phi(x)dx = 1$ and let $\phi_m(x) = m^n\phi(mx)$, then by Proposition 22.34, $h_m := \phi_m * f \in C^\infty(\mathbb{R}^n)$ for all m and

$$\begin{aligned}\partial_v h_m(x) = \partial_v\phi_m * f(x) &= \int_{\mathbb{R}^n}\partial_v\phi_m(x-y)f(y)dy = \langle f,-\partial_v\left[\phi_m(x-\cdot)\right]\rangle \\ &= \langle g,\phi_m(x-\cdot)\rangle = \phi_m * g(x).\end{aligned}$$

By Theorem 22.32, $h_m \to f \in L^2(\mathbb{R}^n)$ and $\partial_v h_m = \phi_m * g \to g$ in $L^2(\mathbb{R}^n)$ as $m\to\infty$. This shows 3. holds except for the fact that h_m need not have compact support. To fix this let $\psi \in C_c^\infty(\mathbb{R}^n,[0,1])$ such that $\psi = 1$ in a neighborhood of 0 and let $\psi_\varepsilon(x) = \psi(\varepsilon x)$ and $(\partial_v\psi)_\varepsilon(x) := (\partial_v\psi)(\varepsilon x)$. Then

$$\partial_v(\psi_\varepsilon h_m) = \partial_v\psi_\varepsilon h_m + \psi_\varepsilon\partial_v h_m = \varepsilon(\partial_v\psi)_\varepsilon h_m + \psi_\varepsilon\partial_v h_m$$

so that $\psi_\varepsilon h_m \to h_m$ in L^2 and $\partial_v(\psi_\varepsilon h_m) \to \partial_v h_m$ in L^2 as $\varepsilon \downarrow 0$. Let $f_m = \psi_{\varepsilon_m}h_m$ where ε_m is chosen to be greater than zero but small enough so that

$$\left\|\psi_{\varepsilon_m}h_m - h_m\right\|_2 + \left\|\partial_v(\psi_{\varepsilon_m}h_m) \to \partial_v h_m\right\|_2 < 1/m.$$

Then $f_m \in C_c^\infty(\mathbb{R}^n)$, $f_m \to f$ and $\partial_v f_m \to g$ in L^2 as $m\to\infty$. 3. $\implies$ 4. By the fundamental theorem of calculus

$$\begin{aligned}\frac{\tau_{-tv}f_m(x)-f_m(x)}{t} &= \frac{f_m(x+tv)-f_m(x)}{t} \\ &= \frac{1}{t}\int_0^1\frac{d}{ds}f_m(x+stv)ds = \int_0^1(\partial_v f_m)(x+stv)ds.\end{aligned} \tag{23.24}$$

Let

$$G_t(x) := \int_0^1\tau_{-stv}g(x)ds = \int_0^1 g(x+stv)ds$$

which is defined for almost every x and is in $L^2(\mathbb{R}^n)$ by Minkowski's inequality for integrals, Theorem 21.27. Therefore

$$\frac{\tau_{-tv} f_m(x) - f_m(x)}{t} - G_t(x) = \int_0^1 \left[(\partial_v f_m)(x + stv) - g(x + stv) \right] ds$$

and hence again by Minkowski's inequality for integrals,

$$\left\| \frac{\tau_{-tv} f_m - f_m}{t} - G_t \right\|_2 \leq \int_0^1 \left\| \tau_{-stv} (\partial_v f_m) - \tau_{-stv} g \right\|_2 ds$$
$$= \int_0^1 \left\| \partial_v f_m - g \right\|_2 ds.$$

Letting $m \to \infty$ in this equation implies $(\tau_{-tv} f - f)/t = G_t$ a.e. Finally one more application of Minkowski's inequality for integrals implies,

$$\left\| \frac{\tau_{-tv} f - f}{t} - g \right\|_2 = \left\| G_t - g \right\|_2 = \left\| \int_0^1 (\tau_{-stv} g - g)\, ds \right\|_2$$
$$\leq \int_0^1 \left\| \tau_{-stv} g - g \right\|_2 ds.$$

By the dominated convergence theorem and Proposition 22.24, the latter term tends to 0 as $t \to 0$ and this proves 4. The proof is now complete since 4. $\Longrightarrow$ 1. is trivial. ■

23.5 *Conditional Expectation

In this section let $(\Omega, \mathcal{F}, P)$ be a probability space, i.e. $(\Omega, \mathcal{F}, P)$ is a measure space and $P(\Omega) = 1$. Let $\mathcal{G} \subset \mathcal{F}$ be a sub – sigma algebra of $\mathcal{F}$ and write $f \in \mathcal{G}_b$ if $f : \Omega \to \mathbb{C}$ is bounded and f is $(\mathcal{G}, \mathcal{B}_{\mathbb{C}})$ – measurable. In this section we will write

$$Ef := \int_\Omega f dP.$$

Definition 23.15 (Conditional Expectation). *Let $E_\mathcal{G} : L^2(\Omega, \mathcal{F}, P) \to L^2(\Omega, \mathcal{G}, P)$ denote orthogonal projection of $L^2(\Omega, \mathcal{F}, P)$ onto the closed subspace $L^2(\Omega, \mathcal{G}, P)$. For $f \in L^2(\Omega, \mathcal{G}, P)$, we say that $E_\mathcal{G} f \in L^2(\Omega, \mathcal{F}, P)$ is the* ***conditional expectation*** *of f.*

Theorem 23.16. *Let $(\Omega, \mathcal{F}, P)$ and $\mathcal{G} \subset \mathcal{F}$ be as above and $f, g \in L^2(\Omega, \mathcal{F}, P)$.*

1. *If $f \geq 0$, P – a.e. then $E_\mathcal{G} f \geq 0$, P – a.e.*
2. *If $f \geq g$, P – a.e. there $E_\mathcal{G} f \geq E_\mathcal{G} g$, P – a.e.*
3. *$|E_\mathcal{G} f| \leq E_\mathcal{G} |f|$, P – a.e.*
4. *$\|E_\mathcal{G} f\|_{L^1} \leq \|f\|_{L^1}$ for all $f \in L^2$. So by the B.L.T. Theorem 8.4, $E_\mathcal{G}$ extends uniquely to a bounded linear map from $L^1(\Omega, \mathcal{F}, P)$ to $L^1(\Omega, \mathcal{G}, P)$ which we will still denote by $E_\mathcal{G}$.*

5. *If $f \in L^1(\Omega, \mathcal{F}, P)$ then $F = E_{\mathcal{G}} f \in L^1(\Omega, \mathcal{G}, P)$ iff*

$$E(Fh) = E(fh) \text{ for all } h \in \mathcal{G}_b.$$

6. *If $g \in \mathcal{G}_b$ and $f \in L^1(\Omega, \mathcal{F}, P)$, then $E_{\mathcal{G}}(gf) = g \cdot E_{\mathcal{G}} f$, P – a.e.*

Proof. By the definition of orthogonal projection for $h \in \mathcal{G}_b$,

$$E(fh) = E(f \cdot E_{\mathcal{G}} h) = E(E_{\mathcal{G}} f \cdot h).$$

So if $f, h \geq 0$ then $0 \leq E(fh) \leq E(E_{\mathcal{G}} f \cdot h)$ and since this holds for all $h \geq 0$ in $\mathcal{G}_b$, $E_{\mathcal{G}} f \geq 0$, P – a.e. This proves (1). Item (2) follows by applying item (1). to $f - g$. If f is real, $\pm f \leq |f|$ and so by Item (2), $\pm E_{\mathcal{G}} f \leq E_{\mathcal{G}} |f|$, i.e. $|E_{\mathcal{G}} f| \leq E_{\mathcal{G}} |f|$, P – a.e. For complex f, let $h \geq 0$ be a bounded and $\mathcal{G}$ – measurable function. Then

$$\begin{aligned} E[|E_{\mathcal{G}} f| h] &= E\left[E_{\mathcal{G}} f \cdot \overline{\operatorname{sgn}(E_{\mathcal{G}} f)} h\right] = E\left[f \cdot \overline{\operatorname{sgn}(E_{\mathcal{G}} f)} h\right] \\ &\leq E[|f| h] = E[E_{\mathcal{G}} |f| \cdot h]. \end{aligned}$$

Since h is arbitrary, it follows that $|E_{\mathcal{G}} f| \leq E_{\mathcal{G}} |f|$, P – a.e. Integrating this inequality implies

$$\|E_{\mathcal{G}} f\|_{L^1} \leq E |E_{\mathcal{G}} f| \leq E[E_{\mathcal{G}} |f| \cdot 1] = E[|f|] = \|f\|_{L^1}.$$

Item (5). Suppose $f \in L^1(\Omega, \mathcal{F}, P)$ and $h \in \mathcal{G}_b$. Let $f_n \in L^2(\Omega, \mathcal{F}, P)$ be a sequence of functions such that $f_n \to f$ in $L^1(\Omega, \mathcal{F}, P)$. Then

$$\begin{aligned} E(E_{\mathcal{G}} f \cdot h) &= E(\lim_{n \to \infty} E_{\mathcal{G}} f_n \cdot h) = \lim_{n \to \infty} E(E_{\mathcal{G}} f_n \cdot h) \\ &= \lim_{n \to \infty} E(f_n \cdot h) = E(f \cdot h). \end{aligned} \tag{23.25}$$

This equation uniquely determines $E_{\mathcal{G}}$, for if $F \in L^1(\Omega, \mathcal{G}, P)$ also satisfies $E(F \cdot h) = E(f \cdot h)$ for all $h \in \mathcal{G}_b$, then taking $h = \overline{\operatorname{sgn}(F - E_{\mathcal{G}} f)}$ in Eq. (23.25) gives

$$0 = E((F - E_{\mathcal{G}} f) h) = E(|F - E_{\mathcal{G}} f|).$$

This shows $F = E_{\mathcal{G}} f$, P – a.e. Item (6) is now an easy consequence of this characterization, since if $h \in \mathcal{G}_b$,

$$E[(g E_{\mathcal{G}} f) h] = E[E_{\mathcal{G}} f \cdot hg] = E[f \cdot hg] = E[gf \cdot h] = E[E_{\mathcal{G}}(gf) \cdot h].$$

Thus $E_{\mathcal{G}}(gf) = g \cdot E_{\mathcal{G}} f$, P – a.e. ■

Proposition 23.17. *If $\mathcal{G}_0 \subset \mathcal{G}_1 \subset \mathcal{F}$. Then*

$$E_{\mathcal{G}_0} E_{\mathcal{G}_1} = E_{\mathcal{G}_1} E_{\mathcal{G}_0} = E_{\mathcal{G}_0}. \tag{23.26}$$

Proof. Equation (23.26) holds on $L^2(\Omega, \mathcal{F}, P)$ by the basic properties of orthogonal projections. It then hold on $L^1(\Omega, \mathcal{F}, P)$ by continuity and the density of $L^2(\Omega, \mathcal{F}, P)$ in $L^1(\Omega, \mathcal{F}, P)$. ■

Example 23.18. Suppose that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are two σ – finite measure spaces. Let $\Omega = X \times Y$, $\mathcal{F} = \mathcal{M} \otimes \mathcal{N}$ and $P(dx, dy) = \rho(x, y)\mu(dx)\nu(dy)$ where $\rho \in L^1(\Omega, \mathcal{F}, \mu \otimes \nu)$ is a positive function such that $\int_{X \times Y} \rho d(\mu \otimes \nu) = 1$. Let $\pi_X : \Omega \to X$ be the projection map, $\pi_X(x, y) = x$, and

$$\mathcal{G} := \sigma(\pi_X) = \pi_X^{-1}(\mathcal{M}) = \{A \times Y : A \in \mathcal{M}\}.$$

Then $f : \Omega \to \mathbb{R}$ is $\mathcal{G}$ – measurable iff $f = F \circ \pi_X$ for some function $F : X \to \mathbb{R}$ which is $\mathcal{N}$ – measurable, see Lemma 18.66. For $f \in L^1(\Omega, \mathcal{F}, P)$, we will now show $E_{\mathcal{G}} f = F \circ \pi_X$ where

$$F(x) = \frac{1}{\bar{\rho}(x)} 1_{(0,\infty)}(\bar{\rho}(x)) \cdot \int_Y f(x, y)\rho(x, y)\nu(dy),$$

$\bar{\rho}(x) := \int_Y \rho(x, y)\nu(dy)$. (By convention, $\int_Y f(x, y)\rho(x, y)\nu(dy) := 0$ if $\int_Y |f(x, y)| \rho(x, y)\nu(dy) = \infty$.)

By Tonelli's theorem, the set

$$E := \{x \in X : \bar{\rho}(x) = \infty\} \cup \left\{x \in X : \int_Y |f(x, y)| \rho(x, y)\nu(dy) = \infty\right\}$$

is a μ – null set. Since

$$\begin{aligned}
E[|F \circ \pi_X|] &= \int_X d\mu(x) \int_Y d\nu(y) |F(x)| \rho(x, y) = \int_X d\mu(x) |F(x)| \bar{\rho}(x) \\
&= \int_X d\mu(x) \left| \int_Y \nu(dy) f(x, y)\rho(x, y) \right| \\
&\leq \int_X d\mu(x) \int_Y \nu(dy) |f(x, y)| \rho(x, y) < \infty,
\end{aligned}$$

$F \circ \pi_X \in L^1(\Omega, \mathcal{G}, P)$. Let $h = H \circ \pi_X$ be a bounded $\mathcal{G}$ – measurable function, then

$$\begin{aligned}
E[F \circ \pi_X \cdot h] &= \int_X d\mu(x) \int_Y d\nu(y) F(x)H(x)\rho(x, y) \\
&= \int_X d\mu(x) F(x)H(x)\bar{\rho}(x) \\
&= \int_X d\mu(x)H(x) \int_Y \nu(dy) f(x, y)\rho(x, y) \\
&= E[hf]
\end{aligned}$$

and hence $E_{\mathcal{G}} f = F \circ \pi_X$ as claimed.

This example shows that conditional expectation is a generalization of the notion of performing integration over a partial subset of the variables in the integrand. Whereas to compute the expectation, one should integrate over all of the variables. See also Exercise 23.25 to gain more intuition about conditional expectations.

Theorem 23.19 (Jensen's inequality). *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function. Assume $f \in L^1(\Omega, \mathcal{F}, P; \mathbb{R})$ is a function such that (for simplicity) $\varphi(f) \in L^1(\Omega, \mathcal{F}, P; \mathbb{R})$, then $\varphi(E_{\mathcal{G}} f) \leq E_{\mathcal{G}} [\varphi(f)]$, P – a.e.*

Proof. Let us first assume that ϕ is C^1 and f is bounded. In this case

$$\varphi(x) - \varphi(x_0) \geq \varphi'(x_0)(x - x_0) \text{ for all } x_0, x \in \mathbb{R}. \tag{23.27}$$

Taking $x_0 = E_{\mathcal{G}} f$ and $x = f$ in this inequality implies

$$\varphi(f) - \varphi(E_{\mathcal{G}} f) \geq \varphi'(E_{\mathcal{G}} f)(f - E_{\mathcal{G}} f)$$

and then applying $E_{\mathcal{G}}$ to this inequality gives

$$\begin{aligned} E_{\mathcal{G}} [\varphi(f)] - \varphi(E_{\mathcal{G}} f) &= E_{\mathcal{G}}[\varphi(f) - \varphi(E_{\mathcal{G}} f)] \\ &\geq \varphi'(E_{\mathcal{G}} f)(E_{\mathcal{G}} f - E_{\mathcal{G}} E_{\mathcal{G}} f) = 0 \end{aligned}$$

The same proof works for general ϕ, one need only use Proposition 21.8 to replace Eq. (23.27) by

$$\varphi(x) - \varphi(x_0) \geq \varphi'_-(x_0)(x - x_0) \text{ for all } x_0, x \in \mathbb{R}$$

where $\varphi'_-(x_0)$ is the left hand derivative of ϕ at x_0. If f is not bounded, apply what we have just proved to $f^M = f 1_{|f| \leq M}$, to find

$$E_{\mathcal{G}} \left[\varphi(f^M)\right] \geq \varphi(E_{\mathcal{G}} f^M). \tag{23.28}$$

Since $E_{\mathcal{G}} : L^1(\Omega, \mathcal{F}, P; \mathbb{R}) \to L^1(\Omega, \mathcal{F}, P; \mathbb{R})$ is a bounded operator and $f^M \to f$ and $\varphi(f^M) \to \phi(f)$ in $L^1(\Omega, \mathcal{F}, P; \mathbb{R})$ as $M \to \infty$, there exists $\{M_k\}_{k=1}^\infty$ such that $M_k \uparrow \infty$ and $f^{M_k} \to f$ and $\varphi(f^{M_k}) \to \phi(f)$, P – a.e. So passing to the limit in Eq. (23.28) shows $E_{\mathcal{G}} [\varphi(f)] \geq \varphi(E_{\mathcal{G}} f)$, P – a.e. ■

23.6 Exercises

Exercise 23.9. Let $(X, \mathcal{M}, \mu)$ be a measure space and $H := L^2(X, \mathcal{M}, \mu)$. Given $f \in L^\infty(\mu)$ let $M_f : H \to H$ be the multiplication operator defined by $M_f g = fg$. Show $M_f^2 = M_f$ iff there exists $A \in \mathcal{M}$ such that $f = 1_A$ a.e.

Exercise 23.10 (Haar Basis). In this problem, let L^2 denote $L^2([0,1], m)$ with the standard inner product,

$$\psi(x) = 1_{[0,1/2)}(x) - 1_{[1/2,1)}(x)$$

and for $k, j \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ with $0 \leq j < 2^k$ let

$$\psi_{kj}(x) = 2^{k/2}\psi(2^k x - j)$$
$$= 2^{k/2}\left(1_{2^{-k}[j,j+1/2)}(x) - 1_{2^{-k}[j+1/2,j+1)}(x)\right).$$

The following pictures shows the graphs of ψ_{00}, $\psi_{1,0}, \psi_{1,1}, \psi_{2,1}, \psi_{2,2}$ and $\psi_{2,3}$ respectively.

Plot of $\psi_0, 0$.

Plot of $\psi_1 0$.

Plot of $\psi_1 1$.

Plot of $\psi_2 0$.

Plot of $\psi_2 1$.

Plot of $\psi_2 2$.

Plot of $\psi_2 3$.

1. For $n \in \mathbb{N}$, let $M_0 = \operatorname{span}(\{\mathbf{1}\})$ and $M_n := \operatorname{span}\left(\{\mathbf{1}\} \cup \left\{\psi_{kj} : 0 \le k < n \text{ and } 0 \le j < 2^k\right\}\right)$ for $n \in \mathbb{N}$, where $\mathbf{1}$ denotes the constant function 1. Show
$$M_n = \operatorname{span}\left(\{1_{[j2^{-n},(j+1)2^{-n})} : \text{and } 0 \le j < 2^n\right).$$

2. Show $\beta := \{\mathbf{1}\} \cup \left\{\psi_{kj} : 0 \le k \text{ and } 0 \le j < 2^k\right\}$ is an orthonormal set. **Hint:** show $\psi_{k+1,j} \in M_k^\perp$ for all $0 \le j < 2^{k+1}$ and show $\left\{\psi_{kj} : 0 \le j < 2^k\right\}$ is an orthonormal set for fixed k.
3. Show $\cup_{n=1}^\infty M_n$ is a dense subspace of L^2 and therefore β is an orthonormal basis for L^2. **Hint:** see Theorem 22.15.
4. For $f \in L^2$, let

$$H_n f := \langle f|\mathbf{1}\rangle \mathbf{1} + \sum_{k=0}^{n-1} \sum_{j=0}^{2^k-1} \langle f|\psi_{kj}\rangle \psi_{kj}.$$

Show (compare with Exercise 23.25)

$$H_n f = \sum_{j=0}^{2^n-1} \left(2^n \int_{j2^{-n}}^{(j+1)2^{-n}} f(x)dx\right) 1_{[j2^{-n},(j+1)2^{-n})}$$

and use this to show $\|f - H_n f\|_\infty \to 0$ as $n \to \infty$ for all $f \in C([0,1])$. **Hint**: Compute orthogonal projection onto M_n using a judiciously chosen basis for M_n.

Exercise 23.11. Let $O(n)$ be the orthogonal groups consisting of $n \times n$ real orthogonal matrices O, i.e. $O^{tr}O = I$. For $O \in O(n)$ and $f \in L^2(\mathbb{R}^n)$ let $U_O f(x) = f(O^{-1}x)$. Show

1. $U_O f$ is well defined, namely if $f = g$ a.e. then $U_O f = U_O g$ a.e.
2. $U_O : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ is unitary and satisfies $U_{O_1}U_{O_2} = U_{O_1 O_2}$ for all $O_1, O_2 \in O(n)$. That is to say the map $O \in O(n) \to \mathcal{U}(L^2(\mathbb{R}^n))$ – the unitary operators on $L^2(\mathbb{R}^n)$ is a group homomorphism, i.e. a "unitary representation" of $O(n)$.
3. For each $f \in L^2(\mathbb{R}^n)$, the map $O \in O(n) \to U_O f \in L^2(\mathbb{R}^n)$ is continuous. Take the topology on $O(n)$ to be that inherited from the Euclidean topology on the vector space of all $n \times n$ matrices. **Hint**: see the proof of Proposition 22.24.

Exercise 23.12. Euclidean group representation and its infinitesimal generators including momentum and angular momentum operators.

Exercise 23.13. Spherical Harmonics.

Exercise 23.14. The gradient and the Laplacian in spherical coordinates.

Exercise 23.15. Legendre polynomials.

23.7 Fourier Series Exercises

Exercise 23.16. Show $\sum_{k=1}^\infty k^{-2} = \pi^2/6$, by taking $f(x) = x$ on $[-\pi, \pi]$ and computing $\|f\|_2^2$ directly and then in terms of the Fourier Coefficients $\tilde{f}$ of f.

Exercise 23.17 (Riemann Lebesgue Lemma for Fourier Series). Show for $f \in L^1([-\pi,\pi]^d)$ that $\tilde{f} \in c_0(\mathbb{Z}^d)$, i.e. $\tilde{f} : \mathbb{Z}^d \to \mathbb{C}$ and $\lim_{k\to\infty} \tilde{f}(k) = 0$. **Hint:** If $f \in H$, this follows form Bessel's inequality. Now use a density argument.

Exercise 23.18. Suppose $f \in L^1([-\pi,\pi]^d)$ is a function such that $\tilde{f} \in \ell^1(\mathbb{Z}^d)$ and set

$$g(x) := \sum_{k\in\mathbb{Z}^d} \tilde{f}(k) e^{ik\cdot x} \text{ (pointwise).}$$

1. Show $g \in C_{per}(\mathbb{R}^d)$.
2. Show $g(x) = f(x)$ for m – a.e. x in $[-\pi,\pi]^d$. **Hint**: Show $\tilde{g}(k) = \tilde{f}(k)$ and then use approximation arguments to show
$$\int_{[-\pi,\pi]^d} f(x)h(x)dx = \int_{[-\pi,\pi]^d} g(x)h(x)dx \;\forall\; h \in C([-\pi,\pi]^d).$$
3. Conclude that $f \in L^1([-\pi,\pi]^d) \cap L^\infty([-\pi,\pi]^d)$ and in particular $f \in L^p([-\pi,\pi]^d)$ for all $p \in [1,\infty]$.

Exercise 23.19. Suppose $m \in \mathbb{N}_0$, α is a multi-index such that $|\alpha| \le 2m$ and $f \in C^{2m}_{per}(\mathbb{R}^d)$[1].

1. Using integration by parts, show (using Notation 22.21) that
$$(ik)^\alpha \tilde{f}(k) = \langle \partial^\alpha f | e_k \rangle \text{ for all } k \in \mathbb{Z}^d.$$
Note: This equality implies
$$\left|\tilde{f}(k)\right| \le \frac{1}{k^\alpha} \|\partial^\alpha f\|_H \le \frac{1}{k^\alpha} \|\partial^\alpha f\|_\infty .$$
2. Now let $\Delta f = \sum_{i=1}^d \partial^2 f/\partial x_i^2$, Working as in part 1) show
$$\langle (1-\Delta)^m f | e_k \rangle = (1+|k|^2)^m \tilde{f}(k). \tag{23.29}$$

Remark 23.20. Suppose that m is an even integer, α is a multi-index and $f \in C^{m+|\alpha|}_{per}(\mathbb{R}^d)$, then

[1] We view $C_{per}(\mathbb{R})$ as a subspace of $H = L^2([-\pi,\pi])$ by identifying $f \in C_{per}(\mathbb{R})$ with $f|_{[-\pi,\pi]} \in H$.

$$\begin{aligned}\left(\sum_{k\in\mathbb{Z}^d}|k^\alpha|\left|\tilde f(k)\right|\right)^2 &= \left(\sum_{k\in\mathbb{Z}^d}|\langle\partial^\alpha f|e_k\rangle|\,(1+|k|^2)^{m/2}(1+|k|^2)^{-m/2}\right)^2\\ &= \left(\sum_{k\in\mathbb{Z}^d}\left|\langle(1-\Delta)^{m/2}\partial^\alpha f|e_k\rangle\right|(1+|k|^2)^{-m/2}\right)^2\\ &\le \sum_{k\in\mathbb{Z}^d}\left|\langle(1-\Delta)^{m/2}\partial^\alpha f|e_k\rangle\right|^2\cdot\sum_{k\in\mathbb{Z}^d}(1+|k|^2)^{-m}\\ &= C_m\left\|(1-\Delta)^{m/2}\partial^\alpha f\right\|_H^2\end{aligned}$$

where $C_m := \sum_{k\in\mathbb{Z}^d}(1+|k|^2)^{-m} < \infty$ iff $m > d/2$. So the smoother f is the faster $\tilde f$ decays at infinity. The next problem is the converse of this assertion and hence smoothness of f corresponds to decay of $\tilde f$ at infinity and visa-versa.

Exercise 23.20 (A Sobolev Imbedding Theorem). Suppose $s\in\mathbb{R}$ and $\left\{c_k\in\mathbb{C}: k\in\mathbb{Z}^d\right\}$ are coefficients such that

$$\sum_{k\in\mathbb{Z}^d}|c_k|^2\,(1+|k|^2)^s < \infty.$$

Show if $s > \frac{d}{2}+m$, the function f defined by

$$f(x) = \sum_{k\in\mathbb{Z}^d} c_k e^{ik\cdot x}$$

is in $C^m_{per}(\mathbb{R}^d)$. **Hint**: Work as in the above remark to show

$$\sum_{k\in\mathbb{Z}^d}|c_k|\,|k^\alpha| < \infty \text{ for all } |\alpha|\le m.$$

Exercise 23.21 (Poisson Summation Formula). Let $F\in L^1(\mathbb{R}^d)$,

$$E := \left\{x\in\mathbb{R}^d : \sum_{k\in\mathbb{Z}^d}|F(x+2\pi k)| = \infty\right\}$$

and set

$$\hat F(k) := (2\pi)^{-d/2}\int_{\mathbb{R}^d} F(x)e^{-ik\cdot x}dx.$$

Further assume $\hat F\in\ell^1(\mathbb{Z}^d)$.

1. Show $m(E) = 0$ and $E + 2\pi k = E$ for all $k\in\mathbb{Z}^d$. **Hint**: Compute $\int_{[-\pi,\pi]^d}\sum_{k\in\mathbb{Z}^d}|F(x+2\pi k)|\,dx$.

2. Let
$$f(x) := \begin{cases} \sum_{k\in\mathbb{Z}^d} F(x+2\pi k) & \text{for } \ x \notin E \\ 0 & \text{if } \ x \in E. \end{cases}$$
Show $f \in L^1([-\pi,\pi]^d)$ and $\tilde{f}(k) = (2\pi)^{-d/2}\,\hat{F}(k)$.
3. Using item 2) and the assumptions on F, show $f \in L^1([-\pi,\pi]^d) \cap L^\infty([-\pi,\pi]^d)$ and
$$f(x) = \sum_{k\in\mathbb{Z}^d} \tilde{f}(k)e^{ik\cdot x} = \sum_{k\in\mathbb{Z}^d} (2\pi)^{-d/2}\,\hat{F}(k)e^{ik\cdot x} \text{ for } m-\text{a.e. } x,$$
i.e.
$$\sum_{k\in\mathbb{Z}^d} F(x+2\pi k) = (2\pi)^{-d/2}\sum_{k\in\mathbb{Z}^d}\hat{F}(k)e^{ik\cdot x} \text{ for } m-\text{a.e. } x. \tag{23.30}$$
4. Suppose we now assume that $F \in C(\mathbb{R}^d)$ and F satisfies 1) $|F(x)| \le C(1+|x|)^{-s}$ for some $s > d$ and $C < \infty$ and 2) $\hat{F} \in \ell^1(\mathbb{Z}^d)$, then show Eq. (23.30) holds for **all** $x \in \mathbb{R}^d$ and in particular
$$\sum_{k\in\mathbb{Z}^d} F(2\pi k) = (2\pi)^{-d/2}\sum_{k\in\mathbb{Z}^d}\hat{F}(k).$$

For notational simplicity, in the remaining problems we will assume that $d = 1$.

Exercise 23.22 (Heat Equation 1.). Let $(t,x) \in [0,\infty)\times\mathbb{R} \to u(t,x)$ be a continuous function such that $u(t,\cdot) \in C_{per}(\mathbb{R})$ for all $t \ge 0$, $\dot{u} := u_t$, u_x, and u_{xx} exists and are continuous when $t > 0$. Further assume that u satisfies the heat equation $\dot{u} = \frac{1}{2}u_{xx}$. Let $\tilde{u}(t,k) := \langle u(t,\cdot)|e_k\rangle$ for $k \in \mathbb{Z}$. Show for $t > 0$ and $k \in \mathbb{Z}$ that $\tilde{u}(t,k)$ is differentiable in t and $\frac{d}{dt}\tilde{u}(t,k) = -k^2\tilde{u}(t,k)/2$. Use this result to show
$$u(t,x) = \sum_{k\in\mathbb{Z}} e^{-\frac{t}{2}k^2}\tilde{f}(k)e^{ikx} \tag{23.31}$$
where $f(x) := u(0,x)$ and as above
$$\tilde{f}(k) = \langle f|e_k\rangle = \int_{-\pi}^{\pi} f(y)e^{-iky}dy = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)e^{-iky}dm\,(y)\,.$$
Notice from Eq. (23.31) that $(t,x) \to u(t,x)$ is C^∞ for $t > 0$.

Exercise 23.23 (Heat Equation 2.). Let $q_t(x) := \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} e^{-\frac{t}{2}k^2}e^{ikx}$. Show that Eq. (23.31) may be rewritten as
$$u(t,x) = \int_{-\pi}^{\pi} q_t(x-y)f(y)dy$$
and

$$q_t(x) = \sum_{k\in\mathbb{Z}} p_t(x + k2\pi)$$

where $p_t(x) := \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}x^2}$. Also show $u(t,x)$ may be written as

$$u(t,x) = p_t * f(x) := \int_{\mathbb{R}^d} p_t(x-y) f(y) dy.$$

Hint: To show $q_t(x) = \sum_{k\in\mathbb{Z}} p_t(x+k2\pi)$, use the Poisson summation formula and the Gaussian integration identity,

$$\hat{p}_t(\omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} p_t(x) e^{i\omega x} dx = \frac{1}{\sqrt{2\pi}} e^{-\frac{t}{2}\omega^2}. \tag{23.32}$$

Equation (23.32) will be discussed in Example **??** below.

Exercise 23.24 (Wave Equation). Let $u \in C^2(\mathbb{R}\times\mathbb{R})$ be such that $u(t,\cdot) \in C_{per}(\mathbb{R})$ for all $t \in \mathbb{R}$. Further assume that u solves the wave equation, $u_{tt} = u_{xx}$. Let $f(x) := u(0,x)$ and $g(x) = \dot{u}(0,x)$. Show $\tilde{u}(t,k) := \langle u(t,\cdot), e_k\rangle$ for $k \in \mathbb{Z}$ is twice continuously differentiable in t and $\frac{d^2}{dt^2}\tilde{u}(t,k) = -k^2\tilde{u}(t,k)$. Use this result to show

$$u(t,x) = \sum_{k\in\mathbb{Z}} \left(\tilde{f}(k)\cos(kt) + \tilde{g}(k)\frac{\sin kt}{k} \right) e^{ikx} \tag{23.33}$$

with the sum converging absolutely. Also show that $u(t,x)$ may be written as

$$u(t,x) = \frac{1}{2}\left[f(x+t) + f(x-t)\right] + \frac{1}{2}\int_{-t}^{t} g(x+\tau) d\tau. \tag{23.34}$$

Hint: To show Eq. (23.33) implies (23.34) use

$$\cos kt = \frac{e^{ikt} + e^{-ikt}}{2},$$
$$\sin kt = \frac{e^{ikt} - e^{-ikt}}{2i}, \text{ and}$$
$$\frac{e^{ik(x+t)} - e^{ik(x-t)}}{ik} = \int_{-t}^{t} e^{ik(x+\tau)} d\tau.$$

23.8 Conditional Expectation Exercises

Exercise 23.25. Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $\mathcal{A} := \{A_i\}_{i=1}^\infty \subset \mathcal{F}$ is a partition of Ω. (Recall this means $\Omega = \coprod_{i=1}^\infty A_i$.) Let $\mathcal{G}$ be the σ – algebra generated by $\mathcal{A}$. Show:

1. $B \in \mathcal{G}$ iff $B = \cup_{i\in\Lambda} A_i$ for some $\Lambda \subset \mathbb{N}$.

2. $g : \Omega \to \mathbb{R}$ is $\mathcal{G}$ – measurable iff $g = \sum_{i=1}^{\infty} \lambda_i 1_{A_i}$ for some $\lambda_i \in \mathbb{R}$.
3. For $f \in L^1(\Omega, \mathcal{F}, P)$, let $E(f|A_i) := E[1_{A_i} f] / P(A_i)$ if $P(A_i) \neq 0$ and $E(f|A_i) = 0$ otherwise. Show

$$E_{\mathcal{G}} f = \sum_{i=1}^{\infty} E(f|A_i) 1_{A_i}.$$

Part VI

Appendices

A

Multinomial Theorems and Calculus Results

Given a multi-index $\alpha \in \mathbb{Z}_+^n$, let $|\alpha| = \alpha_1 + \cdots + \alpha_n$, $\alpha! := \alpha_1! \cdots \alpha_n!$,

$$x^\alpha := \prod_{j=1}^n x_j^{\alpha_j} \text{ and } \partial_x^\alpha = \left(\frac{\partial}{\partial x}\right)^\alpha := \prod_{j=1}^n \left(\frac{\partial}{\partial x_j}\right)^{\alpha_j}.$$

We also write

$$\partial_v f(x) := \frac{d}{dt} f(x + tv)|_{t=0}.$$

A.1 Multinomial Theorems and Product Rules

For $a = (a_1, a_2, \ldots, a_n) \in \mathbb{C}^n$, $m \in \mathbb{N}$ and $(i_1, \ldots, i_m) \in \{1, 2, \ldots, n\}^m$ let $\hat{\alpha}_j (i_1, \ldots, i_m) = \# \{k : i_k = j\}$. Then

$$\left(\sum_{i=1}^n a_i\right)^m = \sum_{i_1, \ldots, i_m = 1}^n a_{i_1} \ldots a_{i_m} = \sum_{|\alpha| = m} C(\alpha) a^\alpha$$

where

$$C(\alpha) = \# \{(i_1, \ldots, i_m) : \hat{\alpha}_j (i_1, \ldots, i_m) = \alpha_j \text{ for } j = 1, 2, \ldots, n\}$$

I claim that $C(\alpha) = \frac{m!}{\alpha!}$. Indeed, one possibility for such a sequence $(a_1, \ldots, a_{i_m})$ for a given α is gotten by choosing

$$(\overbrace{a_1, \ldots, a_1}^{\alpha_1}, \overbrace{a_2, \ldots, a_2}^{\alpha_2}, \ldots, \overbrace{a_n, \ldots, a_n}^{\alpha_n}).$$

Now there are $m!$ permutations of this list. However, only those permutations leading to a distinct list are to be counted. So for each of these $m!$ permutations we must divide by the number of permutation which just rearrange the groups of a_i's among themselves for each i. There are $\alpha! := \alpha_1! \cdots \alpha_n!$ such permutations. Therefore, $C(\alpha) = m!/\alpha!$ as advertised. So we have proved

$$\left(\sum_{i=1}^{n} a_i\right)^m = \sum_{|\alpha|=m} \frac{m!}{\alpha!} a^\alpha. \tag{A.1}$$

Now suppose that $a, b \in \mathbb{R}^n$ and α is a multi-index, we have

$$(a+b)^\alpha = \sum_{\beta \le \alpha} \frac{\alpha!}{\beta!(\alpha-\beta)!} a^\beta b^{\alpha-\beta} = \sum_{\beta+\delta=\alpha} \frac{\alpha!}{\beta!\delta!} a^\beta b^\delta \tag{A.2}$$

Indeed, by the standard Binomial formula,

$$(a_i + b_i)^{\alpha_i} = \sum_{\beta_i \le \alpha_i} \frac{\alpha_i!}{\beta_i!(\alpha_i - \beta_i)!} a^{\beta_i} b^{\alpha_i - \beta_i}$$

from which Eq. (A.2) follows. Eq. (A.2) generalizes in the obvious way to

$$(a_1 + \cdots + a_k)^\alpha = \sum_{\beta_1 + \cdots + \beta_k = \alpha} \frac{\alpha!}{\beta_1! \cdots \beta_k!} a_1^{\beta_1} \dots a_k^{\beta_k} \tag{A.3}$$

where $a_1, a_2, \dots, a_k \in \mathbb{R}^n$ and $\alpha \in \mathbb{Z}_+^n$.

Now let us consider the product rule for derivatives. Let us begin with the one variable case (write $d^n f$ for $f^{(n)} = \frac{d^n}{dx^n} f$) where we will show by induction that

$$d^n(fg) = \sum_{k=0}^{n} \binom{n}{k} d^k f \cdot d^{n-k} g. \tag{A.4}$$

Indeed assuming Eq. (A.4) we find

$$\begin{aligned}
d^{n+1}(fg) &= \sum_{k=0}^{n} \binom{n}{k} d^{k+1} f \cdot d^{n-k} g + \sum_{k=0}^{n} \binom{n}{k} d^k f \cdot d^{n-k+1} g \\
&= \sum_{k=1}^{n+1} \binom{n}{k-1} d^k f \cdot d^{n-k+1} g + \sum_{k=0}^{n} \binom{n}{k} d^k f \cdot d^{n-k+1} g \\
&= \sum_{k=1}^{n+1} \left[\binom{n}{k-1} + \binom{n}{k} \right] d^k f \cdot d^{n-k+1} g + d^{n+1} f \cdot g + f \cdot d^{n+1} g.
\end{aligned}$$

Since

$$\begin{aligned}
\binom{n}{k-1} + \binom{n}{k} &= \frac{n!}{(n-k+1)!(k-1)!} + \frac{n!}{(n-k)!k!} \\
&= \frac{n!}{(k-1)!\,(n-k)!} \left[\frac{1}{(n-k+1)} + \frac{1}{k} \right] \\
&= \frac{n!}{(k-1)!\,(n-k)!} \frac{n+1}{(n-k+1)\,k} = \binom{n+1}{k}
\end{aligned}$$

the result follows.

Now consider the multi-variable case

$$\begin{aligned}\partial^\alpha (fg) &= \left(\prod_{i=1}^n \partial_i^{\alpha_i}\right)(fg) = \prod_{i=1}^n \left[\sum_{k_i=0}^{\alpha_i} \binom{\alpha_i}{k_i} \partial_i^{k_i} f \cdot \partial_i^{\alpha_i - k_i} g\right] \\ &= \sum_{k_1=0}^{\alpha_1} \cdots \sum_{k_n=0}^{\alpha_n} \prod_{i=1}^n \binom{\alpha_i}{k_i} \partial^k f \cdot \partial^{\alpha-k} g = \sum_{k \le \alpha} \binom{\alpha}{k} \partial^k f \cdot \partial^{\alpha-k} g\end{aligned}$$

where $k = (k_1, k_2, \ldots, k_n)$ and

$$\binom{\alpha}{k} := \prod_{i=1}^n \binom{\alpha_i}{k_i} = \frac{\alpha!}{k!(\alpha-k)!}.$$

So we have proved

$$\partial^\alpha (fg) = \sum_{\beta \le \alpha} \binom{\alpha}{\beta} \partial^\beta f \cdot \partial^{\alpha-\beta} g. \tag{A.5}$$

A.2 Taylor's Theorem

Theorem A.1. *Suppose $X \subset \mathbb{R}^n$ is an open set, $x : [0,1] \to X$ is a C^1 – path, and $f \in C^N(X, \mathbb{C})$. Let $v_s := x(1) - x(s)$ and $v = v_1 = x(1) - x(0)$, then*

$$f(x(1)) = \sum_{m=0}^{N-1} \frac{1}{m!} \left(\partial_v^m f\right)(x(0)) + R_N \tag{A.6}$$

where

$$R_N = \frac{1}{(N-1)!} \int_0^1 \left(\partial_{\dot{x}(s)} \partial_{v_s}^{N-1} f\right)(x(s)) ds = \frac{1}{N!} \int_0^1 \left(-\frac{d}{ds} \partial_{v_s}^N f\right)(x(s)) ds. \tag{A.7}$$

and $0! := 1$.

Proof. By the fundamental theorem of calculus and the chain rule,

$$f(x(t)) = f(x(0)) + \int_0^t \frac{d}{ds} f(x(s)) ds = f(x(0)) + \int_0^t \left(\partial_{\dot{x}(s)} f\right)(x(s)) ds \tag{A.8}$$

and in particular,

$$f(x(1)) = f(x(0)) + \int_0^1 \left(\partial_{\dot{x}(s)} f\right)(x(s)) ds.$$

This proves Eq. (A.6) when $N = 1$. We will now complete the proof using induction on N. Applying Eq. (A.8) with f replaced by $\frac{1}{(N-1)!}\left(\partial_{\dot{x}(s)} \partial_{v_s}^{N-1} f\right)$ gives

$$\begin{aligned}\frac{1}{(N-1)!}\left(\partial_{\dot{x}(s)}\partial_{v_s}^{N-1}f\right)(x(s)) &= \frac{1}{(N-1)!}\left(\partial_{\dot{x}(s)}\partial_{v_s}^{N-1}f\right)(x(0)) \\ &\quad + \frac{1}{(N-1)!}\int_0^s \left(\partial_{\dot{x}(s)}\partial_{v_s}^{N-1}\partial_{\dot{x}(t)}f\right)(x(t))dt \\ &= -\frac{1}{N!}\left(\frac{d}{ds}\partial_{v_s}^{N}f\right)(x(0)) - \frac{1}{N!}\int_0^s \left(\frac{d}{ds}\partial_{v_s}^{N}\partial_{\dot{x}(t)}f\right)(x(t))dt\end{aligned}$$

wherein we have used the fact that mixed partial derivatives commute to show $\frac{d}{ds}\partial_{v_s}^{N}f = N\partial_{\dot{x}(s)}\partial_{v_s}^{N-1}f$. Integrating this equation on $s \in [0,1]$ shows, using the fundamental theorem of calculus,

$$\begin{aligned}R_N &= \frac{1}{N!}\left(\partial_v^N f\right)(x(0)) - \frac{1}{N!}\int_{0\le t\le s\le 1}\left(\frac{d}{ds}\partial_{v_s}^{N}\partial_{\dot{x}(t)}f\right)(x(t))dsdt \\ &= \frac{1}{N!}\left(\partial_v^N f\right)(x(0)) + \frac{1}{(N+1)!}\int_{0\le t\le 1}\left(\partial_{w_t}^{N}\partial_{\dot{x}(t)}f\right)(x(t))dt \\ &= \frac{1}{N!}\left(\partial_v^N f\right)(x(0)) + R_{N+1}\end{aligned}$$

which completes the inductive proof. ■

Remark A.2. Using Eq. (A.1) with a_i replaced by $v_i\partial_i$ (although $\{v_i\partial_i\}_{i=1}^n$ are not complex numbers they are commuting symbols), we find

$$\partial_v^m f = \left(\sum_{i=1}^n v_i\partial_i\right)^m f = \sum_{|\alpha|=m}\frac{m!}{\alpha!}v^\alpha\partial^\alpha.$$

Using this fact we may write Eqs. (A.6) and (A.7) as

$$f(x(1)) = \sum_{|\alpha|\le N-1}\frac{1}{\alpha!}v^\alpha\partial^\alpha f(x(0)) + R_N$$

and

$$R_N = \sum_{|\alpha|=N}\frac{1}{\alpha!}\int_0^1\left(-\frac{d}{ds}v_s^\alpha\partial^\alpha f\right)(x(s))ds.$$

Corollary A.3. *Suppose $X \subset \mathbb{R}^n$ is an open set which contains $x(s) = (1-s)x_0 + sx_1$ for $0 \le s \le 1$ and $f \in C^N(X,\mathbb{C})$. Then*

$$\begin{aligned}f(x_1) &= \sum_{m=0}^{N-1}\frac{1}{m!}\left(\partial_v^m f\right)(x_0) + \frac{1}{N!}\int_0^1\left(\partial_v^N f\right)(x(s))d\nu_N(s) \qquad \text{(A.9)} \\ &= \sum_{|\alpha|<N}\frac{1}{\alpha!}\partial^\alpha f(x(0))(x_1-x_0)^\alpha + \sum_{\alpha:|\alpha|=N}\frac{1}{\alpha!}\left[\int_0^1\partial^\alpha f(x(s))d\nu_N(s)\right](x_1-x_0)^\alpha \qquad \text{(A.10)}\end{aligned}$$

where $v := x_1 - x_0$ *and* $d\nu_N$ *is the probability measure on* $[0,1]$ *given by*

$$d\nu_N(s) := N(1-s)^{N-1}ds. \tag{A.11}$$

If we let $x = x_0$ *and* $y = x_1 - x_0$ *(so* $x + y = x_1$*) Eq. (A.10) may be written as*

$$f(x+y) = \sum_{|\alpha|<N} \frac{\partial_x^\alpha f(x)}{\alpha!} y^\alpha + \sum_{\alpha:|\alpha|=N} \frac{1}{\alpha!} \left(\int_0^1 \partial_x^\alpha f(x+sy) d\nu_N(s) \right) y^\alpha. \tag{A.12}$$

Proof. This is a special case of Theorem A.1. Notice that

$$v_s = x(1) - x(s) = (1-s)(x_1 - x_0) = (1-s)v$$

and hence

$$R_N = \frac{1}{N!} \int_0^1 \left(-\frac{d}{ds}(1-s)^N \partial_v^N f \right) (x(s))ds = \frac{1}{N!} \int_0^1 \left(\partial_v^N f \right) (x(s)) N(1-s)^{N-1} ds.$$

■

Example A.4. Let $X = (-1,1) \subset \mathbb{R}$, $\beta \in \mathbb{R}$ and $f(x) = (1-x)^\beta$. The reader should verify

$$f^{(m)}(x) = (-1)^m \beta(\beta-1)\dots(\beta-m+1)(1-x)^{\beta-m}$$

and therefore by Taylor's theorem (Eq. (**??**) with $x = 0$ and $y = x$)

$$(1-x)^\beta = 1 + \sum_{m=1}^{N-1} \frac{1}{m!}(-1)^m \beta(\beta-1)\dots(\beta-m+1)x^m + R_N(x) \tag{A.13}$$

where

$$\begin{aligned} R_N(x) &= \frac{x^N}{N!} \int_0^1 (-1)^N \beta(\beta-1)\dots(\beta-N+1)(1-sx)^{\beta-N} d\nu_N(s) \\ &= \frac{x^N}{N!}(-1)^N \beta(\beta-1)\dots(\beta-N+1) \int_0^1 \frac{N(1-s)^{N-1}}{(1-sx)^{N-\beta}} ds. \end{aligned}$$

Now for $x \in (-1,1)$ and $N > \beta$,

$$0 \le \int_0^1 \frac{N(1-s)^{N-1}}{(1-sx)^{N-\beta}} ds \le \int_0^1 \frac{N(1-s)^{N-1}}{(1-s)^{N-\beta}} ds = \int_0^1 N(1-s)^{\beta-1} ds = \frac{N}{\beta}$$

and therefore,

$$|R_N(x)| \le \frac{|x|^N}{(N-1)!} |(\beta-1)\dots(\beta-N+1)| =: \rho_N.$$

Since

$$\limsup_{N\to\infty} \frac{\rho_{N+1}}{\rho_N} = |x| \cdot \limsup_{N\to\infty} \frac{N-\beta}{N} = |x| < 1$$

and so by the Ratio test, $|R_N(x)| \le \rho_N \to 0$ (exponentially fast) as $N \to \infty$. Therefore by passing to the limit in Eq. (A.13) we have proved

$$(1-x)^\beta = 1 + \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \beta(\beta-1)\dots(\beta-m+1)x^m \tag{A.14}$$

which is valid for $|x| < 1$ and $\beta \in \mathbb{R}$. An important special cases is $\beta = -1$ in which case, Eq. (A.14) becomes $\frac{1}{1-x} = \sum_{m=0}^{\infty} x^m$, the standard geometric series formula. Another another useful special case is $\beta = 1/2$ in which case Eq. (A.14) becomes

$$\begin{aligned}\sqrt{1-x} &= 1 + \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \frac{1}{2}(\frac{1}{2}-1)\dots(\frac{1}{2}-m+1)x^m \\ &= 1 - \sum_{m=1}^{\infty} \frac{(2m-3)!!}{2^m m!} x^m \text{ for all } |x| < 1. \end{aligned} \tag{A.15}$$

B

Zorn's Lemma and the Hausdorff Maximal Principle

Definition B.1. *A partial order $\leq$ on X is a relation with following properties*

(i) If $x \leq y$ and $y \leq z$ then $x \leq z$.
(ii) If $x \leq y$ and $y \leq x$ then $x = y$.
(iii) $x \leq x$ for all $x \in X$.

Example B.2. Let Y be a set and $X = 2^Y$. There are two natural partial orders on X.

1. Ordered by inclusion, $A \leq B$ is $A \subset B$ and
2. Ordered by reverse inclusion, $A \leq B$ if $B \subset A$.

Definition B.3. *Let $(X, \leq)$ be a partially ordered set we say X is* ***linearly*** *a* ***totally*** *ordered if for all $x, y \in X$ either $x \leq y$ or $y \leq x$. The real numbers $\mathbb{R}$ with the usual order $\leq$ is a typical example.*

Definition B.4. *Let $(X, \leq)$ be a partial ordered set. We say $x \in X$ is a* ***maximal*** *element if for all $y \in X$ such that $y \geq x$ implies $y = x$, i.e. there is no element larger than x. An* ***upper bound*** *for a subset E of X is an element $x \in X$ such that $x \geq y$ for all $y \in E$.*

Example B.5. Let

$$X = \left\{ a = \{1\} \; b = \{1, 2\} \; c = \{3\} \; d = \{2, 4\} \; e = \{2\} \right\}$$

ordered by set inclusion. Then b and d are maximal elements despite that fact that $b \nleq a$ and $a \nleq b$. We also have

- If $E = \{a, e, c\}$, then E has **no** upper bound.

Definition B.6. • *If $E = \{a, e\}$, then b is an upper bound.*
• *$E = \{e\}$, then b and d are upper bounds.*

Theorem B.7. *The following are equivalent.*

1. ***The axiom of choice:*** *to each collection,* $\{X_\alpha\}_{\alpha\in A}$ *, of non-empty sets there exists a "choice function,"* $x : A \to \coprod_{\alpha\in A} X_\alpha$ *such that* $x(\alpha) \in X_\alpha$ *for all* $\alpha \in A$*, i.e.* $\prod_{\alpha\in A} X_\alpha \neq \emptyset$.
2. ***The Hausdorff Maximal Principle****: Every partially ordered set has a* ***maximal*** *(relative to the inclusion order) linearly ordered subset.*
3. ***Zorn's Lemma****: If* X *is partially ordered set such that every linearly ordered subset of* X *has an upper bound, then* X *has a maximal element.*[1]

Proof. $(2 \Rightarrow 3)$ Let X be a partially ordered subset as in 3 and let $\mathcal{F} = \{E \subset X : E \text{ is linearly ordered}\}$ which we equip with the inclusion partial ordering. By 2. there exist a maximal element $E \in \mathcal{F}$. By assumption, the linearly ordered set E has an upper bound $x \in X$. The element x is maximal, for if $y \in Y$ and $y \geq x$, then $E \cup \{y\}$ is still an linearly ordered set containing E. So by maximality of E, $E = E \cup \{y\}$, i.e. $y \in E$ and therefore $y \leq x$ showing which combined with $y \geq x$ implies that $y = x$.[2] $(3 \Rightarrow 1)$ Let $\{X_\alpha\}_{\alpha\in A}$ be a collection of non-empty sets, we must show $\prod_{\alpha\in A} X_\alpha$ is not empty. Let $\mathcal{G}$ denote the collection of functions $g : D(g) \to \coprod_{\alpha\in A} X_\alpha$ such that $D(g)$ is a subset of A, and for all $\alpha \in D(g)$, $g(\alpha) \in X_\alpha$. Notice that $\mathcal{G}$ is not empty, for we may let $\alpha_0 \in A$ and $x_0 \in X_\alpha$ and then set $D(g) = \{\alpha_0\}$ and $g(\alpha_0) = x_0$ to construct an element of $\mathcal{G}$. We now put a partial order on $\mathcal{G}$ as follows. We say that $f \leq g$ for $f, g \in \mathcal{G}$ provided that $D(f) \subset D(g)$ and $f = g|_{D(f)}$. If $\Phi \subset \mathcal{G}$ is a linearly ordered set, let $D(h) = \cup_{g\in\Phi} D(g)$ and for $\alpha \in D(g)$ let $h(\alpha) = g(\alpha)$. Then $h \in \mathcal{G}$ is an upper bound for Φ. So by Zorn's Lemma there exists a maximal element $h \in \mathcal{G}$. To finish the proof we need only show that $D(h) = A$. If this were not the case, then let $\alpha_0 \in A \setminus D(h)$ and $x_0 \in X_{\alpha_0}$. We may now define $D(\tilde{h}) = D(h) \cup \{\alpha_0\}$ and

$$\tilde{h}(\alpha) = \begin{cases} h(\alpha) \text{ if } \alpha \in D(h) \\ x_0 \ \text{ if } \ \alpha = \alpha_0. \end{cases}$$

[1] If X is a countable set we may prove Zorn's Lemma by induction. Let $\{x_n\}_{n=1}^\infty$ be an enumeration of X, and define $E_n \subset X$ inductively as follows. For $n = 1$ let $E_1 = \{x_1\}$, and if E_n have been chosen, let $E_{n+1} = E_n \cup \{x_{n+1}\}$ if x_{n+1} is an upper bound for E_n otherwise let $E_{n+1} = E_n$. The set $E = \cup_{n=1}^\infty E_n$ is a linearly ordered (you check) subset of X and hence by assumption E has an upper bound, $x \in X$. I claim that his element is maximal, for if there exists $y = x_m \in X$ such that $y \geq x$, then x_m would be an upper bound for E_{m-1} and therefore $y = x_m \in E_m \subset E$. That is to say if $y \geq x$, then $y \in E$ and hence $y \leq x$, so $y = x$. (Hence we may view Zorn's lemma as a " jazzed" up version of induction.)

[2] Similarly one may show that $3 \Rightarrow 2$. Let $\mathcal{F} = \{E \subset X : E \text{ is linearly ordered}\}$ and order $\mathcal{F}$ by inclusion. If $\mathcal{M} \subset \mathcal{F}$ is linearly ordered, let $E = \cup\mathcal{M} = \bigcup_{A\in\mathcal{M}} A$. If $x, y \in E$ then $x \in A$ and $y \in B$ for some $A, B \subset \mathcal{M}$. Now $\mathcal{M}$ is linearly ordered by set inclusion so $A \subset B$ or $B \subset A$ i.e. $x, y \in A$ or $x, y \in B$. Since A and B are linearly order we must have either $x \leq y$ or $y \leq x$, that is to say E is linearly ordered. Hence by 3. there exists a maximal element $E \in \mathcal{F}$ which is the assertion in 2.

Then $h \leq \tilde{h}$ while $h \neq \tilde{h}$ violating the fact that h was a maximal element. $(1 \Rightarrow 2)$ Let $(X, \leq)$ be a partially ordered set. Let $\mathcal{F}$ be the collection of linearly ordered subsets of X which we order by set inclusion. Given $x_0 \in X$, $\{x_0\} \in \mathcal{F}$ is linearly ordered set so that $\mathcal{F} \neq \emptyset$. Fix an element $P_0 \in \mathcal{F}$. If P_0 is not maximal there exists $P_1 \in \mathcal{F}$ such that $P_0 \subsetneq P_1$. In particular we may choose $x \notin P_0$ such that $P_0 \cup \{x\} \in \mathcal{F}$. The idea now is to keep repeating this process of adding points $x \in X$ until we construct a maximal element P of $\mathcal{F}$. We now have to take care of some details. We may assume with out loss of generality that $\tilde{\mathcal{F}} = \{P \in \mathcal{F} : P \text{ is not maximal}\}$ is a non-empty set. For $P \in \tilde{\mathcal{F}}$, let $P^* = \{x \in X : P \cup \{x\} \in \mathcal{F}\}$. As the above argument shows, $P^* \neq \emptyset$ for all $P \in \tilde{\mathcal{F}}$. Using the axiom of choice, there exists $f \in \prod_{P \in \tilde{\mathcal{F}}} P^*$. We now define $g : \mathcal{F} \to \mathcal{F}$ by

$$g(P) = \begin{cases} P & \text{if } \quad P \text{ is maximal} \\ P \cup \{f(x)\} & \text{if } P \text{ is not maximal.} \end{cases} \tag{B.1}$$

The proof is completed by Lemma B.8 below which shows that g must have a fixed point $P \in \mathcal{F}$. This fixed point is maximal by construction of g. ■

Lemma B.8. *The function $g : \mathcal{F} \to \mathcal{F}$ defined in Eq. (B.1) has a fixed point.*[3]

Proof. The **idea of the proof** is as follows. Let $P_0 \in \mathcal{F}$ be chosen arbitrarily. Notice that $\Phi = \left\{g^{(n)}(P_0)\right\}_{n=0}^{\infty} \subset \mathcal{F}$ is a linearly ordered set and it is therefore easily verified that $P_1 = \bigcup_{n=0}^{\infty} g^{(n)}(P_0) \in \mathcal{F}$. Similarly we may repeat the process to construct $P_2 = \bigcup_{n=0}^{\infty} g^{(n)}(P_1) \in \mathcal{F}$ and $P_3 = \bigcup_{n=0}^{\infty} g^{(n)}(P_2) \in \mathcal{F}$, etc. etc. Then take $P_\infty = \cup_{n=0}^{\infty} P_n$ and start again with P_0 replaced by P_∞. Then keep going this way until eventually the sets stop increasing in size, in which case we have found our fixed point. The problem with this strategy is that we may never win. (This is very reminiscent of constructing measurable sets and the way out is to use measure theoretic like arguments.) Let us now start the **formal proof.** Again let $P_0 \in \mathcal{F}$ and let $\mathcal{F}_1 = \{P \in \mathcal{F} : P_0 \subset P\}$. Notice that $\mathcal{F}_1$ has the following properties:

1. $P_0 \in \mathcal{F}_1$.
2. If $\Phi \subset \mathcal{F}_1$ is a totally ordered (by set inclusion) subset then $\cup \Phi \in \mathcal{F}_1$.
3. If $P \in \mathcal{F}_1$ then $g(P) \in \mathcal{F}_1$.

Let us call a general subset $\mathcal{F}' \subset \mathcal{F}$ satisfying these three conditions a tower and let

$$\mathcal{F}_0 = \cap \{\mathcal{F}' : \mathcal{F}' \text{ is a tower}\}.$$

[3] Here is an easy proof if the elements of $\mathcal{F}$ happened to all be finite sets and there existed a set $P \in \mathcal{F}$ with a maximal number of elements. In this case the condition that $P \subset g(P)$ would imply that $P = g(P)$, otherwise $g(P)$ would have more elements than P.

Standard arguments show that $\mathcal{F}_0$ is still a tower and clearly is the smallest tower containing P_0. (Morally speaking $\mathcal{F}_0$ consists of all of the sets we were trying to constructed in the "idea section" of the proof.) We now claim that $\mathcal{F}_0$ is a linearly ordered subset of $\mathcal{F}$. To prove this let $\Gamma \subset \mathcal{F}_0$ be the linearly ordered set

$$\Gamma = \{C \in \mathcal{F}_0 : \text{ for all } A \in \mathcal{F}_0 \text{ either } A \subset C \text{ or } C \subset A\}.$$

Shortly we will show that $\Gamma \subset \mathcal{F}_0$ is a tower and hence that $\mathcal{F}_0 = \Gamma$. That is to say $\mathcal{F}_0$ is linearly ordered. Assuming this for the moment let us finish the proof. Let $P \equiv \cup\mathcal{F}_0$ which is in $\mathcal{F}_0$ by property 2 and is clearly the largest element in $\mathcal{F}_0$. By 3. it now follows that $P \subset g(P) \in \mathcal{F}_0$ and by maximality of P, we have $g(P) = P$, the desired fixed point. So to finish the proof, we must show that Γ is a tower. First off it is clear that $P_0 \in \Gamma$ so in particular Γ is not empty. For each $C \in \Gamma$ let

$$\Phi_C := \{A \in \mathcal{F}_0 : \text{either } A \subset C \text{ or } g(C) \subset A\}.$$

We will begin by showing that $\Phi_C \subset \mathcal{F}_0$ is a tower and therefore that $\Phi_C = \mathcal{F}_0$. 1. $P_0 \in \Phi_C$ since $P_0 \subset C$ for all $C \in \Gamma \subset \mathcal{F}_0$. 2. If $\Phi \subset \Phi_C \subset \mathcal{F}_0$ is totally ordered by set inclusion, then $A_\Phi := \cup\Phi \in \mathcal{F}_0$. We must show $A_\Phi \in \Phi_C$, that is that $A_\Phi \subset C$ or $C \subset A_\Phi$. Now if $A \subset C$ for all $A \in \Phi$, then $A_\Phi \subset C$ and hence $A_\Phi \in \Phi_C$. On the other hand if there is some $A \in \Phi$ such that $g(C) \subset A$ then clearly $g(C) \subset A_\Phi$ and again $A_\Phi \in \Phi_C$. 3. Given $A \in \Phi_C$ we must show $g(A) \in \Phi_C$, i.e. that

$$g(A) \subset C \text{ or } g(C) \subset g(A). \tag{B.2}$$

There are three cases to consider: either $A \subsetneq C$, $A = C$, or $g(C) \subset A$. In the case $A = C$, $g(C) = g(A) \subset g(A)$ and if $g(C) \subset A$ then $g(C) \subset A \subset g(A)$ and Eq. (B.2) holds in either of these cases. So assume that $A \subsetneq C$. Since $C \in \Gamma$, either $g(A) \subset C$ (in which case we are done) or $C \subset g(A)$. Hence we may assume that

$$A \subsetneq C \subset g(A).$$

Now if C were a proper subset of $g(A)$ it would then follow that $g(A)\setminus A$ would consist of at least two points which contradicts the definition of g. Hence we must have $g(A) = C \subset C$ and again Eq. (B.2) holds, so Φ_C is a tower. It is now easy to show Γ is a tower. It is again clear that $P_0 \in \Gamma$ and Property 2. may be checked for Γ in the same way as it was done for Φ_C above. For Property 3., if $C \in \Gamma$ we may use $\Phi_C = \mathcal{F}_0$ to conclude for all $A \in \mathcal{F}_0$, either $A \subset C \subset g(C)$ or $g(C) \subset A$, i.e. $g(C) \in \Gamma$. Thus Γ is a tower and we are done. ■

C

Nets

In this section (which may be skipped) we develop the notion of nets. Nets are generalization of sequences. Here is an example which shows that for general topological spaces, sequences are not always adequate.

Example C.1. Equip $\mathbb{C}^{\mathbb{R}}$ with the topology of pointwise convergence, i.e. the product topology and consider $C(\mathbb{R},\mathbb{C}) \subset \mathbb{C}^{\mathbb{R}}$. If $\{f_n\} \subset C(\mathbb{R},\mathbb{C})$ is a sequence which converges such that $f_n \to f \in \mathbb{C}^{\mathbb{R}}$ pointwise then f is a Borel measurable function. Hence the sequential limits of elements in $C(\mathbb{R},\mathbb{C})$ is necessarily contained in the Borel measurable functions which is properly contained in $\mathbb{C}^{\mathbb{R}}$. In short the sequential closure of $C(\mathbb{R},\mathbb{C})$ is a proper subset of $\mathbb{C}^{\mathbb{R}}$. On the other hand we have $\overline{C(\mathbb{R},\mathbb{C})} = \mathbb{C}^{\mathbb{R}}$. Indeed a typical open neighborhood of $f \in \mathbb{C}^{\mathbb{R}}$ is of the form

$$N = \{g \in \mathbb{C}^{\mathbb{R}} : |g(x) - f(x)| < \varepsilon \text{ for } x \in \Lambda\},$$

where $\varepsilon > 0$ and Λ is a finite subset of $\mathbb{R}$. Since $N \cap C(\mathbb{R},\mathbb{C}) \neq \emptyset$ it follows that $f \in \overline{C(\mathbb{R},\mathbb{C})}$.

Definition C.2. *A* ***directed set*** $(A, \leq)$ *is a set with a relation "$\leq$" such that*

1. $\alpha \leq \alpha$ *for all* $\alpha \in A$.
2. *If* $\alpha \leq \beta$ *and* $\beta \leq \gamma$ *then* $\alpha \leq \gamma$.
3. A *is* ***cofinite****, i.e.* $\alpha, \beta \in A$ *there exists* $\gamma \in A$ *such that* $\alpha \leq \gamma$ *and* $\beta \leq \gamma$.

A ***net*** *is function* $x : A \to X$ *where* A *is a directed set. We will often denote a net* x *by* $\{x_\alpha\}_{\alpha \in A}$.

Example C.3 (Directed sets).

1. $A = 2^X$ ordered by inclusion, i.e. $\alpha \leq \beta$ if $\alpha \subset \beta$. If $\alpha \leq \beta$ and $\beta \leq \gamma$ then $\alpha \subset \beta \subset \gamma$ and hence $\alpha \leq \gamma$. Similalry if $\alpha, \beta \in 2^X$ then $\alpha, \beta \leq \alpha \cup \beta =: \gamma$.
2. $A = 2^X$ ordered by reverse inclusion, i.e. $\alpha \leq \beta$ if $\beta \subset \alpha$. If $\alpha \leq \beta$ and $\beta \leq \gamma$ then $\alpha \supseteq \beta \supseteq \gamma$ and so $\alpha \leq \gamma$ and if $\alpha, \beta \in A$ then $\alpha, \beta \leq \alpha \cap \beta$.

3. Let $A = \mathbb{N}$ equipped with the usual ordering on $\mathbb{N}$. In this case nets are simply sequences.

Definition C.4. *Let $\{x_\alpha\}_{\alpha \in A} \subset X$ be a net then:*

1. *x_α **converges to** $x \in X$ (written $x_\alpha \to x$) iff for all $V \in \tau_x$, $x_\alpha \in V$ **eventually**, i.e. there exists $\beta = \beta_V \in A$ such that $x_\alpha \in V$ for all $\alpha \geq \beta$.*
2. *x is a **cluster point** of $\{x_\alpha\}_{\alpha \in A}$ if for all $V \in \tau_x$, $x_\alpha \in V$ **frequently**, i.e. for all $\beta \in A$ there exists $\alpha \geq \beta$ such that $x_\alpha \in V$.*

Proposition C.5. *Let X be a topological space and $E \subset X$. Then*

1. *x is an accumulation point of E (see Definition 10.29) iff there exists net $\{x_\alpha\} \subset E \setminus \{x\}$ such that $x_\alpha \to x$.*
2. *$x \in \bar{E}$ iff there exists $\{x_\alpha\} \subset E$ such that $x_\alpha \to x$.*

Proof.

1. Suppose x is an accumulation point of E and let $A = \tau_x$ be ordered by reverse set inclusion. To each $\alpha \in A = \tau_x$ choose $x_\alpha \in (\alpha \setminus \{x\}) \cap E$ which is possible sine x is an accumulation point of E. Then given $V \in \tau_x$ for all $\alpha \geq V$ (i.e. and $\alpha \subset V$), $x_\alpha \in V$ and hence $x_\alpha \to x$. Conversely if $\{x_\alpha\}_{\alpha \in A} \subset E \setminus \{x\}$ and $x_\alpha \to x$ then for all $V \in \tau_x$ there exists $\beta \in A$ such that $x_\alpha \in V$ for all $\alpha \geq \beta$. In particular $x_\alpha \in (E \setminus \{x\}) \cap V \neq \emptyset$ and so $x \in \mathrm{acc}(E)$ – the accumulation points of E.
2. If $\{x_\alpha\} \subset E$ such that $x_\alpha \to x$ then for all $V \in \tau_x$ there exists $\beta \in A$ such that $x_\alpha \in V \cap E$ for all $\alpha \geq \beta$. In particular $V \cap E \neq \emptyset$ for all $V \in \tau_x$ and this implies $x \in \bar{E}$. For the converse recall Proposition 10.31 implies $\overline{E} = E \cup \mathrm{acc}(E)$. If $x \in \mathrm{acc}(E)$ there exists a net $\{x_\alpha\} \subset E$ such that $x_\alpha \to x$ by item 1. If $x \in E$ we may simply take $x_n = x$ for all $n \in A := \mathbb{N}$.

■

Proposition C.6. *Let X and Y be two topological spaces and $f : X \to Y$ be a function. Then f is continuous at $x \in X$ iff $f(x_\alpha) \to f(x)$ for all nets $x_\alpha \to x$.*

Proof. If f is continuous at x and $x_\alpha \to x$ then for any $V \in \tau_{f(x)}$ there exists $W \in \tau_x$ such that $f(W) \subset V$. Since $x_\alpha \in W$ eventually, $f(x_\alpha) \in V$ eventually and we have shown $f(x_\alpha) \to f(x)$. Conversely, if f is **not** continuous at x then there exists $W \in \tau_{f(x)}$ such that $f(V) \nsubseteq W$ for all $V \in \tau_x$. Let $A = \tau_x$ be ordered by reverse set inclusion and for $V \in \tau_x$ choose (axiom of choice) $x_V \in V$ such that $f(x_V) \notin W$. Then $x_V \to x$ since for any $U \in \tau_x$, $x_V \in U$ if $V \geq U$ (i.e. $V \subset U$). On the over hand $f(x_V) \notin W$ for all $V \in \tau_x$ showing $f(x_V) \nrightarrow f(x)$. ■

Definition C.7 (Subnet). *A net $\langle y_\beta \rangle_{\beta \in B}$ is a **subnet** of a net $\langle x_\alpha \rangle_{\alpha \in A}$ if there exists a map $\beta \in B \to \alpha_\beta \in A$ such that*

1. $y_\beta = x_{\alpha_\beta}$ *for all* $\beta \in B$ *and*
2. *for all* $\alpha_0 \in A$ *there exists* $\beta_0 \in B$ *such that* $\alpha_\beta \geq \alpha_0$ *whenever* $\beta \geq \beta_0$, *i.e.* $\alpha_\beta \geq \alpha_0$ *eventually.*

Proposition C.8. *A point* $x \in X$ *is a* ***cluster*** *point of a net* $\langle x_\alpha \rangle_{\alpha \in A}$ *iff there exists a subnet* $\langle y_\beta \rangle_{\beta \in B}$ *such that* $y_\beta \to x$.

Proof. Suppose $\langle y_\beta \rangle_{\beta \in B}$ is a subnet of $\langle x_\alpha \rangle_{\alpha \in A}$ such that $y_\beta = x_{\alpha_\beta} \to x$. Then for $W \in \tau_x$ and $\alpha_0 \in A$ there exists $\beta_0 \in B$ such that $y_\beta = x_{\alpha_\beta} \in W$ for all $\beta \geq \beta_0$. Choose $\beta_1 \in B$ such that $\alpha_\beta \geq \alpha_0$ for all $\beta \geq \beta_1$ then choose $\beta_3 \in B$ such that $\beta_3 \geq \beta_1$ and $\beta_3 \geq \beta_2$ then $\alpha_\beta \geq \alpha_0$ and $x_{\alpha_\beta} \in W$ for all $\beta \geq \beta_3$ which implies $x_\alpha \in W$ frequently. Conversely assume x is a cluster point of a net $\langle x_\alpha \rangle_{\alpha \in A}$. We mak $B := \tau_x \times A$ into a directed set by defining $(U, \alpha) \leq (U', \alpha')$ iff $\alpha \leq \alpha'$ and $U \supseteq U'$. For all $(U, \gamma) \in B = \tau_x \times A$, choose $\alpha_{(U,\gamma)} \geq \gamma$ in A such that $y_{(U,\gamma)} = x_{\alpha_{(U,\gamma)}} \in U$. Then if $\alpha_0 \in A$ for all $(U', \gamma') \geq (U, \alpha_0)$, i.e. $\gamma' \geq \alpha_0$ and $U' \subset U$, $\alpha_{(U',\gamma')} \geq \gamma' \geq \alpha_0$. Now if $W \in \tau_x$ is given, then $y_{(U,\gamma)} \in U \subset W$ for all $U \subset W$. Hence fixing $\alpha \in A$ we see if $(U, \gamma) \geq (W, \alpha)$ then $y_{(U,\gamma)} = x_{\alpha_{(U,\gamma)}} \in U \subset W$ showing that $y_{(U,\gamma)} \to x$. ■

Exercise C.1 (Folland #34, p. 121). Let $\langle x_\alpha \rangle_{\alpha \in A}$ be a net in a topological space and for each $\alpha \in A$ let $E_\alpha \equiv \{x_\beta : \beta \geq \alpha\}$. Then x is a cluster point of $\langle x_\alpha \rangle$ iff $x \in \bigcap_{\alpha \in A} \overline{E}_\alpha$.

Solution to Exercise (C.1). If x is a cluster point, then given $W \in \tau_x$ we know $E_\alpha \cap W \neq \emptyset$ for all $\alpha \in E$ since $x_\beta \in W$ frequently thus $x \in \overline{E}_\alpha$ for all α, i.e. $x \in \bigcap_{\alpha \in A} \overline{E}_\alpha$. Conversely if x is **not** a cluster point of $\langle x_\alpha \rangle$ then there exists $W \in \tau_x$ and $\alpha \in A$ such that $x_\beta \notin W$ for all $\beta \geq \alpha$, i.e. $W \cap E_\alpha = \emptyset$. But this shows $x \notin \overline{E}_\alpha$ and hence $x \notin \bigcap_{\alpha \in A} \overline{E}_\alpha$.

Theorem C.9. *A topological space* X *is compact iff every net has a cluster point iff every net has a convergent subnet.*

Proof. Suppose X is compact, $\langle x_\alpha \rangle_{\alpha \in A} \subset X$ is a net and let $F_\alpha := \overline{\{x_\beta : \beta \geq \alpha\}}$. Then F_α is closed for all $\alpha \in A$, $F_\alpha \subset F_{\alpha'}$ if $\alpha \geq \alpha'$ and $F_{\alpha_1} \cap \cdots \cap F_{\alpha_n} \supseteq F_\gamma$ whenever $\gamma \geq \alpha_i$ for $i = 1, \ldots, n$. (Such a γ always exists since A is a directed set.) Therefore $F_{\alpha_1} \cap \cdots \cap F_{\alpha_n} \neq \emptyset$ i.e. $\{F_\alpha\}_{\alpha \in A}$ has the finite intersection property and since X is compact this implies there exists $x \in \bigcap_{\alpha \in a} F_\alpha$ By Exercise C.1, it follows that x is a cluster point of $\langle x_\alpha \rangle_{\alpha \in A}$. Conversely, if X is not compact let $\{U_j\}_{j \in J}$ be an infinite cover with no finite subcover. Let A be the directed set $A = \{\alpha \subset J : \#(\alpha) < \infty\}$ with $\alpha \leq \beta$ iff $\alpha \subset \beta$. Define a net $\langle x_\alpha \rangle_{\alpha \in A}$ in X by choosing

$$x_\alpha \in X \setminus \left(\bigcup_{j \in \alpha} U_j \right) \neq \emptyset \text{ for all } \alpha \in A.$$

This net has no cluster point. To see this suppose $x \in X$ and $j \in J$ is chosen so that $x \in U_j$. Then for all $\alpha \geq \{j\}$ (i.e. $j \in \alpha$), $x_\alpha \notin \bigcup_{\gamma \in \alpha} U_\alpha \supseteq U_j$ and in particular $x_\alpha \notin U_j$. This shows $x_\alpha \notin U_j$ frequently and hence x is not a cluster point. ■

D

Study Guides

D.1 Study Guide For Math 240A: Fall 2003

D.1.1 Basic things you should know about numbers and limits

1. I am taking for granted that you know the basic properties of $\mathbb{R}$ and $\mathbb{C}$ and that they are complete.
2. Should know how to compute $\lim a_n$, $\limsup a_n$ and $\liminf a_n$ and their basic properties. See Lemma 4.2 and Proposition 4.5 for example.

D.1.2 Basic things you should know about topological and measurable spaces:

1. You should know the basic definitions, Definition 10.1 and Definition 18.1.
2. It would be good to understand the notion of generating a topology or a σ – algebra by either a collection of sets or functions. This is key to understanding product topologies and product σ – algebras. See Propositions 10.7, 10.21 and 18.4 and Definition 18.24 and Proposition 18.25.
3. You should be able to check whether a given function is continuous or measurable. **Hints:**
   a) If possible avoid going back to the definition of continuity or measurability. Do this by using the stability properties of continuous (measurable) functions. For example continuous (measurable) functions are stable under compositions and algebraic operations, under uniform (pointwise) limits and sums. Measurable functions are also stable under taking sup, inf $\liminf$ and $\limsup$ of a sequence of measruable functions, see Proposition 18.36. Also recall if we are using the Borel σ – algebras, then continuous functions are automatically measurable.
   b) It is also possible to check continuity and measurability by splitting the space up and checking continuity and measurability on the individual pieces. See Proposition 10.19 and Exercise 10.7 and Proposition 18.29.

c) If you must go back to first principles, then the fact that $\sigma\left(f^{-1}(\mathcal{E})\right) = f^{-1}(\sigma(\mathcal{E}))$ and $\tau\left(f^{-1}(\mathcal{E})\right) = f^{-1}(\tau(\mathcal{E}))$ is key, see Lemma 18.22 and 10.14 respectively.

4. Dynkin's multiplicative system Theorems 18.51 and 18.52 are extremely useful for understanding the structure of measurable functions. They are also very useful for proving general theorems which are to hold for all bounded measurable functions. See the examples following Theorem 18.52 and the examples in Section 19.7.

D.1.3 Basic things you should know about Metric Spaces

1. The associated topology, see Example 10.3.
2. How to find the closure of a set. I typically would use the sequential definition of closure here.
3. Continuity is equivalent to the sequential notion of continuity, see Section 6.1.
4. The continuity properties of the metric, see Lemma 6.6.
5. The notions of Cauchy sequences and completeness.

D.1.4 Basic things you should know about Banach spaces

1. They are complete normed spaces.
2. $\ell^p(\mu)$ – spaces are Banach spaces, see Theorems 5.6, 5.8, and 7.5. Later we will see that all of these theorem hold for more general $L^p(\mu)$ – spaces as well.
3. $BC(X)$ is a closed subset of the Banach space $\ell^\infty(X)$ and hence is a Banach space, see Lemma 7.3.
4. The space of operators $L(X,Y)$ between two Banach spaces is a Banach space. In particular the dual space X^* is a Banach space, see Proposition 7.12.
5. How to find the norm of an operator and the basic properties of the operator norm, Lemma 7.10.
6. Boundedness of an operator is equivalent to continuity, Proposition 7.8.
7. Small perturbations of an invertible operator is still invertible, see Proposition 7.19 and Corollary 7.20.

D.1.5 The Riemann integral

The material on Riemann integral in Chapter 8 served as an illustration of much of the general Banach space theory described above. We also saw interesting applications to linear ODE.

However the **most important** result from Chapter 8 is the Weierstrass Approximation Theorem 8.34 and its complex version in Corollary 8.36.

D.1.6 Basic things you should know about Lebesgue integration theory and infinite sums

Recall that the Lebesgue integral relative to a counting type measure corresponds to an infinite sum, see Lemma 19.15. As a rule one does not need to go back to the definitions of integrals to work with them. The key points to working with integrals (and hence sums as well) are the following facts.

1. The integral is linear and satisfies the monotonicity properties: $\int f \le \int g$ if $f \le g$ a.e. and $\left|\int f\right| \le \int |f|$.
2. The monotone convergence Theorem 19.16 and its Corollary 19.18 about interchanging sums and integrals.
3. The dominated convergence Theorem 19.38 and its Corollary 19.39 about interchanging sums and integrals.
4. Fatou's Lemma 19.28 is used to a lesser extent.
5. Fubini and Tonelli theorems for computing multiple integrals. We have not done this yet for integrals, but the result for sums is in Theorems 4.22 and 4.23.
6. To compute integrals involving Lebesgue measure you will need to know the basic properties of Lebesgue measure, Theorem 19.10 and the fundamental theorem of calculus, Theorem 19.40.
7. You should understand when it is permissible to differentiate past the integral, see Corollary 19.43.

Remark D.1. Again let me stress that the above properties are typically all that are needed to work with integrals (sums). In particular to understand $\int_X f d\mu$ for a general measurable f it suffices to understand:

1. If $A \in \mathcal{M}$, then $\int_X 1_A d\mu = \mu(A)$. By linearity of the integral this determines $\int_X f d\mu$ on simple functions f.
2. Using either the monotone or dominated convergence theorem along with the approximation Theorem 18.42, $\int_X f d\mu$ may be written as a limit of integrals of simple functions.

D.2 Study Guide For Math 240B: Winter 2004

D.2.1 Basic things you should know about Multiple Integrals:

1. Product measures, Fubini and Tonelli theorems for computing multiple integrals, see Theorems 20.8 and 20.9. Keep in mind Driver's "rule;" if you see a multiple integral you should probably try to change the order of integration.
2. Lebesgue Measure on $\mathbb{R}^d$ and the change of variables formula, see Theorem 20.19. Also how to work in "abstract polar" coordinates, see Theorem 20.28.

D.2.2 Basic things you should know about L^p – spaces

1. L^p – spaces are Banach spaces, Theorems 21.19 and 21.20.
2. Key inequalities:
   a) Holder inequality, Theorem 21.2.
   b) Minkowski's Inequality, Theorem 21.4.
   c) Jensen's Inequality, Theorem 21.10.
   d) Chebyshev's inequality, Lemma 21.14.
   e) Minkowski's Inequality for Integrals, Theorem 21.27.

   You should be able to use these inequalities in basic situations.
3. Recall that the $L^p(\mu)$ – norm controls two types of behaviors of f, namely the "behavior at infinity" and the behavior of "local singularities." See the comments after Theorem 21.20.
4. You should have some feeling for the different modes of convergence, see Section 21.2.

D.2.3 Additional Basic things you should know about topological spaces:

1. The operations of closure, boundary and interior and in particular the interaction of closure with relative topologies. See Proposition 10.31 and Lemma 10.32.
2. The basic definitions of first countability, second countability, separability, density, etc., see Section 10.4.
3. The basic properties of connected sets, Theorems 10.48, 10.49, 10.50 and Proposition 10.53.
4. Compactness:
   a) The continuous image of compact sets are compact, Exercise 11.2.
   b) Dini's Theorem, Exercise 11.3.
   c) Equivalent characterizations of compactness in metric spaces, Theorem 11.7. Also see Corollary 11.9. You should be able to check compactness of a set in basic situations.

d) Extreme value theorem (Exercise 11.5), uniform continuity (Exercise 11.6).
e) The consequences for normed vector spaces, see Theorem 11.12, Corollary 11.13, Corollary 11.14 and Theorem 11.15.
f) Ascoli-Arzela Theorem 11.29 for checking function space compactness.
g) The definition of a compact operator, Definition 11.16.
h) The notions of locally and σ - compact spaces, Section 11.3.
i) Tychonoff's Theorem 11.34, i.e. the product of compact sets is still compact.

D.2.4 Things you should know about Locally Compact Hausdorff Spaces:

1. Know the definition.
2. They have lots of open sets and lots of continuous functions, see Propositions 12.5 and 12.7 and Urysohn's Lemma 12.8 for LCH Spaces and the Locally Compact Tietz Extension Theorem12.9.
3. Basic knowledge of partitions of unity, Section 12.2.
4. Alexanderov Compactification, Proposition 12.24. (Probably will not appear on any test.)
5. The Stone-Weierstrass Theorem, see Theorem 12.31 and Corollary 12.32.

D.2.5 Approximation Theorems and Convolutions

1. The density of $C_c(X)$ in $L^p(\mu)$ for all $p \in [1,\infty)$ when (X,τ) is a second countable locally compact Hausdorff space and $\mu : \mathcal{B}_X \to [0,\infty]$ be a K-finite measure, see Theorem 22.8. See the important special cases in Corollaries 22.9 and 22.10. Also see the closely related Lemma 22.11.
2. Density of smaller spaces of functions by using the results in item 1. with the Stone Weierstrass theorem, see Exercises 22.11 – 22.14.
3. The density of $\mathbb{S}_f(\mathcal{A},\mu)$ in $L^p(\mu)$ when μ is σ - finite on $\mathcal{A}$ and $\mathcal{M} = \sigma(\mathcal{A})$, see Theorem 22.14. Also see Theorem 22.15 on the separability of L^p - spaces and Example 22.16.
4. Convolution
   a) Know the Definition 22.20
   b) Know $\|f * g\|_p \le \|f\|_1 \|g\|_p$, Proposition 22.23.
   c) Understand the basic properties of convolution in Lemma 22.27.
   d) Understand Theorem 22.32 about approximate $\delta -$ functions.
   e) Know that $f * g$ is smooth if $g \in C_c^\infty(\mathbb{R}^d)$, see Proposition 22.34. Coupling this with Theorem 22.32 shows (for example): 1) continuous functions may be locally approximated by C^∞ - functions, 2) $C_c^\infty(\mathbb{R}^d)$ is dense in $L^p(\mathbb{R}^d,\mu)$ where $p \in [1,\infty)$ and μ is any $K -$ finite measure on $\mathcal{B}_{\mathbb{R}^d}$ (see Corollary 22.38 more generally), 3) there are C^∞ versions of Urysohn's Lemma (Corollary 22.35) and smooth versions of partitions of unity, see Section 22.2.1.

f) The integration by parts Lemma 22.36 is also often very useful.

D.2.6 Things you should know about Hilbert Spaces

1. The definition and the fact that $L^2(\mu)$ is an example.
2. The Schwarz Inequality Theorem 14.2 and the fact that the Hilbert norm is a norm, Corollary 14.3.
3. The notions of orthogonality, see Proposition 14.5.
4. The Best Approximation Theorem 14.10 and the Projection Theorem 14.13, see also Corollary 14.14.
5. The **very important** Riesz Theorem 14.15.
6. The notion of the adjoint of operators and their properties in Proposition 14.16 and Lemma 14.17.
7. The notions of orthonormal bases on Hilbert Spaces and their basic properties, see Section 14.1. Basically the results of this section, show you may manipulate with orthonormal bases on Hilbert spaces as you would in finite dimensional inner product spaces. Understand the examples in Example 23.2 and the important Fourier Series example in Theorem 23.9.
8. Many of the basic properties about Hilbert spaces can easily be deduced from your knowledge about $\ell^2(X)$ and the fact that every Hilbert space is unitarily equivalent (see Definition 14.29) to such a Hilbert space, see Exercise 14.7.
9. The notion of the spectrum of an operator, Definition 14.30.

References

1. Lynn H. Loomis, *An introduction to abstract harmonic analysis*, D. Van Nostrand Company, Inc., Toronto-New York-London, 1953. MR 14,883c

Index