Contents

Part I Numbers

1 Natural, integer, and rational Numbers
   1.1 Limits in Q ................................................................. 3
   1.2 The Problem with Q .................................................. 4
   1.3 Peano’s arithmetic (Highly Optional) ......................... 5

2 Fields .................................................................................. 11
   2.1 Basic Properties of Fields ........................................... 11
   2.2 Ordered Fields ............................................................... 13

3 Real Numbers ........................................................................ 17
   3.1 Extended real numbers ............................................... 20
   3.2 Limsups and Liminfs .................................................... 21
   3.3 Partitioning the Real Numbers ..................................... 25
   3.4 The Decimal Representation of a Real Number .......... 25
   3.5 Summary of Key Facts about Real Numbers ............... 26
   3.6 (Optional) Proofs of Theorem 3.6 and Theorem 3.3 .... 27
   3.7 Supremum and Infinums of sets ................................. 28

4 Complex Numbers ................................................................ 31
   4.1 A Matrix Perspective (Optional) ................................. 33

5 Set Operations, Functions, and Counting .......................... 35
   5.1 Set Operations and Functions .................................... 35
      5.1.1 Exercises ............................................................... 36
   5.2 Cardinality ................................................................. 36
   5.3 Finite Sets ................................................................. 37
   5.4 Countable and Uncountable Sets ............................... 38
# 10 Differential Calculus in One Real Variable

10.1 Basics of the Derivative .................................................. 107
10.2 First Derivative Test and the Mean Value Theorems .................. 109
10.3 Limsup and Liminf revisited .............................................. 110
10.4 Differentiating past infinite sums ...................................... 112
10.5 Taylor’s Theorem .......................................................... 116
10.6 Inverse Function Theorem I .............................................. 118
10.7 L’hopital’s Rule ............................................................ 121

# 11 Simple Integration Theory

11.1 Banach Spaces of Bounded Functions ................................. 123
11.2 Algebras of Sets ............................................................ 124
11.3 Finitely additive measure on $\mathcal{A}$ .................................. 126
11.4 Simple Functions .......................................................... 127
11.5 Simple Integration ....................................................... 128
11.6 Product Measure .......................................................... 131

# 12 Extending the Integral by Uniform Limits

12.1 The Fundamental Theorem of Calculus ............................... 133
12.2 Taylor’s Theorem revisited .............................................. 138
12.3 Interchanging Limits ........................................................ 142
12.4 Stirling’s Formula .......................................................... 143
12.4.1 A primitive Stirling type approximation ......................... 147
12.4.2 Bone Yard ............................................................. 148

# Part IV A Little Functional Analysis

# 13 Weierstrass Approximation Theorem

13.1 Classical Weierstrass Approximation Theorem ..................... 153
13.2 Stone-Weierstrass Theorem for Compact $X$ ....................... 155
13.3 Stone-Weierstrass Theorem for Locally Compact $X$ ............... 157

# 14 Inner Product Spaces and Orthonormal Bases ......................... 161

# 15 Fourier Series

15.1 Fourier Basis ............................................................. 167
15.2 Fourier Series Exercises .................................................. 167
15.3 Dirichlet, Fejér and Kernels .............................................. 170
15.4 The Dirichlet Problems on $D$ and the Poisson Kernel ............. 171

# 16 Arzela-Ascoli Theorem (Compactness in $C(X)$)

16.0.1 Arzela-Ascoli Theorem Problems ................................... 177
## Part V Appendices

### A Appendix: Notation and Logic

- Overview
- Exercises

### B Appendix: More Set Theoretic Properties (highly optional)

- Appendix: Zorn’s Lemma and the Hausdorff Maximal Principle (optional)
- Cardinality
- Formalities of Counting

### C Appendix: Aspects of General Topological Spaces

- Compactness
- Exercises
  - General Topological Space Problems
  - Metric Spaces as Topological Spaces
  - Compactness Problems
- Connectedness in General Topological Spaces
- Constructions of Metric Spaces
- Supplementary Remarks
  - Riemannian Metrics

### D Appendix: Lim sup and lim inf of functions

### E Appendix: Countably Additive Measures

- Overview
- A Density Result
- Construction of Measures
- Radon Measures on \( \mathbb{R} \)
  - Lebesgue Measure

### F Appendix: Math 140A Topics

- Summary of Key Facts about Real Numbers
- Test 2 Review Topics
- After Test 2 Review Topics
Part I

Numbers
Natural, integer, and rational Numbers

Notation 1.1 Let \( \mathbb{N} = \{1, 2, \ldots \} \) denote the natural numbers, \( \mathbb{N}_0 = \{0\} \cup \mathbb{N} \), 
\[ \mathbb{Z} := \{0, \pm1, \pm2, \ldots \} = \{\pm n : n \in \mathbb{N}_0\} \]
be the integers, and 
\[ \mathbb{Q} := \left\{ \frac{m}{n} : m \in \mathbb{Z} \text{ and } n \in \mathbb{N} \right\} \]
be the rationale numbers.

I am going to assume that the reader is familiar with all the standard arithmetic operations (addition, multiplication, inverses, etc.) on \( \mathbb{N}_0, \mathbb{Z}, \) and \( \mathbb{Q}. \) However let us review the important induction axiom of the natural numbers.

**Induction Axiom** If \( S \subset \mathbb{N} \) is a subset such that \( 1 \in S \) and \( n+1 \in S \) whenever \( n \in S \), then \( S = \mathbb{N}. \)

This axiom takes on two other useful forms which we describe in the next Propositions.

**Proposition 1.2 (Strong form of Induction).** Suppose \( S \subset \mathbb{N} \) is a subset such that \( 1 \in S \) and \( n+1 \in S \) whenever \( \{1, 2, \ldots, n\} \subset S \), then \( S = \mathbb{N}. \)

**Proof.** Let \( T := \{n \in \mathbb{N} : \{1, 2, \ldots, n\} \subset S \}. \) Then \( 1 \in T \) and if \( n \in T \) then \( n+1 \in T \) by assumption. Therefore by the induction axiom, \( T = \mathbb{N} \) so that \( \{1, 2, \ldots, n\} \subset S \) in for all \( n \in \mathbb{N}. \) This suffices to show \( S = \mathbb{N}. \)

**Proposition 1.3 (Well ordering principle).** Suppose \( S \subset \mathbb{N} \) is a non-empty subset, then there exists a smallest element \( m \) of \( S. \)

**Proof.** Let \( S \) be a subset of \( \mathbb{N} \) for which there is no smallest element, \( m \in S. \) Let 
\[ T = \{n \in \mathbb{N} : n < s \text{ for all } s \in S \}. \]
If \( 1 \notin T \), then \( 1 \in S \) and \( 1 \) would be a smallest element of \( S. \) Hence we must have \( 1 \in T. \) Now suppose that \( n \in T \) so that \( n < s \) for all \( s \in S. \) If \( n+1 \notin T \) then there exists \( s \in S \) such that \( n < s \leq n+1 \) which would force \( s = n+1 \in S. \) But we would then have \( n+1 \) is the minimal element of \( S \) which is assumed not to exist. So we have shown if \( n \in T \) then \( n+1 \in T. \) So by the induction axiom of \( \mathbb{N} \) it follows that \( T = \mathbb{N} \) and therefore \( n \notin S \) for all \( n \in \mathbb{N}, \) i.e. \( S = \emptyset. \)

**Remark 1.4.** Let us further observe that the well ordering principle implies the induction axiom. Indeed, suppose that \( S \subset \mathbb{N} \) is a subset such that \( 1 \in S \) and \( n+1 \in S \) whenever \( n \in S. \) For sake of contradiction suppose that \( S \neq \mathbb{N} \) so that \( T := \mathbb{N} \setminus S \) is not empty. By the well ordering principle there \( T \) has a unique minimal element \( m \) and in particular \( T \subset \{m, m+1, \ldots \}. \) This implies that \( \{1, 2, \ldots, m-1\} \subset S \) and then by assumption that \( \{1, 2, \ldots, m\} \subset S. \) But this then implies \( T \subset \{m+1, \ldots \} \) and therefore \( m \notin T \) which violates \( m \) being the minimal element of \( T. \) We have arrived at the desired contradiction and therefore conclude that \( S = \mathbb{N}. \)

**Remark 1.5.** Recall that, for \( q \in \mathbb{Q}, \) we define
\[ |q| = \begin{cases} q & \text{if } q \geq 0 \\ -q & \text{if } q \leq 0. \end{cases} \]
Recall that, for all \( a, b \in \mathbb{Q}, \)
\[ |a + b| \leq |a| + |b|, \quad |ab| = |a||b|, \quad \text{and } \left| \frac{1}{a} \right| = \frac{1}{|a|} \text{ when } a \neq 0. \]
It is also often useful to keep in mind that the following statements are equivalent for \( a, b \in \mathbb{Q} \) with \( b \geq 0; \)
1. \( |a| \leq b, \)
2. \( -b \leq a \leq b, \) and
3. \( a \leq b, \) i.e. \( a \leq b \) and \( -a \leq b. \)

**Lemma 1.6.** If \( a, b \in \mathbb{Q}, \) then 
\[ ||b| - |a|| \leq |b - a|. \] (1.1)

**Proof.** Since both sides of Eq. (1.1) are symmetric in \( a \) and \( b, \) we may assume that \( |b| \geq |a| \) so that \( ||b| - |a|| = |b| - |a|. \) Since 
\[ |b| = |b - a + a| \leq |b - a| + |a|, \] it follows that 
\[ ||b| - |a|| = |b - a| \leq |b - a|. \]
1.1 Limits in \( \mathbb{Q} \)

In this course we will often use the abbreviations, i.o. and a.a. which stand for \textbf{infinitely often} and \textbf{almost always} respectively. For example \( a_n \leq b_n \) a.a. \( n \) means there exists an \( N \in \mathbb{N} \) such that \( a_n \leq b_n \) for all \( n \geq N \) while \( a_n \leq b_n \) i.o. \( n \) means for all \( N \in \mathbb{N} \) there exists a \( n \geq N \) such that \( a_n \leq b_n \). So for example, \( 1/n \leq 1/100 \) for a.a. \( n \) while and \((-1)^n \geq 0 \) i.o. By the way, it should be clear that if something happens for a.a. \( n \) then it also happens i.o. \( n \).

**Definition 1.9.** A sequence \( \{a_n\}_{n=1}^{\infty} \subset \mathbb{Q} \) \textbf{converges to} \( 0 \in \mathbb{Q} \) if for all \( \varepsilon > 0 \) in \( \mathbb{Q} \) there exists \( N \in \mathbb{N} \) such that \( |a_n| \leq \varepsilon \) for all \( n \geq N \). Alternatively put, for all \( \varepsilon > 0 \) we have \( |a_n| \leq \varepsilon \) for a.a. \( n \). This may also be stated as for all \( M \in \mathbb{N} \), \( |a_n| \leq \frac{\varepsilon}{M} \) for a.a. \( n \).

**Definition 1.10.** A sequence \( \{a_n\}_{n=1}^{\infty} \subset \mathbb{Q} \) \textbf{converges to} \( a \in \mathbb{Q} \) if \( |a - a_n| \to 0 \) as \( n \to \infty \), i.e. if for all \( N \in \mathbb{N} \), \( |a - a_n| \leq \frac{1}{N} \) for a.a. \( n \). As usual if \( \{a_n\}_{n=1}^{\infty} \) converges to \( a \) we will write \( a_n \rightarrow a \) as \( n \to \infty \) or \( a = \lim_{n \to \infty} a_n \).

**Proposition 1.11.** If \( \{a_n\}_{n=1}^{\infty} \subset \mathbb{Q} \) converges to \( a \in \mathbb{Q} \), then \( \lim_{n \to \infty} |a_n| = |a| \).

**Proof.** From Lemma 1.6 we have,

\[
||a| - |a_n|| \leq |a - a_n|.
\]

Thus if \( \varepsilon > 0 \) is given, by definition of \( a_n \rightarrow a \) there exists \( N \in \mathbb{N} \) such that \( |a - a_n| < \varepsilon \) for all \( n \geq N \). From the previously displayed equation, it follows that \( ||a| - |a_n|| < \varepsilon \) for all \( n \geq N \) and hence we may conclude that \( \lim_{n \to \infty} |a_n| \) exists and is equal to \( |a| \).

**Lemma 1.12 (Convergent sequences are bounded).** If \( \{a_n\}_{n=1}^{\infty} \subset \mathbb{Q} \) converges to \( a \in \mathbb{Q} \), then there exists \( M \in \mathbb{Q} \) such that \( |a_n| \leq M \) for all \( n \in \mathbb{N} \).

**Proof.** Taking \( \varepsilon = 1 \) in the definition of \( a = \lim_{n \to \infty} a_n \) implies there exists \( N \in \mathbb{N} \) such that \( |a_n - a| \leq 1 \) for all \( n \geq N \). Therefore,

\[
|a_n| = |a_n - a + a| \leq |a_n - a| + |a| \leq 1 + |a| \ 	ext{for} \ n \geq N.
\]

We may now take \( M := \max\left(\left\{ |a_n|_{n=1}^{N}\right\} \cup \{1 + |a|\} \right) \).

**Theorem 1.13.** If \( \{a_n\}_{n=1}^{\infty} \subset \mathbb{Q} \) converges to \( a \in \mathbb{Q} \setminus \{0\} \), then

\[
\frac{1}{|a_n|} = \frac{1}{a}.
\]  

(1.2)

It is possible that \( a_n = 0 \) for small \( n \) so that \( \frac{1}{a_n} \) is not defined but for large \( n \) this can not happen and therefore it makes sense to talk about the limit which only depends on the tail of the sequences.
Proof. Since \( a \neq 0 \) we know that \(|a| > 0\). Hence, there exists \( M := M_{|a|} \in \mathbb{N} \) such that \(|a_n - a| < \frac{|a|}{2} \) for all \( n \geq M \). Therefore for \( n \geq M \)

\[
|a| = |a - a_n + a_n| \leq |a - a_n| + |a_n| < \frac{|a|}{2} + |a_n|
\]

from which it follows\(^3\) that \(|a_n| > \frac{|a|}{2}\) for all \( n \geq M \). If \( \varepsilon > 0 \) is given arbitrarily, we may choose \( N \geq M \) such that \(|a_n - a| < \varepsilon\) for all \( n \geq M \). Then for \( n \geq N \) we have,

\[
\frac{1}{a_n} - 1 = \frac{a - a_n}{a_n a} = \frac{|a - a_n|}{|a_n| a} < \varepsilon = \frac{|a|}{2} \cdot \frac{1}{|a|^2} = 2\varepsilon.
\]

As \( \varepsilon > 0 \) is arbitrary it follows that \( 2\varepsilon \frac{|a|}{|a|^2} = 2\varepsilon \) is arbitrarily small as well (replace \( \varepsilon \) by \( \varepsilon |a|^2 / 2 \) if you feel it is necessary), and hence we may conclude that Eq. 1.2 holds.

Variation on the method. In order to make these arguments more routine, it is often a good idea to write \( a_n = a + \delta_n \), where \( \delta_n := a_n - a \) is the error between \( a_n \) and \( a \). By assumption, \( \lim_{n \to \infty} \delta_n = 0 \) and so for any \( \delta > 0 \) given there exists \( N(\delta) \in \mathbb{N} \) such that \( |\delta_n| \leq \delta \). With this notation we have,

\[
\frac{1}{a_n} - 1 = \frac{1}{a + \delta_n} - 1 = \frac{-\delta_n}{a + \delta_n} \cdot \frac{1}{a} \leq \frac{\delta}{|a||a| - |a|}.
\]

So if we assume that \( \delta \leq |a|/2 \) we find that

\[
\frac{1}{a_n} - 1 \leq 2\varepsilon/|a|^2 \text{ for all } n \geq N(\delta).
\]

Taking \( \delta = \delta(\varepsilon) = \min\left(|a|/2, |a|^2 \varepsilon/2\right) \) in Eq. 1.3 shows for \( n \geq N(\delta(\varepsilon)) \) that \( \left|\frac{1}{a_n} - \frac{1}{a}\right| \leq \varepsilon \). Since \( \varepsilon > 0 \) is arbitrary we may conclude that \( \lim_{n \to \infty} \frac{1}{a_n} = \frac{1}{a} \).

End of Lecture 1, 9/28/2012

Definition 1.14. A sequence \( \{a_n\}_{n=1}^\infty \subset \mathbb{Q} \) is Cauchy if \( |a_n - a_m| \to 0 \) as \( m, n \to \infty \). More precisely we require for each \( \varepsilon > 0 \) in \( \mathbb{Q} \) that \( |a_m - a_n| \leq \varepsilon \) for a.e. pairs \( (m, n) \), i.e. there should exists \( N \in \mathbb{N} \) such that \( |a_m - a_n| \leq \varepsilon \) for all \( m, n \geq N \).

The idea is very simple here. If \( a_n \) is near \( a \) and \( a \neq 0 \) then \( a_n \) must stay away from zero. You should draw the picture to go along with the proof.

Exercise 1.1. Show that all convergent sequences \( \{a_n\}_{n=1}^\infty \subset \mathbb{Q} \) are Cauchy.

Exercise 1.2. Show all Cauchy sequences \( \{a_n\}_{n=1}^\infty \) are bounded – i.e. there exists \( M \in \mathbb{N} \) such that

\[
|a_n| \leq M \text{ for all } n \in \mathbb{N}.
\]

Exercise 1.3. Suppose \( \{a_n\}_{n=1}^\infty \) and \( \{b_n\}_{n=1}^\infty \) are Cauchy sequences in \( \mathbb{Q} \). Show \( \{a_n + b_n\}_{n=1}^\infty \) and \( \{a_n \cdot b_n\}_{n=1}^\infty \) are Cauchy.

Exercise 1.4. Assume that \( \{a_n\}_{n=1}^\infty \) and \( \{b_n\}_{n=1}^\infty \) are convergent sequences in \( \mathbb{Q} \). Show \( \{a_n + b_n\}_{n=1}^\infty \) and \( \{a_n \cdot b_n\}_{n=1}^\infty \) are convergent in \( \mathbb{Q} \) and

\[
\lim_{n \to \infty} (a_n + b_n) = \lim_{n \to \infty} a_n + \lim_{n \to \infty} b_n \quad \text{and} \quad \lim_{n \to \infty} (a_n b_n) = \lim_{n \to \infty} a_n \cdot \lim_{n \to \infty} b_n.
\]

Exercise 1.5. Assume that \( \{a_n\}_{n=1}^\infty \) and \( \{b_n\}_{n=1}^\infty \) are convergent sequences in \( \mathbb{Q} \) such that \( a_n \leq b_n \) for all \( n \). Show \( A := \lim_{n \to \infty} a_n \leq \lim_{n \to \infty} b_n =: B \).

Exercise 1.6 (Sandwich Theorem). Assume that \( \{a_n\}_{n=1}^\infty \) and \( \{b_n\}_{n=1}^\infty \) are convergent sequences in \( \mathbb{Q} \) such that \( \lim_{n \to \infty} a_n = \lim_{n \to \infty} b_n \). If \( \{x_n\}_{n=1}^\infty \) is another sequence in \( \mathbb{Q} \) that satisfies \( a_n \leq x_n \leq b_n \) for all \( n \), then

\[
\lim_{n \to \infty} x_n = a := \lim_{n \to \infty} a_n = \lim_{n \to \infty} b_n.
\]

Please note that that main part of the problem is to show that \( \lim_{n \to \infty} x_n \) exists in \( \mathbb{Q} \). Hint: start by showing; if \( a \leq x \leq b \) then \( |x| \leq \max(|a|, |b|) \).

Definition 1.15 (Subsequence). We say a sequence, \( \{y_k\}_{k=1}^\infty \subset \mathbb{Q} \) is a subsequence of another sequence, \( \{x_n\}_{n=1}^\infty \), provided there exists a strictly increasing function, \( \mathbb{N} \ni k \to n_k \in \mathbb{N} \) such that \( y_k = x_{n_k} \) for all \( k \in \mathbb{N} \). [Example, \( n_k = k^2 + 3 \), and \( \{y_k := x_{k^2+3}\}_{k=1}^\infty \) would be a subsequence of \( \{x_n\}_{n=1}^\infty \).]

Exercise 1.7. Suppose that \( \{x_n\}_{n=1}^\infty \) is a Cauchy sequence in \( \mathbb{Q} \) (or \( \mathbb{R} \)) which has a convergent subsequence, \( \{y_k = x_{n_k}\}_{k=1}^\infty \subset \mathbb{Q} \) (or \( \mathbb{R} \)). Show that \( \lim_{n \to \infty} x_n \) exists and is equal to \( \lim_{k \to \infty} y_k \).

1.2 The Problem with \( \mathbb{Q} \)

The problem with \( \mathbb{Q} \) is that it is full of “holes.” To be more precise, \( \mathbb{Q} \) is not “complete,” i.e. not all Cauchy sequences are convergent. In fact, according to Corollary 5.51\(^4\) below, “most” Cauchy sequences of rational numbers do not converge to a rational number. Let us demonstrate some examples pointing out this flaw. We first pause to recall how to sum geometric series.

---

\(^3\) The idea is very simple here. If \( a_n \) is near \( a \) and \( a \neq 0 \) then \( a_n \) must stay away from zero. You should draw the picture to go along with the proof.

\(^4\) Theorem 5.51: Every Cauchy sequence of rational numbers converges to a limit in \( \mathbb{Q} \).

---
Lemma 1.16 (Geometric Series). Let $\alpha \in \mathbb{Q}, m, n \in \mathbb{Z}$ with $n \leq m$, and $S := \sum_{k=n}^{m} \alpha^k$. Then

$$S = \begin{cases} m - n + 1 & \text{if } \alpha = 1 \\
\frac{\alpha^{m+1} - \alpha^n}{\alpha - 1} & \text{if } \alpha \neq 1.
\end{cases}$$

Moreover if $0 \leq \alpha < 1$, then

$$\sum_{k=n}^{m} \alpha^k = \alpha^n \frac{1 - \alpha^{m-n+1}}{1 - \alpha} \leq \alpha^n \frac{1}{1 - \alpha}. \quad (1.4)$$

Proof. When $\alpha = 1$,

$$S = \sum_{k=n}^{m} 1^k = m - n + 1.$$

If $\alpha \neq 1$, then

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

Solving for $S$ gives

$$S = \sum_{k=n}^{m} \alpha^k = \frac{\alpha^{m+1} - \alpha^n}{\alpha - 1} \text{ if } \alpha \neq 1. \quad (1.5)$$

Example 1.17. Let $S_n := \sum_{k=0}^{n} \frac{1}{n!} \in \mathbb{Q}$ for all $n \in \mathbb{N}$. For $n > m$ in $\mathbb{N}$ we have,

$$0 \leq S_n - S_m = \sum_{k=m+1}^{n} \frac{1}{k!} = \sum_{j=1}^{n-m} \frac{1}{(m+j)!}$$

$$= \frac{1}{(m+1)!} + \cdots + \frac{1}{m!}$$

$$\leq \frac{1}{m!} \left[ \frac{1}{m+1} + \left( \frac{1}{m+1} \right)^2 + \cdots + \left( \frac{1}{m+1} \right)^{n-m} \right]$$

$$\leq \frac{1}{m!} \frac{1}{m+1} \frac{1}{1 - \left( \frac{1}{m+1} \right)} = \frac{1}{m \cdot m!}. \quad (1.6)$$

wherein we have used Eq. (1.4) for the last inequality. From this inequality it follows that $\{S_n\}_{n=0}^{\infty}$ is a Cauchy sequence and we also have,

$$\frac{1}{(m+1)!} \leq S_n - S_m \leq \frac{1}{m \cdot m!} \text{ for all } n > m. \quad (1.7)$$

Suppose that $e := \lim_{n \to \infty} S_n$ were to exist in $\mathbb{Q}$. Then letting $n \to \infty$ in Eq. (1.7) would show,

$$0 < \frac{1}{(m+1)!} \leq e - S_m \leq \frac{1}{m \cdot m!}.$$

Multiplying this inequality by $m!$ then implies,

$$0 < m! e - m! S_m \leq \frac{1}{m}.$$

However for $m$ sufficiently large $m! e \in \mathbb{N}$ (as $e$ is assumed to be rational) and $m! S_m$ is always in $\mathbb{N}$ and therefore $k := m! e - m! S_m \in \mathbb{N}$. But there is no element $k \in \mathbb{N}$ such that $0 < k < \frac{1}{m} \text{ and hence we must conclude } \lim_{n \to \infty} S_n \text{ can not exist in } \mathbb{Q}$.

Moral: the number $e = \sum_{n=0}^{\infty} \frac{1}{n!} = \lim_{n \to \infty} (1 + \frac{1}{n})^n$ that you learned about in calculus is not in $\mathbb{Q}$!

Plotting the partial sums $\sum_{k=0}^{n} \frac{1}{n!}$ (black curve) and $(1 + \frac{1}{n})^n$ (red curve) which are both converging to “$e$.”

Example 1.18 (Square roots need not exist). The square root, $\sqrt{2}$, of $2$ does not exist in $\mathbb{Q}$. Indeed, if $\sqrt{2} = \frac{m}{n}$ where $m$ and $n$ have no common factors (in particular no common factors of 2 so that either $m$ or $n$ is odd), then

$$\frac{m^2}{n^2} = 2 \implies m^2 = 2n^2.$$

This shows that $m^2$ is even which would then imply that $m = 2k$ is even (since odd-odd=odd). However this implies $4k^2 = 2n^2$ from which it follows that $n^2 = 2k^2$ is even and hence $n$ is even. But this contradicts the assumption that $m$ and $n$ had no common factors (of 2).
Exercise 1.8. Use the following outline to construct another 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 \( q^2 = 2 \). To each \( n \in \mathbb{N} \) let \( m_n \in \mathbb{N} \) be chosen so that
   \[
   \frac{m_n^2}{n^2} < 2 < \left( \frac{m_n + 1}{n} \right)^2
   \tag{1.8}
   \]
   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 \( \mathbb{Q} \).

Example 1.19. It is also a fact that \( \pi \notin \mathbb{Q} \) where
   \[
   \pi = 2 \int_0^\infty \frac{1}{1 + x^2} \, dx = 2 \lim_{N \to \infty} \int_0^N \frac{1}{1 + x^2} \, dx
   = 2 \lim_{N \to \infty} \sum_{k=1}^{N^2} \frac{1}{1 + \left( \frac{k}{N} \right)^2} \cdot \frac{1}{N}
   = \lim_{N \to \infty} \sum_{k=1}^{N^2} \frac{2N^2}{N^2 + k^2}.
   
   The point is that the basic operations from calculus tend to produce “real numbers” which are not rational even though we start with only rational numbers.

End of Lecture 2, 10/1/2012

1.3 Peano’s arithmetic (Highly Optional)

This section is for those who want to understand \( \mathbb{N} \) at a more fundamental level. Here we start with Peano’s rather minimalist axioms for \( \mathbb{N} \) and show how they lead to all the standard properties you are used to using for \( \mathbb{N} \). Here are the axioms:

- non-empty \( \mathbb{N}_0 \) is a non-empty set which contains a distinguished element, 0.
- We let \( \mathbb{N} := \mathbb{N}_0 \setminus \{0\} \) and call these the natural numbers.

Successor Function There is an injective function, \( s : \mathbb{N}_0 \to \mathbb{N} \) and we let
   \[ 1 := s(0) \in \mathbb{N} . \]

Induction hypothesis If \( S \subset \mathbb{N}_0 \) is a set such that \( 0 \in S \) and \( s(n) \in S \) whenever \( n \in S \), then \( S = \mathbb{N}_0 \).

Assuming these axioms one may develop all of the properties or \( \mathbb{N}_0 \) that you are accustomed to seeing. I will develop the basic properties of addition, multiplication, and the ordering on \( \mathbb{N}_0 \) in this section. For more on this point and then the further construction of \( \mathbb{Z} \) and \( \mathbb{Q} \) from \( \mathbb{N}_0 \), the reader is referred to the notes; \textit{“Numbers” by M. Taylor}. You may also consult E. Landau’s book \([\textcopyright] \) for a very detailed (but perhaps too long winded) exposition of these topics.

Lemma 1.20. The map \( s : \mathbb{N}_0 \to \mathbb{N} \) is a bijection.

Proof. Let \( S := s(\mathbb{N}_0) \cup \{0\} \subset \mathbb{N}_0 \). Then \( 0 \in S \) and \( s(0) \in s(\mathbb{N}_0) \subset S \). Moreover if \( x \in \mathbb{N} \cap S \) then \( s(x) \in s(\mathbb{N}_0) \subset S \) so that \( x \in S \implies s(x) \in S \) and hence \( S = \mathbb{N}_0 \) and therefore \( s(\mathbb{N}_0) = \mathbb{N} \).

Theorem 1.21 (Addition). There exists a function \( p : \mathbb{N}_0 \times \mathbb{N}_0 \to \mathbb{N}_0 \) such that \( p(x,0) = x \) for all \( x \in \mathbb{N}_0 \) and \( p(x,s(y)) = s(p(x,y)) \) for all \( x, y \in \mathbb{N}_0 \). Moreover, we may construct \( p \) so that \( p(s(x),y) = p(x,s(y)) \) for all \( x, y \in \mathbb{N}_0 \).

This function \( p \) satisfies the following properties:
1. \( p(x,0) = x = p(0,x) \) for all \( x \in \mathbb{N}_0 \),
2. \( p(x,1) = p(1,x) = s(x) \) for all \( x \in \mathbb{N}_0 \),
3. \( p(x,y) = p(y,x) \) for all \( x, y \in \mathbb{N}_0 \),
4. \( p(x,p(y,z)) = p(x,y,z) \) for all \( x, y, z \in \mathbb{N}_0 \).

Proof. We will construct \( p \) inductively. Let
   \[ S := \{ x \in \mathbb{N} : \exists p_x : \mathbb{N}_0 \to \mathbb{N}_0 \ni p_x(0) = x \text{ and } p_x(s(y)) = s(p_x(y)) \forall y \in \mathbb{N}_0 \} . \]
Taking \( p_0(y) = y \) shows \( 0 \in S \). Moreover if \( x \in S \) we define
   \[ p_{s(x)}(y) := s(p_x(y)) \text{ for all } y \in \mathbb{N}_0 . \]
We then have \( p_{s(x)}(0) = s(p_x(0)) = s(x) \) and
   \[ p_{s(x)}(s(y)) := s(p_x(s(y))) = s(p_x(y)) = s(p_{s(x)}(y)) \]
which shows \( s(x) \in S \). Thus we may conclude \( S = \mathbb{N}_0 \) and we may now define \( p(x,y) := p_x(y) \) for all \( x, y \in \mathbb{N}_0 \). By construction this function satisfies,
   \[ p(s(x),y) = s(p(x,y)) = p(x,s(y)) . \]

We now verify the properties in items 1. – 4.
1. By construction \( p(x,0) = x \) for all \( x \in \mathbb{N}_0 \). Let \( S = \{ x \in \mathbb{N} : p(0,x) = x \} \), then \( 0 \in S \) and if \( x \in S \) we have \( p(0,s(x)) = s(p(0,x)) = s(x) \) so that \( s(x) \in S \). Therefore \( S = \mathbb{N}_0 \) and the first item holds.

2. \( p(x,1) = p(x,s(0)) = s(p(x,0)) = s(x) \) and \( p(1,x) = p(s(0),x) = s(p(0,x)) = s(x) \) so that item 2. is proved.

3. Let \( S = \{ x \in \mathbb{N}_0 : p(x,\cdot) = p(\cdot,x) \} \). Then by items 1 and 2. it follows that \( 0,1 \in S \). Moreover if \( x \in S \), then for all \( y \in \mathbb{N}_0 \) we find,

\[
p(s(x),y) = s(p(x,y)) = s(p(y,x)) = p(y,s(x))
\]

which shows \( s(x) \in S \). Therefore \( S = \mathbb{N}_0 \) and item 3. is proved.

4. Let

\[
S := \{ x \in \mathbb{N}_0 : p(x,p(y,z)) = p(p(x,y),z) \ \forall y,z \in \mathbb{N}_0 \}.
\]

Then \( 0 \in S \) and if \( x \in S \) we find,

\[
p(s(x),p(y,z)) = s(p(x,p(y,z))) = s(p(p(x,y),z)) = p(s(p(x,y),z)) = p(p(s(x),y),z)
\]

which shows that \( s(x) \in S \) and therefore \( S = \mathbb{N}_0 \) and item 4. is proved.

**Notation 1.22** We now write \( x + y \) for \( p(x,y) \) and refer to the symmetric binary operator, +, as addition.

To summarize we have now shown addition satisfies for all \( x, y, z \in \mathbb{N}_0 \):

1. \( x + 0 = 0 + x = x \),
2. \( s(x) + 1 = x + 1 = 1 + x \),
3. \( x + y = y + x \),
4. \( (x + y) + z = x + (y + z) \).
5. The induction hypothesis may now be written as; if \( S \subset \mathbb{N}_0 \) and \( n + 1 \in S \) whenever \( n \in S \), then \( S = \mathbb{N}_0 \).

**Proposition 1.23** (Additive Cancellation). If \( x, y, z \in \mathbb{N}_0 \) and \( x + z = y + z \), then \( x = y \).

**Proof.** Let \( S \) be those \( z \in \mathbb{N}_0 \) for which the statement \( x + z = y + z \) implies \( x = y \) holds. It is clear that \( 0 \in S \). Moreover if \( z \in S \) and \( x + (z + 1) = y + (z + 1) \) then \( (x+1) + z = (y+1) + z \) and so by the inductive hypothesis \( s(x) = x + 1 = y + 1 = s(y) \). Recall that \( s \) is one to one by assumption and therefore we may conclude \( x = y \) and we have shown \( s(z) \in S \). Therefore \( S = \mathbb{N}_0 \) and the proposition is proved.

**Definition 1.24.** Given \( x, y \in \mathbb{N}_0 \), we say \( x < y \) iff \( y = x + n \) for some \( n \in \mathbb{N} \) and \( x \leq y \) iff \( y = x + n \) for some \( n \in \mathbb{N}_0 \). We further let

\[
R_x := \{ x + n : n \in \mathbb{N}_0 \}
\]

so that \( y \geq x \) iff \( y \in R_x \).

**Proposition 1.25.** If \( x, y \in \mathbb{N}_0 \) and \( x \leq y \) and \( y \leq x \) then \( x = y \). Moreover if \( x \leq y \) then either \( x < y \) or \( x = y \).

**Proof.** By assumption there exists \( m, n \in \mathbb{N}_0 \) such that \( y = x + m \) and \( x = y + n \) and therefore \( y = y + (m + n) \). Hence by cancellation it follows that \( m + n = 0 \). If \( n \neq 0 \) then \( n = s(x) \) for some \( x \in \mathbb{N}_0 \) and we have \( m + n = m + s(x) = s(m + x) \in \mathbb{N} \) which would imply \( m + n \neq 0 \). Thus we conclude that \( m = 0 = n \) and therefore \( x = y \).

If \( x \leq y \) and \( y \neq x \) then \( y = x + n \) for some \( n \in \mathbb{N}_0 \) with \( n \neq 0 \), i.e. \( x < y \).

**Proposition 1.26.** If \( x, y \in \mathbb{N}_0 \) then precisely one of the following three choices must hold, 1) \( x < y \), 2) \( x = y \), 3) \( y < x \).

**Proof.** Suppose that \( x \leq y \) does not hold, i.e. \( y \notin R_x \). We wish to show that \( y < x \), i.e. that \( x = y + n \) for some \( n \in \mathbb{N} \). We do this by induction on \( y \). That is let \( S \) be the the set of \( y \in \mathbb{N}_0 \) such that the statement \( y \notin R_x \) implies \( y < x \) holds. If \( y = 0 \notin R_x \) implies \( n := x \neq 0 \) so that \( x = y + n \), i.e. \( y = 0 < x \). This shows \( 0 \in S \). Now suppose that \( y \in S \) and that \( y + 1 \notin R_x = \{ x + m : m \in \mathbb{N}_0 \} \). It follows that \( y + 1 \neq x + m + 1 \) for all \( m \in \mathbb{N}_0 \) and hence that \( y \neq x + m \) for all \( m \in \mathbb{N}_0 \), i.e. \( y \notin R_x \). So by induction \( y < x \) and therefore \( x = y + k \) for some \( k \in \mathbb{N} \). Since \( k \in \mathbb{N} \) we know there exists \( k' \in \mathbb{N}_0 \) such that \( k = k' \) and it follows that \( x = y + 1 + k' \), i.e. \( y + 1 \leq x \). Since \( y + 1 \notin R_x \) we may conclude that in fact \( y + 1 < x \) and therefore \( y + 1 \in S \). So by induction \( S = \mathbb{N}_0 \) and we have shown if \( x < y \) does not hold iff \( y \leq x \). Combining this statement with the Proposition 1.25 completes the proof.

We have now set up a satisfactory addition operations and ordering on \( \mathbb{N}_0 \). Our next goal is to define multiplication on \( \mathbb{N}_0 \).

**Theorem 1.27.** There exists a function \( M : \mathbb{N}_0 \times \mathbb{N}_0 \to \mathbb{N}_0 \) such that \( M(x,0) = 0 \) for all \( x \in \mathbb{N}_0 \) and \( M(x,y+1) = M(x,y) + x \) for all \( x, y \in \mathbb{N}_0 \). This function \( M \) satisfies the following properties:

1. \( M(x,0) = 0 = M(0,x) \) for all \( x \in \mathbb{N}_0 \),
2. \( M(x,1) = M(1,x) = x \) for all \( x \in \mathbb{N}_0 \),
3. \( M(x,y) = M(y,x) \) for all \( x, y \in \mathbb{N}_0 \),
4. \( M(x,y+z) = M(x,y) + M(x,z) \) for all \( x, y, z \in \mathbb{N}_0 \),
5. \( M(x, M(y,z)) = M(M(x,y),z) \) for all \( x, y, z \in \mathbb{N}_0 \).
4. Let \( M \) be a function from \( \mathbb{N}_0 \) to \( \mathbb{N}_0 \) satisfying \( M(0) = 0 \) and \( M(y + 1) = M(y) + x \) for all \( y \in \mathbb{N}_0 \). Taking \( M(0) := 0 \) shows \( 0 \in S \). Moreover if \( x \in S \) we define \( M_{x+1}(y) := M_x(y) + y \). Then \( M_{x+1}(0) = 0 \) and

\[
M_{x+1}(y + 1) = M_x(y + 1) + y + 1 = M_x(y) + x + y + 1
\]

while

\[
M_{x+1}(y) + (x + 1) = M_x(y) + y + x + 1 = M_{x+1}(y + 1).
\]

This shows that \( x + 1 \in S \) and so by induction \( S = \mathbb{N}_0 \) and we may now define \( M(x, y) := M_x(y) \) for all \( x, y \in \mathbb{N}_0 \). We now prove the properties of \( M \) stated above.

1. By construction \( M(x, 0) = 0 \) for all \( x \). Let \( S := \{ x \in \mathbb{N}_0 : M(0, x) = 0 \} \). Then \( 0 \in S \) and if \( x \in S \) we have

\[
M(0, x + 1) = M(0, x) + 0 = 0 + 0 = 0
\]

which shows \( x + 1 \in S \). Therefore by induction \( S = \mathbb{N}_0 \) and \( M(0, x) = 0 \) for all \( x \in \mathbb{N}_0 \).

2. \( M(x, 1) = M(x, 0 + 1) = M(x, 0) + x = 0 + x = x \) for all \( x \in \mathbb{N}_0 \). Let \( S := \{ x \in \mathbb{N}_0 : M(1, x) = x \} \). Then \( 0 \in S \) and if \( x \in S \) we have

\[
M(1, x + 1) = M(1, x) + 1 = x + 1
\]

which shows \( x + 1 \in S \). Therefore \( S = \mathbb{N}_0 \) and \( M(1, x) = x \) for all \( x \in \mathbb{N}_0 \).

3. Let \( S := \{ x \in \mathbb{N}_0 : M(x, \cdot) = M(\cdot, x) \} \). Then by items 1. and 2. we know that \( 0, 1 \in S \). Now suppose that \( x \in S \), then by construction,

\[
M(x + 1, y) = M_{x+1}(y) = M(x, y) + y
\]

while

\[
M(y, x + 1) = M(y, x) + y.
\]

The last two displayed equations along with the induction hypothesis shows \( x + 1 \in S \) and therefore \( S = \mathbb{N}_0 \) and item 3. is proved.

4. Let \( S \) denotes those \( x \in S \) such that \( M(x, y + z) = M(x, y) + M(x, z) \) for all \( y, z \in \mathbb{N}_0 \). Then \( 0, 1 \in S \) and if \( x \in S \) we have,

\[
M(x + 1, y + z) = M(x, y + z) + y + z
\]

\[
= M(x, y) + M(x, z) + y + z
\]

\[
= M(x, y) + y + M(x, z) + z
\]

\[
= M(x + 1, y) + M(x + 1, z)
\]

which shows \( x + 1 \in S \). Therefore \( S = \mathbb{N}_0 \) and we have proved item 4.

5. Let

\[
S := \{ x \in \mathbb{N}_0 : M(x, M(y, z)) = M(M(x, y), z) \forall y, z \in \mathbb{N}_0 \}.
\]

Then \( 0 \in S \) and if \( x \in S \) we find,

\[
M(x + 1, M(y, z)) = M(x, M(y, z)) + M(y, z)
\]

while

\[
M(M(x + 1, y), z) = M(M(x, y) + y, z) = M(M(x, y) + y, z) + M(y, z).
\]

The last two equations along with the induction hypothesis shows \( x + 1 \in S \) and therefore \( S = \mathbb{N}_0 \) and item 5. is proved.

\[\Box\]

**Notation 1.28** We now write \( x \cdot y \) for \( M(x, y) \) and refer to the symmetric binary operator, \( \cdot \), as multiplication.

To summarize Theorem 1.27 we have shown multiplication satisfies for all \( x, y, z \in \mathbb{N}_0 \):

1. \( x \cdot 0 = 0 = 0 \cdot x \),
2. \( x \cdot 1 = x = 1 \cdot x \),
3. \( x \cdot y = y \cdot x \),
4. \( x \cdot (y + z) = x \cdot y + x \cdot z \),
5. \( (x \cdot y) \cdot z = x \cdot (y \cdot z) \).

**Proposition 1.29 (Multiplicative Cancellation).** If \( x, y \in \mathbb{N}_0 \) and \( z \in \mathbb{N} \) such that \( x \cdot z = y \cdot z \), then \( x = y \).

**Proof.** If \( x \neq y \), say \( x < y \), then \( y = x + n \) for some \( n \in \mathbb{N} \) and therefore

\[
y \cdot z = (x + n) \cdot z = x \cdot z + n \cdot z.
\]

Hence if \( x \cdot z = y \cdot z \), then by additive cancellation we must have \( n \cdot z = 0 \). As \( n, x \in \mathbb{N} \) we may write \( n = n' + 1 \) and \( z = z' + 1 \) with \( n', z' \in \mathbb{N}_0 \) and therefore,

\[
0 = n \cdot z = (n' + 1) \cdot (z' + 1) = n' \cdot z' + n' + z' + 1 \neq 0
\]

which is a contradiction. \[\Box\]

**Remark 1.30 (Base 10 counting).** The typical method of counting is to use base 10 enumeration of \( \mathbb{N}_0 \). The rules are;

\[
0 = 0, \quad 1 = 1, \quad 2 := 1 + 1, \quad 3 := 2 + 1, \quad 4 := 3 + 1, \quad 5 := 4 + 1
\]

\[
6 := 5 + 1, \quad 7 := 6 + 1, \quad 8 := 7 + 1, \quad 9 := 8 + 1, \quad 10 := 9 + 1.
\]
Once these elements of $\mathbb{N}_0$ have been defined, then given $a_0, \ldots, a_n \in \{0, 1, \ldots, 9\}$ with $a_n \neq 0$, we let

$$a_na_{n-1} \ldots a_0 := \sum_{k=0}^{n} a_k 10^k.$$ 

For example, $35 = 3 \cdot 10 + 5 = 34 + 1$, etc.

As mentioned above one can formalize $\mathbb{Z}$ and $\mathbb{Q}$ using $\mathbb{N}_0$ constructed above. I will omit the details here and refer the reader to the references already mentioned.
Fields

The basic question we want to eventually address is: What are the real numbers? Our answer is going to be; the real numbers is the essentially unique complete ordered field, see Theorem 3.3 below. In order to make sense of this answer we need to explain the terms, “complete,” “ordered,” and “field.” We will start with the notion of a field which loosely stated means something that can reasonably be interpreted a “numbers.”

Definition 2.1 (Fields, i.e. “numbers”). A field is a non-empty set \( F \) equipped with two operations called addition and multiplication, and denoted by + and \( \cdot \), respectively, such that the following axioms hold; (subtraction and division are defined implicitly in terms of the inverse operations of addition and multiplication:)

1. **Closure of \( F \) under addition and multiplication.** For all \( a, b \in F \), both \( a + b \) and \( a \cdot b \) are in \( F \) (or more formally, + and \( \cdot \) are binary operations on \( F \)).

2. **Associativity of addition and multiplication.** For all \( a, b, \) and \( c \in F \), the following equalities hold: \( a + (b + c) = (a + b) + c \) and \( a \cdot (b \cdot c) = (a \cdot b) \cdot c \).

3. **Commutativity of addition and multiplication.** For all \( a \) and \( b \) in \( F \), the following equalities hold: \( a + b = b + a \) and \( a \cdot b = b \cdot a \).

4. **Additive and multiplicative identity.** There exists an element of \( F \), called the additive identity element and denoted by \( 0 = 0_F \), such that for all \( a \in F \), \( a + 0 = a \). Likewise, there is an element, called the multiplicative identity element and denoted by \( 1 = 1_F \), such that for all \( a \in F \), \( a \cdot 1 = a \). It is assumed that \( 0_F \neq 1_F \).

5. **Additive and multiplicative inverses.** For every \( a \) in \( F \), there exists an element \( -a \) in \( F \), such that \( a + (-a) = 0 \). Similarly, for any \( a \in F \) other than \( 0 \), there exists an element \( a^{-1} \) in \( F \), such that \( a \cdot a^{-1} = 1 \). (The elements \( a + (-b) \) and \( a \cdot b^{-1} \) are also denoted \( a - b \) and \( a/b \), respectively.) In other words, subtraction and division operations exist.

6. **Distributivity of multiplication over addition.** For all \( a, b \) and \( c \) in \( F \), the following equality holds: \( a \cdot (b + c) = (a \cdot b) + (a \cdot c) \).

(\(Note that all but the last axiom are exactly the axioms for a commutative group, while the last axiom is a compatibility condition between the two operations.)

2.1 Basic Properties of Fields

Here are some sample properties about fields. For more information about Fields see 5-8 of Rudin.

**Lemma 2.2.** Let \( F \) be a field, then;

1. There is only one additive and multiplicative inverses.
2. If \( x, y, z \in F \) with \( x \neq 0 \) and \( xy = xz \) then \( y = z \).
3. \( 0 \cdot x = 0 \) for all \( x \in F \).
4. If \( x, y \in F \) such that \( xy = 0 \) then \( x = 0 \) or \( y = 0 \).
5. \( (−x)y = −(xy) \).
6. \( −(−x) = x \) for all \( x \in F \).
7. \( (−x)(−y) = xy \) for all \( x, y \in F \).

**Proof.** We take each item in turn.

1. Suppose that \( x + y = 0 = x + y' \), then adding \( −x \) to both sides of this equation shows \( y = y' \). Taking \( y = −x \) then shows \( y = −x = y' \), i.e. additive inverses are unique. Similarly if \( x \neq 0 \) and \( xy = 1 \) then multiplying this equation by \( x^{-1} \) shows \( y = x^{-1} \) and so there is only one multiplicative inverse.

2. If \( xy = xz \) then multiplying this equation by \( x^{-1} \) shows \( y = z \).

3. \[ 0 \cdot x + x = 0 \cdot x + 1 \cdot x = (0 + 1) \cdot x = 1 \cdot x = x. \]

Adding \( −x \) to both side of this equation using associativity and commutativity of addition then implies \( 0 \cdot x = 0 \).

4. If \( x \in F \setminus \{0\} \) and \( y \in F \) such that \( xy = 0 \), then

\[ 0 = x^{-1} \cdot 0 = x^{-1} (xy) = (x^{-1}x) y = 1y = y. \]

5. \( (−x)y + xy = (−x + x) \cdot y = 0 \cdot y = 0 \implies (−x)y = −(xy) \).

6. Since \( (−x) + x = 0 \) we have \( −(−x) = x \).

7. \( (−x)(−y) = −(x \cdot (−y)) = −(−(xy)) = xy \) by 6.

**Example 2.3.** Here are a few examples of Fields;
1. \(F_2 = \{0, 1\}\) with \(+0 = 0 = 1 + 1\), and \(0 + 1 = 1 + 0 = 0\) and \(0 \cdot 1 = 1 \cdot 0 = 0 \cdot 0 = 0\) and \(1 \cdot 1 = 1\). In this case \(-1 = 1, 1^{-1} = 1\) and \(-0 = 0\).

2. \(\mathbb{Q}\) — the rational numbers with the usual addition and multiplication of fractions. \(\left(\frac{m}{n}\right)^{-1} = \frac{n}{m}\) if \(m \neq 0\) and \(-\frac{m}{n} = -\frac{n}{m}\).

3. \(\mathbb{Q}(t)\) where

\[
\mathbb{Q}(t) = \left\{ \frac{p(t)}{q(t)} \mid p(t) \text{ and } q(t) \text{ are polynomials over } \mathbb{Q} \ni q(t) \neq 0 \right\}.
\]

Again the multiplication and addition are as usual.

**Example 2.4.** \(\mathbb{Z}\) is not a field. For example, 2 has no multiplicative inverse in \(\mathbb{Z}\). The inverse to 2, 2\(^{-1}\), should be \(\frac{1}{2}\) but this is not in \(\mathbb{Z}\).

**Definition 2.5.** We say a map \(\varphi : \mathbb{Z} \rightarrow F\) is a (ring) homomorphism iff \(\varphi(1) = 1_{\mathbb{Z}}, \varphi(0) = 0_{\mathbb{Z}}\), and for all \(x, y \in \mathbb{Z}\):

\[\varphi(x + y) = \varphi(x) + \varphi(y) \text{ and } \varphi(xy) = \varphi(x) \varphi(y).\]

The assumption that \(\varphi(0) = 0\) is redundant since \(\varphi(0) = \varphi(0 + 0) = \varphi(0) + \varphi(0)\) and therefore \(\varphi(0) = 0\).

**Lemma 2.6.** For every field \(F\) there a unique (ring) homomorphism, \(\varphi : \mathbb{Z} \rightarrow F\). In fact, \(\varphi(n) = n1_{\mathbb{Z}}\) for all \(n \in \mathbb{Z}\) where \(0 \cdot 1_{\mathbb{Z}} = 0\),

\[
\varphi(n) = n1_{\mathbb{Z}} := \underbrace{1_{\mathbb{Z}} + \cdots + 1_{\mathbb{Z}}}_{n \text{ times}} \text{ if } n \in \mathbb{N} \text{ and }
\]

\[
(-n)1_{\mathbb{Z}} := -(n1_{\mathbb{Z}}) \text{ if } n \in \mathbb{N}.
\]

The map \(\varphi\) need not be injective as is seen by taking \(F = F_2\).

**Proof.** Let us first work on \(\mathbb{N}_0 \subset \mathbb{Z}\). We must define \(\varphi(0) = 0\) and \(\varphi(1) = 1\) and then \(\varphi\) inductively by \(\varphi(n + 1) = \varphi(n) + \varphi(1) = \varphi(n) + 1_{\mathbb{Z}}\) so that

\[
\varphi(n) = \underbrace{1_{\mathbb{Z}} + \cdots + 1_{\mathbb{Z}}}_{n \text{ times}}.
\]

We now write \(n1_{\mathbb{Z}}\) for \(\varphi(n)\) with the convention that \(01_{\mathbb{Z}} = 0_{\mathbb{Z}}\). For \(n \in \mathbb{N}\) we must set \(\varphi(-n) = -\varphi(n) = -(n1_{\mathbb{Z}})\). Thus we have \(\varphi(n) = n1_{\mathbb{Z}}\) for all \(n \in \mathbb{Z}\).

Now we must show \(\varphi\) is a homomorphism.

**Additive homomorphism:** First suppose that \(m, n \in \mathbb{N}_0\) and let

\[
S := \{m \in \mathbb{N}_0 : \varphi(mn) = \varphi(m) \varphi(n) \text{ for all } n \in \mathbb{N}_0\}.
\]

One easily sees that \(0 \in S\) and that \(1 \in S\) by construction. Moreover if \(m \in S\), then

\[
\varphi((m + 1) + n) = \varphi(m + n + 1) = \varphi(m) + \varphi(n + 1)
\]

\[
= \varphi(m) + \varphi(n) + 1_f = \varphi(m) + 1_f + \varphi(n) = \varphi(m + 1) + \varphi(n)
\]

which shows \(m + 1 \in S\). Therefore by induction, \(S = \mathbb{N}_0\) and \(\varphi(m + n) = \varphi(m) + \varphi(n)\) for all \(m, n \in \mathbb{N}_0\).

If \(m \in \mathbb{N}_0\) we have \(\varphi(-m) = -\varphi(m)\) by construction. If \(n > m \in \mathbb{N}_0\), then

\[
\varphi(n + (-m)) = \varphi(n) + (-\varphi(m)) = \varphi(n) + \varphi(-m).
\]

so that

\[
\varphi(n + (-m)) = \varphi(n) + (-\varphi(m)) = \varphi(n) + \varphi(-m).
\]

If \(n < m \in \mathbb{N}_0\), then

\[
\varphi(n + (-m)) = \varphi(n) - \varphi(m) = \varphi(n) + \varphi(-m).
\]

and if \(m, n \in \mathbb{N}_0\), then

\[
\varphi(-n + (-m)) = \varphi(-n + m) = -\varphi(n + m)
\]

\[
= -\varphi(n) + \varphi(m) = -\varphi(n) - \varphi(m)
\]

\[
= \varphi(-n) + \varphi(-m).
\]

Putting all of this together shows \(\varphi\) is an additive homomorphism.

**Multiplicative homomorphism:** First suppose that \(m, n \in \mathbb{N}_0\) and let

\[
S := \{m \in \mathbb{N}_0 : \varphi(mn) = \varphi(m) \varphi(n) \text{ for all } n \in \mathbb{N}_0\}.
\]

It is easily seen that 0, 1 \(\in S\). Moreover if \(m \in S\) and \(n \in \mathbb{N}_0\), then

\[
\varphi((m + 1) n) = \varphi(mn + n) = \varphi(mn) + \varphi(n)
\]

\[
= \varphi(m) \varphi(n) + \varphi(n) = (\varphi(m) + 1_f) \varphi(n)
\]

\[
= \varphi(m + 1) \varphi(n),
\]

which shows \(m + 1 \in S\). Therefore by induction, \(S = \mathbb{N}_0\) and \(\varphi(mn) = \varphi(m) \varphi(n)\) for all \(m, n \in \mathbb{N}_0\).

If \(m, n \in \mathbb{N}_0\), then

\[
\varphi((-m)n) = \varphi(-mn) = -\varphi(mn) = -[\varphi(m) \varphi(n)] = -\varphi(m) \varphi(n) = \varphi(-m) \varphi(n)
\]

and

\[
\varphi((-m)(-n)) = \varphi(mn) = \varphi(mn) = (\varphi(m))(-\varphi(n)) = -\varphi(m) \varphi(-n)
\]

which completes the verification that \(\varphi\) is a multiplicative homomorphism.
2.2 Ordered Fields

Definition 2.7 (Ordered Field). We say $\mathbb{F}$ is an ordered field if there exists, $P \subset \mathbb{F}$, called the positive elements, such that

Ord 1. $\mathbb{F}$ is the disjoint union of $P$, $\{0\}$, and $-P$, i.e. if $x \in \mathbb{F}$ then precisely one of following happens; $x \in P$, $x = 0$, or $-x \in P$.

Ord 2. $P + P \subset P$ and $P \cdot P \subset P$.

Lemma 2.8. Let $(\mathbb{F}, P)$ be an ordered field, then;

1. For all $x \in \mathbb{F} \setminus \{0\}$, $x^2 \in P$. In particular $1 = 1^2 \in P$.
2. If $x \in P$ and $y \in -P$ then $xy \in -P$.
3. If $x \in P$ then $x^{-1} \in P$.

Proof. If $x \in P$ then $x^2 \in P : P \subset P$ while if $x \in -P$ then $-x \in P$ and $x^2 = (-x)^2 \in P$. For item 3. we have $x \cdot x^{-1} = 1$.

Example 2.9. The field $\mathbb{F} = \{0, 1\}$ is not ordered. The only possible choice for $P$ is $P = \{1\}$ which does not work since $1 + 1 = 0 \notin P$.

Example 2.10. Take $\mathbb{F} = \mathbb{Q}$ and $P = \left\{ \frac{m}{n} : m, n > 0 \right\}$. This is in fact the unique choice we can make for $P$ in this case. Indeed suppose that $P$ is any order on $\mathbb{Q}$. By Lemma 2.8 we know $1 \in P$ and then by induction it follows that $\mathbb{N} \subset P$. Then again by Lemma 2.8 we must have $m \cdot n^{-1} \in P$ for all $m, n \in \mathbb{Q}$.

Example 2.11. Take $\mathbb{F} = \mathbb{Q}(t)$ and

$$P = \left\{ \frac{p(t)}{q(t)} : \frac{p(t)}{q(t)} > 0 \text{ for } t > 0 \text{ large} \right\},$$

i.e. $\frac{p(t)}{q(t)} \in P$ iff the highest order coefficients of $p(t)$ and $q(t)$ have the same sign. For example $\frac{t^2 - 25t + 7}{t^4 - 10t^2 + 1} \in P$ while $\frac{t^2 + 25t + 7}{t^4 - 10t^2 + 1} \in -P$.

Notice that $t > n$ for all $n \in \mathbb{N}$ and $\frac{1}{2} < \frac{1}{n}$ for all $n \in \mathbb{N}$. This is kind of strange and explains why you have to prove the “obvious” in this course!!!

Moral: obvious statements are often false.

Notation 2.12 (Max and Min) We will often use the following notation in the sequel. If $a, b$ are elements of an ordered field, let

$$a \wedge b := \min \{a, b\} = \begin{cases} a & \text{if } a \leq b \\ b & \text{if } b \leq a \end{cases}$$

and

$$a \lor b := \max \{a, b\} = \begin{cases} b & \text{if } a \leq b \\ a & \text{if } b \leq a \end{cases}.$$

More generally if $\left\{ a_i \right\}_{i=1}^n \subset \mathbb{F}$ we let

$$a_1 \wedge \cdots \wedge a_n := \min \{a_1, \ldots, a_n\} \text{ and } a_1 \lor \cdots \lor a_n := \max \{a_1, \ldots, a_n\}.$$

be the smallest and largest element in the finite list $(a_1, \ldots, a_n)$.

Definition 2.13. Suppose that $\mathbb{F}$ and $\mathbb{G}$ are fields. A map, $\varphi : \mathbb{F} \to \mathbb{G}$ is a (field) homomorphism iff $\varphi(1_F) = 1_G$, $\varphi(0_F) = 0_G$, and for all $x, y \in \mathbb{F}$;

$$\varphi(x + y) = \varphi(x) + \varphi(y) \text{ and } \varphi(xy) = \varphi(x) \varphi(y).$$

Lemma 2.14 ($\mathbb{Q}$ embeds into an ordered field). For every ordered field $\mathbb{F}$, there a unique field homomorphism, $\varphi : \mathbb{Q} \to \mathbb{F}$. In fact,

$$\varphi \left( \frac{m}{n} \right) = \frac{m}{n} \cdot 1_F := m1_F \cdot (n1_F)^{-1} \quad (2.1)$$

$n$ times

where $n1_F := 1_F + \cdots + 1_F$ and $(-n)1_F := -(n1_F)$ for $n \in \mathbb{N}$ and $0 \cdot 1_F = 0_F$. Moreover;

1. $\varphi(x) \in P$ whenever $x > 0$,
2. and $\varphi$ is injective. Thus we may identify $\mathbb{Q}$ with $\varphi(\mathbb{Q})$ and consider $\mathbb{Q}$ as a sub-field of $\mathbb{F}$.

[In particular, ordered fields must be fields with an infinite number of elements in it.]

Proof. From Lemma 2.6 we know there is a unique ring homomorphism, $\varphi : \mathbb{Z} \to \mathbb{F}$, given by $\varphi(m) = m \cdot 1_F$. So for $m \in \mathbb{Z}$ and $n \in \mathbb{N}$ we must have

$$\varphi \left( \frac{m}{n} \right) \cdot n1_F = \varphi \left( \frac{m}{n} \right) \cdot \varphi(n) = \varphi \left( \frac{m}{n} \cdot n \right) = \varphi(m) = m1_F$$

which forces us to define $\varphi$ as in Eq. (2.1). Notice that is easy to verify by induction that $n1_F = \varphi(n) \in P$ for all $n \in \mathbb{N}$ and in particular $n1_F \neq 0$ for $n \in \mathbb{N}$. In particular if $x = m/n > 0$ then $\varphi \left( \frac{m}{n} \right) = m1_F \cdot (n1_F)^{-1} \in P$ by Lemma 2.8. We must still check that $\varphi$ is well defined homomorphism.

Well defined. Suppose that $k \in \mathbb{N}$, we must show

$$(km)1_F \cdot ((kn)1_F)^{-1} = m1_F \cdot (n1_F)^{-1}.$$

By cross multiplying, this will happen iff

$$(km)1_F \cdot (n1_F) = ((kn)1_F) \cdot m1_F.$$
which is the case as $\varphi : \mathbb{Z} \to F$ is a ring homomorphism.

**Homomorphism property.** We have

$$\varphi \left( \frac{m}{n} + \frac{p}{n} \right) = \varphi \left( \frac{m + p}{n} \right) = \varphi (m + p) \cdot \varphi (n)^{-1}$$

$$= [\varphi (m) + \varphi (p)] \cdot \varphi (n)^{-1}$$

$$= \varphi (m) \cdot \varphi (n)^{-1} + \varphi (p) \cdot \varphi (n)^{-1}$$

$$= \varphi \left( \frac{m}{n} \right) + \varphi \left( \frac{p}{n} \right)$$

and

$$\varphi \left( \frac{m}{n} \right) \varphi \left( \frac{q}{p} \right) = \varphi (m) \varphi (n)^{-1} \varphi (q) \varphi (p)^{-1}$$

$$= \varphi (m) \varphi (q) [\varphi (n) \varphi (p)]^{-1}$$

$$= \varphi (mq) [\varphi (np)]^{-1} = \varphi \left( \frac{mq}{np} \right).$$

**Injectivity.** If $0 = \varphi \left( \frac{m}{n} \right)$ then

$$0 = \varphi (m) \cdot \varphi (n)^{-1}$$

which implies $\varphi (m) = 0$ which happens iff $m = 0$, i.e. $m/n = 0$.

---

**End of Lecture 3, 10/3/2012**

**Notation 2.15** If $(F, P)$ is an ordered field we write $x > y$ if $x - y \in P$. We also write $x \geq y$ if $x - y \in P$ or $x = y$.

Notice that if $x, y \in \mathbb{F}$ then either $x - y = 0$ (i.e. $x = y$), or $x - y \in \mathbb{P}$ (i.e. $x > y$), or $x - y \in -\mathbb{P}$ (i.e. $y - x \in \mathbb{P}$ and $y > x$). Also in this notation we have $P = \{x \in F : x > 0\}$,

$$-P = \{x \in F : x < 0 \text{ (i.e. } 0 > x)\}.$$  

**Lemma 2.16.** Suppose that $x < y$ and $y < z$ and $a > 0$. Then $x < z$ and $ax < ay$.

**Proof.** By assumption $y - x \in P$ and $z - y \in P$, therefore $z - x = (y - x) + (z - y) \in P$, i.e. $z > x$. Moreover, $a \in P$ and $(y - x) \in P$ implies

$$P \ni a(y - x) = ay - ax.$$

That is $ay > ax$.

**Exercise 2.1.** Let $(F, P)$ be an ordered field and $x, y \in F$ with $y > x$. Show:

1. $y + a > x + a$ for all $a \in \mathbb{F}$,
2. $-x > -y$,
3. if we further suppose $x > 0$, show $\frac{1}{x} > \frac{1}{y}$.

**Definition 2.17.** Given $x \in \mathbb{F}$, we say that $y \in \mathbb{F}$ is a square root of $x$ if $y^2 = x$.

[From Lemma 2.8 it follows that if $x \in \mathbb{F}$ has a square root then $x \geq 0$.]

**Lemma 2.18.** Suppose $x, y \in \mathbb{F}$ with $x^2 = y^2$, then either $x = y$ or $x = -y$. In particular, there are at most 2 square roots of any number $x \geq 0$ in $\mathbb{F}$.

**Proof.** Observe that

$$(x - y)(x + y) = (x - y) x + (x - y) y$$

$$= x^2 - xy + xy - y^2 = x^2 - y^2 = 0.$$

Thus it follows that either $x - y = 0$ or $x + y = 0$, i.e. $x = y$ or $x = -y$.

**Definition 2.19.** If $x > 0$ admits a square root we let $\sqrt{x}$ be the unique positive root. We also define $\sqrt{0} = 0$.

**Lemma 2.20.** Suppose that $0 < x < y$, i.e. $x, y - x \in P$, then $x^2 < y^2$.

**Proof.** By Lemma 2.16 we know $x \cdot x < x \cdot y$ and $x \cdot y < y \cdot y$ and therefore $x^2 < y^2$.

**Corollary 2.21.** If $0 \leq x < y$ and $\sqrt{x}$ and $\sqrt{y}$ exists, then $0 \leq \sqrt{x} < \sqrt{y}$.

**Proof.** If $\sqrt{x} = \sqrt{y}$ then $x = (\sqrt{x})^2 = (\sqrt{y})^2 = y$ which is impossible. Similarly if $\sqrt{x} > \sqrt{y}$ then

$$x = (\sqrt{x})^2 > (\sqrt{y})^2 = y$$

which is again false.

**Alternatively:** starting with $y^2 - x^2 = (y - x)(y + x)$ and then replacing $y$ and $x$ by $\sqrt{y}$ and $\sqrt{x}$ respectively (assuming they exist) shows,

$$y - x = (\sqrt{y} - \sqrt{x})(\sqrt{y} + \sqrt{x}) \implies \sqrt{y} - \sqrt{x} = (y - x)(\sqrt{y} + \sqrt{x})^{-1}$$

from which it follows that $\sqrt{y} - \sqrt{x} \in P$ if $(y - x) \in P$. More importantly this shows $\sqrt{y}$ depends “continuously” in on $y$.

**Definition 2.22.** The absolute value, $|x|$, of $x$ in ordered field $F$ is defined by

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

**Alternatively we may define**

$$|x| = \sqrt{x^2}.$$
Proposition 2.23. For all $x, y \in \mathbb{F}$, then

1. $|x| \geq 0$
2. $|xy| = |x| |y|$
3. $|x + y| \leq |x| + |y|.$

Proof. 1 holds by definition since $-x > 0$ if $x < 0$.
2. As $|x| |y| \geq 0$ and $(|x| |y|)^2 = |x|^2 |y|^2 = x^2 y^2 = (xy)^2$, we have
   $$|x| |y| = \sqrt{(xy)^2} = |xy|.$$
3. It suffices to show $|x + y|^2 \leq (|x| + |y|)^2$. However,
   $$|x + y|^2 = (x + y)^2 = x^2 + y^2 + 2xy \leq x^2 + y^2 + 2|x||y| \quad (xy \leq |xy|)$$
   $$= |x|^2 + |y|^2 + 2|x||y|
   = (|x| + |y|)^2.$$

Definition 2.24. Let $(\mathbb{F}, P)$ be an ordered field and $S$ be a subset of $\mathbb{F}$.

1. We say that $S \subset \mathbb{F}$ is bounded from above (below) if there exists $x \in \mathbb{F}$ such that $x \geq s \ (x \leq s)$ for all $s \in S$. Any such $x$ is called an upper (lower) bound of $S$.
2. If $S$ is bounded from above (below), we say that $y \in \mathbb{F}$ is a least upper bound (greatest lower bound) for $S$ if $y$ is an upper (lower) bound for $S$ and $y \leq x \ (y \geq x)$ for any other upper (lower) bound, $x$, of $S$.

Notice that least upper bounds and greatest lower bounds are unique if they exist. We will write and

$$y = \text{l.u.b.} \ (S) = \sup \ (S)$$
if $y$ is the least upper bound for $S$ and

$$y = \text{g.l.b.} \ (S) = \inf \ (S)$$
if $y$ is the greatest lower bound for $S$.

Example 2.25. Let $\mathbb{F} = \mathbb{Q}$, then;

1. maximal $(a, b)$ and minimal $(a, b)$ are least upper respectively greatest lower bounds respectively for $S = \{a, b\}$. More generally, if $S = \{a_1, \ldots, a_n\}$, then
   $$\sup \ (S) = a_1 \lor \cdots \lor a_n := \max \ (a_1, \ldots, a_n) \text{ and}$$
   $$\inf \ (S) = a_1 \land \cdots \land a_n := \min \ (a_1, \ldots, a_n).$$

2. $S = \mathbb{N}$ is not bounded from above while $\inf \ (S) = 1$.
3. $S = \{1 - \frac{1}{n} : n \in \mathbb{N}\}$ is bounded from above and $1 = \sup \ (S)$ while $\inf \ (S) = 0$.
4. Let
   $$S = \{1, 1.4, 1.41, 1.414, 1.4142, 1.41421, 1.414213, 1.4142135, \ldots\}$$
where I am getting these digits from the decimal expansion of $\sqrt{2};$

$$\sqrt{2} \approx 1.41421356237309504880168872420969807565967187537694807317667973799.$$

In this case $S$ is bounded above by 2, or 1.42, or 1.415, etc. Nevertheless $\sqrt{2} = \sup \ (S)$ does not exist in $\mathbb{Q}$.

Example 2.26. Now let $\mathbb{F} = \mathbb{Q} \ (t)$ be the field of rational functions described in Example 2.11 then; $S = \mathbb{N}$ is bounded from above. For example $t$ is an upper bound but there is not least upper bound. For example $\frac{1}{m} t$ is also an upper bound for $S$.

End of Lecture 4, 10/5/2012

Definition 2.27 (Dedekind Cuts). A subset $\alpha \subset \mathbb{Q}$ is called a cut (see [?, p. 17]) if:

1. $\alpha$ is a proper subset of $\mathbb{Q}$, i.e. $\alpha \neq \emptyset$ and $\alpha \neq \mathbb{Q}$,
2. if $p \in \alpha$ and $q \in \mathbb{Q}$ and $q < p$, then $q \in \alpha$,
3. if $p \in \alpha$, then there exists $r \in \alpha$ with $r > p$.

Example 2.28. To each $a \in \mathbb{Q}$, let $\alpha_a := \{q \in \mathbb{Q} : q < a\}$. Then $\alpha_a$ is a cut and $a$ is the least upper bound of $\alpha_a$ in $\mathbb{Q}$.

Example 2.29. Let $\{S_n\}_{n=0}^\infty \subset \mathbb{Q}$ be any bounded sequence such that $S_n \leq S_{n+1}$ for all $n$. Then

$$\alpha := \bigcup_{n=0}^\infty \alpha_{S_n} = \{q \in \mathbb{Q} : q < S_n \text{ a.a. } n\}$$
is a cut as the reader should verify. Let us further suppose that $\lim_{n \to \infty} S_n$ does not exist in $\mathbb{Q}$. [For example from Example 1.17 we may take $S_n := \sum_{k=0}^n \frac{1}{k} \in \mathbb{Q}$.] If $m \in \mathbb{Q}$ is an upper bound for $\alpha$, then $m \geq S_n$ for all $n$ since if $m < S_n$ for some $n$ then $q := \frac{1}{2} (m + S_n) \in \alpha$ with $q > m$. Since $\lim_{n \to \infty} S_n \not= m$ as $m \in \mathbb{Q}$ there must exists $\varepsilon > 0$ such that

$$m - S_n = |m - S_n| \geq \varepsilon \text{ i.o. } n.$$

As $m - S_n$ is decreasing we may conclude that $m - S_n \geq \varepsilon$ for all $n$, i.e. $S_n \leq m - \varepsilon$ for all $n$. From this it now follows that $m - \varepsilon$ is an upper bound for $\alpha$ which is strictly smaller that $m$. So there can be no least upper bound.
We then define \( \alpha \) numbers.

\( \{ a \in \mathbb{R} : \alpha > 0 \} \) where we say \( \alpha > 0 \) iff there exists an \( N \in \mathbb{N} \) such that \( a_n > \frac{1}{N} \) for a.a. \( n \).

3. The ordered field \( (\mathbb{R}, +) \) is complete, i.e. has the least upper bound property.

The proof of Theorem 3.6 and Theorem 3.3 will be relegated to Section 3.6 at the end of this chapter. For an alternative existence proof of \( \mathbb{R} \) using Dedekind cuts as the elements of \( \mathbb{R} \) is covered in Rudin [? , pages 17-21]. One may also construct the Real numbers using decimal expansions, see [L. Gower’s notes on real numbers as decimals]. We will prove the uniqueness assertion of Theorem 3.3 in Section at the end of this section. From now on we are going to take Theorem 3.3 for granted and derive from this the “familiar” properties of the real numbers.

Observe that \( \mathbb{Q} \), \( \mathbb{Q}(t) \), \( \mathbb{R}(t) \) are not complete and hence are not the real numbers, \( \mathbb{R} \). For example \( \mathbb{N} \subset \mathbb{Q}(t) \) (or \( \mathbb{N} \subset \mathbb{R}(t) \) ) is bounded by \( t \) say but has no least upper bound. However, we do know that \( \mathbb{Q} \subset \mathbb{R} \) by Lemma 2.14. We will soon see that \( \mathbb{Q} \) is “dense” in \( \mathbb{R} \). We now pause to discuss some of the basic properties of \( \mathbb{R} \).

**Theorem 3.7.** Suppose that \( \mathbb{R} \) is a complete ordered field which we assume we have already embedded \( \mathbb{Q} \) into \( \mathbb{R} \) as in Lemma 2.14. Then:

1. With these definitions, \( \mathbb{R} \), satisfies the axioms of a field.
2. Moreover, we can make this into an ordered field by setting \( P := \{ \alpha \in \mathbb{R} : \alpha > 0 \} \) where we say \( \alpha > 0 \) iff there exists an \( N \in \mathbb{N} \) such that \( a_n > \frac{1}{N} \) for a.a. \( n \).
3. The ordered field \( (\mathbb{R}, +) \) is complete, i.e. has the least upper bound property.

### Real Numbers

As we saw in Section 1.2, \( \mathbb{Q} \) is full of holes and calculus tends to produce answers which live in these holes. So it is imperative that we fill the holes. Doing so will lead to the real numbers provided we fill in the holes without adding too much extra filler along the way. One good answer to the question, What are the real numbers?, is contained in the statement of Theorem 3.3.

**Definition 3.1.** An order preserving field isomorphism between two ordered fields, \( (\mathbb{F}, +) \) and \( (\mathbb{F}', +) \), is a bijection, \( f : \mathbb{F} \to \mathbb{F}' \) such that \( f(0) = 0, f(1) = 1, f(P_1) = P_2, \) and

\[
f(x + y) = f(x) + f(y) \quad \text{and} \quad f(xy) = f(x)f(y) \quad \text{for all} \quad x, y \in \mathbb{F}.
\]

**Definition 3.2.** An ordered field \( (\mathbb{F}, +) \) is has the least upper bound property if every non-empty subset, \( S \subset \mathbb{F} \), which is bounded from above possesses a least upper bound in \( \mathbb{F} \). [As we have seen in examples above, \( \mathbb{Q} \) does not have the least upper bound property.]

**Theorem 3.3 (The real numbers).** Up to order preserving field isomorphism (see Definition 3.7), there is exactly one complete ordered field. It is this field that we refer to as the real numbers and denote by \( \mathbb{R} \).

**Definition 3.4.** We say two Cauchy sequences \( \{a_n\}_{n=1}^{\infty} \) and \( \{b_n\}_{n=1}^{\infty} \) of rational numbers are equivalent and write \( \{a_n\}_{n=1}^{\infty} \sim \{b_n\}_{n=1}^{\infty} \) iff

\[
\lim_{n \to \infty} |a_n - b_n| = 0.
\]

We then define \( \alpha := \{\{a_n\}_{n=1}^{\infty}\} \) to be the equivalence class of the Cauchy sequence \( \{a_n\}_{n=1}^{\infty} \) and refer to the collection of these equivalence classes as the real numbers. The set of real numbers will be denoted by \( \mathbb{R} \).

**Notation 3.5** Let \( i : \mathbb{Q} \to \mathbb{R} \) be defined by \( i(a) := \{[a, a, a, \ldots]\}, \) i.e. \( i(a) \) is the equivalence class of the constant sequence \( a \).

Notice that if \( i(a) = i(b) \) iff \( a = \lim_{n \to \infty} a = \lim_{n \to \infty} b = b \). Thus the map, \( i : \mathbb{Q} \to \mathbb{R} \) is injective and we will often simply identify \( a \) with \( i(a) \) and in this way consider \( \mathbb{Q} \) as a subset of \( \mathbb{R} \).

**Theorem 3.6.** Let \( \mathbb{R} \) be as in Theorem 3.3. For \( \alpha := \{\{a_n\}_{n=1}^{\infty}\} \) and \( \beta := \{\{b_n\}_{n=1}^{\infty}\} \) in \( \mathbb{R} \) we define

\[
\alpha + \beta = \{\{a_n + b_n\}_{n=1}^{\infty}\} \quad \text{and} \quad \alpha \cdot \beta = \{\{a_n \cdot b_n\}_{n=1}^{\infty}\}.
\]

1. For all \( x \geq 0 \) there exists \( n \in \mathbb{N} \) such that \( n \geq x \).
2. For all \( \varepsilon > 0 \) in \( \mathbb{R} \) there exists \( n \in \mathbb{N} \) such that \( 0 < \frac{1}{n} \leq \varepsilon \).
3. If \( \varepsilon \geq 0 \) satisfies \( \varepsilon \leq 1/n \) for all \( n \in \mathbb{N} \) then \( \varepsilon = 0 \).
4. If \( a, b \in \mathbb{R} \) and \( a \leq b + \frac{1}{n} \) for all \( n \in \mathbb{N} \) or \( a \leq b + \varepsilon \) for all \( \varepsilon > 0 \), then \( a \leq b \).

**Proof.** We take each item in turn.

1. If \( n < x \) for all \( n \in \mathbb{N} \), then \( \mathbb{N} \) is bounded from above and so \( a := \sup(\mathbb{N}) \) exists in \( \mathbb{R} \) by the completeness axiom. As \( a \) is the least upper bound for \( \mathbb{N} \) there must be an \( n \in \mathbb{N} \) such that \( n > a - 1 \). However this implies \( n + 1 > a \) which violates \( a \) be an upper bound for \( \mathbb{N} \).

Roughly speaking here, you should think of \( \alpha = \lim_{n \to \infty} a_n \) and so \( \alpha > 0 \) should happen iff \( \alpha > \frac{1}{N} \) for some \( N \in \mathbb{N} \) which then implies \( a_n \geq \frac{1}{N} \) for a.a. \( n \).
2. If $\varepsilon > 0$ in $\mathbb{R}$ and $\frac{1}{n} > \varepsilon$ for all $n \in \mathbb{N}$, then $n < \frac{1}{\varepsilon}$ for all $n \in \mathbb{N}$ which is impossible by item 1.

3. If there exists $\varepsilon > 0$ such that $\varepsilon \leq \frac{1}{n}$ for all $n$ then $n \leq 1/\varepsilon$ for all $n$ which is again impossible by item 1.

4. It suffices to prove the first assertion. We may assume $a \geq b$ for otherwise we are done. If $a \leq b + \frac{1}{n}$ for all $n$, then $0 \leq a - b \leq \frac{1}{n}$ for all $n \in \mathbb{N}$ and hence $a = b$ and in particular $a \leq b$.

Proposition 3.8. If $\mathbb{R}$ is a complete ordered field, then every subset $S \subset \mathbb{R}$ which is bounded from below has a greatest lower bound, $\text{glb}(S) = \inf(S)$. In fact,

$$\inf(S) = -\sup(-S).$$

Proof. We let $m := -\sup(-S)$. Then we have $-s \leq m$ for all $s \in S$, i.e. $s \geq m$ for all $s \in S$ so that $m$ is a lower bound for $S$. Moreover if $\varepsilon > 0$ is given there exists $s \in S$ such that $-s \geq -m - \varepsilon$, i.e. $s \leq m + \varepsilon$. This shows that any lower bound, $m$ of $S$ must satisfy, $k \leq m + \varepsilon$ for all $\varepsilon > 0$, i.e. $k \leq m$. This shows that $m$ is the greatest lower bound for $S$.

Let me sketch one way to construct $\mathbb{R}$ based on Cauchy sequences of rational numbers.

Definition 3.9. A sequence $\{q_n\}_{n=1}^{\infty} \subset \mathbb{R}$ converges to $0 \in \mathbb{R}$ if for all $\varepsilon > 0$ in $\mathbb{R}$ there exists $N \in \mathbb{N}$ such that $|q_n| \leq \varepsilon$ for all $n \geq N$. Alternatively put, for all $M \in \mathbb{R}$ we have $|q_n| \leq \frac{1}{M}$ for a.a. $n$.

Definition 3.10. A sequence $\{q_n\}_{n=1}^{\infty} \subset \mathbb{R}$ converges to $q \in \mathbb{R}$ 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 3.11. A sequence $\{q_n\}_{n=1}^{\infty} \subset \mathbb{R}$ is Cauchy if $|q_n - q_m| \to 0$ as $m, n \to \infty$. More precisely we require for each $\varepsilon > 0$ in $\mathbb{R}$ that $|q_m - q_n| \leq \varepsilon$ for a.a. pairs $(m, n)$, i.e. there should exists $N \in \mathbb{N}$ such that $|q_m - q_n| \leq \varepsilon$ for all $m, n \geq N$.

The next few results are analogous to what you have already shown in the case $\mathbb{Q}$ is replaced by $\mathbb{Q}$. As the proofs are essentially identical to those of Theorem 1.13 and Exercise 1.6.

Proposition 3.12. If $\{a_n\}_{n=1}^{\infty} \subset \mathbb{R}$ is a convergent sequence then it is Cauchy\(^2\) \footnote{We are going to show shortly that the converse is true as well!}

If $\{a_n\}_{n=1}^{\infty}$ is Cauchy sequence then it is bounded.

---

Theorem 3.13 (Basic Limit Results). Suppose that $\{a_n\}$ and $\{b_n\}$ are sequences of real numbers such that $A := \lim_{n \to \infty} a_n$ and $B := \lim_{n \to \infty} b_n$ exists in $\mathbb{R}$. Then:

1. $\lim_{n \to \infty} |a_n| = |A|.$
2. If $A \neq 0$ then $\lim_{n \to \infty} \frac{1}{a_n} = \frac{1}{A}.$
3. $\lim_{n \to \infty} (a_n + b_n) = A + B.$
4. $\lim_{n \to \infty} (a_n b_n) = A \cdot B.$
5. If $a_n \leq b_n$ for all $n$, then $A \leq B$.
6. If $\{x_n\} \subset \mathbb{R}$ is another sequence such that $a_n \leq x_n \leq b_n$ and $A = B$, then $\lim_{n \to \infty} x_n = A = B$.

Theorem 3.14. If $S \subset \mathbb{R}$ is a non-empty set which is bounded from above, then there exists $\{x_n\}_{n=1}^{\infty} \subset S$ such that $x_n \to \sup S$ as $n \to \infty$, i.e. $x_n \leq x_{n+1}$ for all $n$ and $\lim_{n \to \infty} x_n = \sup S$.

Proof. Let $M := \sup S$. For each $n \in \mathbb{N}$, there exists $y_n \in S$ such that $M \geq y_n \geq M - \frac{1}{n}$. We now let $x_n := \max \{y_1, \ldots, y_n\}$ in which case $M \geq x_n \geq M - \frac{1}{n}$ and $x_n$ is increasing. By the Sandwich theorem it follows that $\lim_{n \to \infty} x_n = M$.

---

End of Lecture 5, 10/8/2012

Theorem 3.15. If $\{x_n\}_{n=1}^{\infty} \subset \mathbb{R}$ is bounded from above and $x_n$ is non-decreasing, then $\lim_{n \to \infty} x_n = \sup_n x_n$. Similarly if $\{x_n\}_{n=1}^{\infty} \subset \mathbb{R}$ is bounded from below and $x_n$ is non-increasing, then $\lim_{n \to \infty} x_n = \inf_n x_n$.

Proof. Let $M := \sup_n x_n$, then for all $\varepsilon > 0$, there exists $N_\varepsilon \in \mathbb{N}$ such that $M \geq x_n \geq M - \varepsilon$. As $x_n$ is non-decreasing it follows that $M \geq x_n \geq M - \varepsilon$ for all $n \geq N_\varepsilon$, i.e.

$$|M - x_n| \leq \varepsilon$$

for all $n \geq N_\varepsilon$. 
As \( \varepsilon > 0 \) was arbitrary, we may conclude that \( \lim_{n \to \infty} x_n = M \). If \( x_n \) is decreasing instead, then \( -x_n \uparrow \) and we have \( -\lim_{n \to \infty} x_n = \sup_n (-x_n) = -\inf_n x_n \).

**Theorem 3.16.** Suppose that \( \mathbb{R} \) is a complete ordered field which we assume we have already embedded \( \mathbb{Q} \) into \( \mathbb{R} \) as in Lemma 2.14. Then:

1. For all \( m \in \mathbb{R} \), if \( \alpha_m := \{ y \in \mathbb{Q}: y < m \} \),
   then \( \sup \alpha_m = m \).

2. If \( a, b \in \mathbb{R} \) with \( a < b \), then there exists \( q \in \mathbb{Q} \) such that \( a < q < b \).

3. If \( \alpha \subseteq \mathbb{Q} \) is a cut and \( m := \sup \alpha \), then \( \alpha = \alpha_m \).

**Proof.** We take each item in turn.

1. **Proof 1.** Let \( \alpha_m := \{ y \in \mathbb{Q}: y < m \} \) and \( M := \sup \alpha_m \in \mathbb{R} \). Then \( M \leq m \).
   If \( M \neq m \) then \( M < m \). To see this last case is not possible \( \varepsilon := m - M > 0 \) and choose \( n \in \mathbb{N} \) such that \( 0 < \frac{1}{n} \leq \varepsilon \). Then choose \( y \in \mathbb{Q} \) such that
   \[
   M - \frac{1}{2n} < y < M.
   \]
   From this it follows that
   \[
   M < y + \frac{1}{2n} < M + \frac{1}{2n} < m
   \]
   which shows \( y + \frac{1}{2n} \in \alpha_m \) is greater than \( M \) violating the assumption that \( M \) is an upper bound for \( \alpha_m \).

2. **Proof 2.** [Here is a slight rewriting of the above argument.] Choose \( y_m \in \alpha_m \) such that \( y_m \uparrow M \) as \( m \to \infty \). Choose \( n \in \mathbb{N} \) so that \( m - M > \frac{1}{n} \). Then
   \[
   y_m + \frac{1}{n} > M + \frac{1}{n} < m \quad \text{as} \quad m \to \infty.
   \]
   So for large \( m \), \( y_m + \frac{1}{n} < m \) while \( y_m + \frac{1}{n} \uparrow M \), i.e. \( y_m + \frac{1}{n} \in \alpha_m \) yet \( y_m + \frac{1}{n} > M \). This violates the assumption that \( M \) is an upper bound for \( \alpha_m \).

3. **Proof 3.** Choose \( n \in \mathbb{N} \) so that \( \alpha \in \alpha_m \).

**Exercise 3.1.** Suppose that \( \alpha \subseteq \mathbb{Q} \) is a cut as in Definition 2.27. Show \( \alpha \) is bounded from above. Then let \( m := \sup \alpha \) and show that \( \alpha = \alpha_m \), where \( \alpha_m := \{ y \in \mathbb{Q}: y < m \} \).

Also verify that \( \alpha_m \) is a cut for all \( m \in \mathbb{R} \). [In this way we see that we may identify \( \mathbb{R} \) with the cuts of \( \mathbb{Q} \). This should motivate Dedekind’s construction of the real numbers as described in Rudin.]

**Proposition 3.17 (Rationals are dense in the reals).** For all \( b \in \mathbb{R} \), there exists \( q_n \in \mathbb{Q} \) such that \( q_n \uparrow b \). Similarly there exists \( p_n \in \mathbb{Q} \) such that \( p_n \downarrow b \).

**Proof.** Given \( b \in \mathbb{Q} \) we know that \( b = \sup \alpha_b \) by Theorem 3.16. Then by Theorem 3.14 there exists \( q_n \in \alpha_b \) such that \( q_n \uparrow b \) as \( n \to \infty \). The second assertion can be proved in much the same way as the first. Alternatively, let \( q_n \in \mathbb{Q} \) such that \( q_n \uparrow b \) and set \( p_n := -q_n \in \mathbb{Q} \). Then \( p_n \downarrow b \).

**Definition 3.18.** The real numbers which are not rational are called irrational\(^3\) so the irrational numbers are \( \mathbb{R} \setminus \mathbb{Q} \).

**Example 3.19 (Euler’s number).** Let \( S_n := \sum_{k=0}^{n} \frac{1}{k!} \) for all \( n \in \mathbb{N}_0 \). We define Euler’s number to be,

\[
e := \lim_{n \to \infty} S_n = \sup \{ S_n : n \in \mathbb{N}_0 \} \in \mathbb{R}.
\]

From Example 1.1.7 we have seen that \( e \in \mathbb{R} \setminus \mathbb{Q} \).

**Theorem 3.20 (\( n^{th} \) roots).** Let \( n \in \mathbb{N} \) and \( x > 0 \) in \( \mathbb{R} \), then there exists a unique \( y \in \mathbb{R}_{+} \) such that \( y^n = x \). We of course denote \( y \) by \( x^{1/n} \) for \( \sqrt[n]{x} \). The function \( x \to x^{1/n} \) is increasing. [See Rudin for more properties of \( x^{1/n} \) and \( x^{m/n} \) where \( m \in \mathbb{Z} \) and \( n \in \mathbb{N} \).

**Proof. Uniqueness.** First of \( t > s \geq 0 \) then \( t^n > s^n \geq 0 \) as can be proved by induction.\(^4\) Thus if \( x, y \geq 0 \) and \( x^n = y^n \) then \( x = y \) for otherwise \( x > y \) or

\[^3\text{For what it is worth, as dictionary definition of irrational is "not consistent with or using reason." Let's try to use irrational numbers in a rational way!}\]

\[^4\text{The statement holds for } n = 1 \text{ by assumption and if } t^n > s^n \text{, then } t^{n+1} > ts^n > s^{n+1}. \text{ For the last equality we used } t > s \text{ implies } ts^n > s^n.\]
In which case \(x^n > y^n\) or \(y^n > x^n\) respectively. This shows that there is at most one \(n^{th}\) root if it exists. I also assert that \(x^{1/n} < y^{1/n}\) if \(x < y\). If not then \(x^{1/n} \geq y^{1/n}\) and this would then imply \(x = (x^{1/n})^n \geq (y^{1/n})^n = y\), which contradicts \(x < y\).

**Existence.** Let \(\Lambda := \{t \in \mathbb{R} : t^n \leq x\}\). If \(t = \frac{1}{1 + x} \in (0,1)\), then \(t^n \leq t \leq x\), so that \(t \in \Lambda\) and \(\Lambda \neq \emptyset\). If \(t = 1 + x\), then \(t^n = (1 + x)^n \geq 1 + nx > x\) and therefore \(\Lambda\) is bounded from above. Hence we may define \(y := \sup \Lambda\). We will now show that \(y^n = x\).

By Theorem 3.14 there exists \(t_k \in \Lambda\) such that \(t_k \uparrow y\) as \(k \to \infty\). By definition of \(\Lambda\), \(t_k^n \leq x\) for all \(k\). Passing to the limit as \(k \to \infty\) in this inequality implies \(y^n = \lim_{k \to \infty} t_k^n \leq x\).

If \(y^n < x\) then (using the Binomial theorem) and properties of limits,

\[
\left( y + \frac{1}{m} \right)^n = y^n + \sum_{k=1}^{n} \binom{n}{k} \left( \frac{1}{m} \right)^k \to y^n < x \text{ as } m \to \infty.
\]

Hence for sufficiently large \(m\) we will have \((y + \frac{1}{m})^n < x\). But this shows that \(y + \frac{1}{m} \in \Lambda\) which violates \(y\) being an upper bound for \(\Lambda\). Therefore we conclude that \(y^n = x\).

**End of Lecture 6, 10/10/2012.**

### 3.1 Extended real numbers

**Notation 3.21** 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 \(-\infty \) and \(+\infty\). We use the following conventions, \(\pm \infty \cdot a = \pm \infty\) if \(a \in \mathbb{R}\) with \(a > 0\), \(\pm \infty \cdot a = -\infty\) if \(a \in \mathbb{R}\) with \(a < 0\), \(\pm \infty + a = \pm \infty\) for any \(a \in \mathbb{R}\), \(\infty + \infty = \infty\) and \(-\infty - \infty = -\infty\) while the following expressions are not defined:

\[-\infty - \infty, -\infty + \infty, \infty / \infty, 0 \cdot \infty, \text{ and } \infty \cdot 0.\]

A sequence \(a_n \in \bar{\mathbb{R}}\) is said to converge to \(\infty\) (\(-\infty\)) 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 \(n \geq m\). In these cases we write \(\lim_{n \to \infty} a_n = \pm \infty\) or \(a_n \to \pm \infty\) as \(n \to \infty\).

For any subset \(\Lambda \subset \bar{\mathbb{R}}\), let \(\sup \Lambda\) and \(\inf \Lambda\) denote the least upper bound and greatest lower bound of \(\Lambda\) respectively. The convention being that \(\sup \Lambda = -\infty\) if \(\infty \in \Lambda\) or \(\Lambda\) is not bounded from above and \(\inf \Lambda = -\infty\) if \(-\infty \in \Lambda\) or \(\Lambda\) is not bounded from below. We will also use the conventions that \(\sup \emptyset = -\infty\) and \(\inf \emptyset = +\infty\). The next theorem is a fairly simple but often useful result about computing least upper bounds.

**Theorem 3.22 (Sup Sup Theorem).** Suppose that \(\Lambda\) is a subset of \(\mathbb{R}\) such that \(\Lambda = \bigcup_{\alpha \in I} \Lambda_\alpha\) where \(\Lambda_\alpha \subset \mathbb{R}\) and \(I\) is some index set. Then

\[
\sup \Lambda = \sup_{\alpha \in I} \Lambda_\alpha.
\]

The convention here is that the supremum of a set which is not bounded from above is \(\infty\) and the sup \(\emptyset = -\infty\).

**Proof.** Let \(M := \sup \Lambda\) and \(M_\alpha := \sup \Lambda_\alpha\) for all \(\alpha \in I\). As \(\Lambda_\alpha \subset \mathbb{R}\) we have \(M_\alpha \leq M\) for all \(\alpha \in I\) and therefore \(\sup_{\alpha \in I} M_\alpha \leq M\). Conversely, if \(\Lambda \subset \mathbb{R}\), then \(\lambda \in \Lambda_\alpha\) for some \(\alpha \in I\) and therefore \(\lambda \leq M_\alpha\). From this it follows that \(\lambda \leq \sup_{\alpha \in I} M_\alpha\) and as \(\lambda \in \Lambda\) is arbitrary we may conclude that \(M = \sup \Lambda = \sup_{\alpha \in I} M_\alpha\).

The next corollary records a typical way the Sup Sup theorem is used.

**Corollary 3.23.** Suppose that \(X\) and \(Y\) are sets and \(S : X \times Y \to \mathbb{R}\) is a function. Then

\[
\sup_{x \in X, y \in Y} S(x, y) = \sup_{x \in X} \sup_{y \in Y} S(x, y) = \sup_{y \in Y} \sup_{x \in X} S(x, y).
\]

**Proof.** Let \(X := \{x \in X \times Y\}\), and for \(x \in X\) let \(A_x := \{y \in Y\}\). Then \(A = \bigcup_{x \in X} A_x\) and therefore,

\[
\sup_{(x, y) \in X \times Y} S(x, y) = \sup_{x \in X} \sup_{y \in Y} S(x, y).
\]

The same reasoning also shows,

\[
\sup_{(x, y) \in X \times Y} S(x, y) = \sup_{y \in Y} \sup_{x \in X} S(x, y).
\]

The next Lemma records some basic limit theorems involving the extended real numbers.

**Lemma 3.24.** 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 \(c \in \mathbb{R}\), \(\lim_{n \to \infty} (ca_n) = c \lim_{n \to \infty} a_n\).

[5] The only sequences that do not converge in \(\mathbb{R}\) are those which oscillate too much.
3. If \( \{a_n + b_n\} \) is convergent and

\[
\lim_{n \to \infty} (a_n + b_n) = \lim_{n \to \infty} a_n + \lim_{n \to \infty} b_n
\]

provided the right side is not of the form \( \infty - \infty \).

4. \( \{a_n b_n\} \) is convergent and

\[
\lim_{n \to \infty} (a_n b_n) = \lim_{n \to \infty} a_n \cdot \lim_{n \to \infty} b_n
\]

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

Before going to the proof consider the simple example where \( a_n = n \) and \( b_n = -\alpha n + c \) with \( \alpha > 0 \) and \( c \in \mathbb{R} \). Then\(^6\)

\[
\lim (a_n + b_n) = \begin{cases} 
\infty & \text{if } \alpha < 1 \\
c & \text{if } \alpha = 1 \\
 \infty & \text{if } \alpha > 1 
\end{cases}
\]

while

\[
\lim a_n + \lim b_n = \infty - \infty.
\]

This shows that the requirement that the right side of Eq. 3.1 is not of form \( \infty - \infty \) is necessary in Lemma 3.24. 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. 3.2 is not of form \( \infty \cdot 0 \).

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

**Proof of Eq. 3.2.** 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 \( \lim_{n \to \infty} b_n = \infty \). Let us also suppose that \( a > 0 \) (the case \( a < 0 \) is handled similarly) and let \( \alpha := \min \left( \frac{2}{3}, 1 \right) \). 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.

**Exercise 3.2.** Show \( \lim_{n \to \infty} a^n = \infty \) and \( \lim_{n \to \infty} \frac{1}{a^n} = 0 \) whenever \( \alpha > 1 \).

**Exercise 3.3.** Suppose \( \alpha > 1 \) and \( k \in \mathbb{N} \), show there is a constant \( c = c(\alpha, k) > 0 \) such that \( a^n \geq c n^k \) for all \( n \in \mathbb{N} \). [In words, \( \alpha^n \) grows in \( n \) faster than any polynomial in \( n \].

**Lemma 3.25.** Suppose that \( \{a_n\} \subset \mathbb{R} \) and \( \lim_{n \to \infty} a_n = A \in \mathbb{R} \). Then every subsequence, \( \{b_k := a_{n_k}\}_{k=1}^\infty \) also converges to \( A \).

**Exercise 3.4.** Prove Lemma 3.25.

### 3.2 Limsups and Liminfs

**Notation 3.26.** Suppose that \( \{x_n\} \subset \mathbb{R} \) is a sequence of numbers. Then

\[
\begin{align*}
\liminf_{n \to \infty} x_n &= \liminf_{n \to \infty} \{x_k : k \geq n\} \quad \text{and} \\
\limsup_{n \to \infty} x_n &= \limsup_{n \to \infty} \{x_k : k \geq n\}.
\end{align*}
\]

We will also write \( \liminf \) for \( \lim\inf \) and \( \limsup \) for \( \lim\sup \).

**Remark 3.27.** 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. (3.3) and Eq. (3.4) always exist in \( \mathbb{R} \) (see Theorem 3.15) and

\[
\begin{align*}
\liminf_{n \to \infty} x_n &= \sup\{x_k : k \geq n\} \\
\limsup_{n \to \infty} x_n &= \inf\{x_k : k \geq n\}.
\end{align*}
\]

Owing to the following exercise, one may reduce properties of the \( \liminf \) to those of the \( \limsup \).

**Exercise 3.5.** Show \( \liminf_{n \to \infty} (-a_n) = -\limsup_{n \to \infty} a_n \).
Exercise 3.6. Let \( \{a_n\}_{n=1}^{\infty} \) be the sequence given by,
\[
(-1, 2, 3, -1, 2, 3, -1, 2, 3, \ldots).
\]
Find \( \limsup_{n \to \infty} a_n \) and \( \liminf_{n \to \infty} a_n \).

Exercise 3.7. If \( \{a_n\}_{n=1}^{\infty} \) and \( \{b_n\}_{n=1}^{\infty} \) are two sequences such that \( a_n \leq b_n \) for a.a. \( n \), then
\[
\limsup_{n \to \infty} a_n \leq \limsup_{n \to \infty} b_n \quad \text{and} \quad \liminf_{n \to \infty} a_n \leq \liminf_{n \to \infty} b_n. \tag{3.5}
\]
The following proposition contains some basic properties of liminfs and limsups.

Proposition 3.28. Let \( \{a_n\}_{n=1}^{\infty} \) and \( \{b_n\}_{n=1}^{\infty} \) be two sequences of real numbers. Then
1. \( \liminf_{n \to \infty} a_n \leq \limsup_{n \to \infty} a_n \).
2. \( \lim_{n \to \infty} a_n \) exists in \( \mathbb{R} \) iff
\[
\liminf_{n \to \infty} a_n = \limsup_{n \to \infty} a_n \in \mathbb{R}.
\]
3. \( \limsup_{n \to \infty} (a_n + b_n) \leq \limsup_{n \to \infty} a_n + \limsup_{n \to \infty} b_n \) \tag{3.6}

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
\[
\limsup_{n \to \infty} (a_nb_n) \leq \limsup_{n \to \infty} a_n \cdot \limsup_{n \to \infty} b_n, \tag{3.7}
\]
provided the right hand side of (3.7) is not of the form \( 0 \cdot \infty \) or \( \infty \cdot 0 \).

Proof. Items 1. and 2. will be proved here leaving the remaining items as an exercise to the reader. For item 1. we have
\[
\inf\{a_k : k \geq n\} \leq \sup\{a_k : k \geq n\} \quad \forall n,
\]
and therefore by the Sandwich theorem, \( \liminf_{n \to \infty} a_n \leq \limsup_{n \to \infty} a_n \).

2. (\( \Leftarrow \)) Let \( A := \liminf_{n \to \infty} a_n = \limsup_{n \to \infty} a_n \in \mathbb{R} \). Since
\[
\inf_{k \geq n} a_k \leq a_n \leq \sup_{k \geq n} a_k,
\]
if \( A \in \mathbb{R} \) then it follows by the sandwich theorem that \( \lim_{n \to \infty} a_n = A \). If \( A = \infty \), then for all \( M \in \mathbb{N} \) we have \( M \leq \inf_{k \geq n} a_k \) for a.a. \( n \). Therefore \( a_k \geq M \) for a.a. \( k \) and we have shown \( \lim_{k \to \infty} a_k = \infty \). If \( A = -\infty \) then for all \( M \in \mathbb{N} \) we have \( \sup_{k \geq n} a_k \leq -M \) for a.a. \( n \). Therefore \( a_k \leq -M \) for a.a. \( k \) and we have shown \( \lim_{k \to \infty} a_k = -\infty \).

(\( \Rightarrow \)) Conversely, suppose that \( \lim_{n \to \infty} a_n = A \in \mathbb{R} \) exists. If \( A \in \mathbb{R} \), then for every \( \varepsilon > 0 \) there exists \( N(\varepsilon) \in \mathbb{N} \) such that \( |A - a_n| \leq \varepsilon \) for all \( n \geq N(\varepsilon) \), i.e.
\[
A - \varepsilon \leq a_n \leq A + \varepsilon \quad \text{for all} \quad n \geq N(\varepsilon).
\]
From this we learn that
\[
A - \varepsilon \leq \inf_{k \geq n} a_k \leq \sup_{k \geq n} a_k \leq A + \varepsilon \quad \text{for a.a.} \quad n
\]
and so passing to the limit as \( n \to \infty \) implies
\[
A - \varepsilon \leq \liminf_{n \to \infty} a_n \leq \limsup_{n \to \infty} a_n \leq A + \varepsilon.
\]
Since \( \varepsilon > 0 \) is arbitrary, it follows that
\[
A \leq \liminf_{n \to \infty} a_n \leq \limsup_{n \to \infty} a_n \leq A,
\]
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 \( N = N(M) \) such that \( a_n \geq M \) for all \( n \geq N \). This show that \( \liminf_{n \to \infty} a_n \geq M \) and since \( M \) is arbitrary it follows that
\[
\liminf_{n \to \infty} a_n \leq \limsup_{n \to \infty} a_n.
\]
The proof for the case \( A = -\infty \) is analogous to the \( A = \infty \) case.

Exercise 3.8. Show that
\[
\limsup_{n \to \infty}(a_n + b_n) \leq \limsup_{n \to \infty} a_n + \limsup_{n \to \infty} b_n,
\]
provided that the right side of Eq. (3.8) 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 3.9. Suppose that \( a_n \geq 0 \) and \( b_n \geq 0 \) for all \( n \in \mathbb{N} \). Show
\[
\limsup_{n \to \infty}(a_nb_n) \leq \limsup_{n \to \infty} a_n \cdot \limsup_{n \to \infty} b_n,
\]
provided the right hand side of (3.9) is not of the form \( 0 \cdot \infty \) or \( \infty \cdot 0 \).

Exercise 3.10. Suppose that \( \{a_n\}_{n=1}^{\infty} \) and \( \{b_n\}_{n=1}^{\infty} \) are two non-negative sequences and assume \( A = \lim_{n \to \infty} a_n \) exists in \( (0, \infty) \). Show
\[
\limsup_{n \to \infty} a_nb_n = A \cdot \limsup_{n \to \infty} b_n.
\]
Definition 3.29. A sequence, \( \{a_n\}_{n=1}^{\infty} \), of positive numbers is said to have \textit{sub-geometric growth} iff for all \( \alpha > 1 \) there exists \( c = c(\alpha) < \infty \) such that
\[
a_n \leq c a^n \text{ for a.a. } n \in \mathbb{N}.
\] (3.10)

Lemma 3.30. Suppose that \( \{a_n\}_{n=1}^{\infty} \) is a sequence of non-negative numbers.

1. If \( \{a_n\}_{n=1}^{\infty} \) has sub-geometric growth, then \( \limsup_{n \to \infty} a_n^{1/n} \leq 1 \).
2. If \( a_n > 0 \) for a.a. \( n \) and \( \{a_n\}_{n=1}^{\infty} \) has sub-geometric growth (i.e. for all \( \alpha > 1 \) there exists \( \varepsilon = \varepsilon(\alpha) < \infty \) such that \( a_n \geq \varepsilon a^{-n} \) for a.a. \( n \)), then \( \liminf_{n \to \infty} a_n^{1/n} \geq 1 \).
3. In particular if \( a_n > 0 \) for a.a. \( n \) and both \( \{a_n\}_{n=1}^{\infty} \) and \( \{a_n^{-1}\}_{n=1}^{\infty} \) have sub-geometric growth, then \( \lim_{n \to \infty} a_n^{1/n} = 1 \).

Proof. 1. For any \( \beta > 1 \) choose \( \alpha \in (1, \beta) \) and let \( c = c(\alpha) \). Since \( \lim_{n \to \infty} c \left( \frac{\alpha}{\beta} \right)^n = 0 < 1 \), it follows that \( a_n \leq c a^n = c \left( \frac{\alpha}{\beta} \right)^n \beta^n \leq \beta^n \) for a.a. \( n \) and therefore
\[
a_n^{1/n} \leq \beta \text{ for a.a. } n \implies \limsup_{n \to \infty} a_n^{1/n} \leq \beta.
\]
As \( \beta > 1 \) is arbitrary we may conclude that \( \limsup_{n \to \infty} a_n^{1/n} \leq 1 \).

2. This may be proved similarly to item 1, or it can be reduced to item 1 as we show here. By item 1. applies to \( \{a_n^{-1}\} \),
\[
\limsup_{n \to \infty} \frac{1}{a_n^{1/n}} = \limsup_{n \to \infty} a_n^{-1/n} \leq 1.
\]
The proof is completed because \( \limsup_{n \to \infty} \frac{1}{a_n^{1/n}} = \frac{1}{\liminf_{n \to \infty} a_n^{1/n}} \) as the reader should prove.

3. From items 1. and 2. we learn that
\[
1 \leq \liminf_{n \to \infty} a_n^{1/n} \leq \limsup_{n \to \infty} a_n^{1/n} \leq 1
\]
from which the result follows.

Example 3.31. If \( c^{-1/n-p} \leq a_n \leq cn^p \) for some \( p \in \mathbb{N} \) and \( c < \infty \), then \( \lim_{n \to \infty} a_n^{1/n} = 1 \). Indeed using Exercise 3.3 we may easily verify the hypothesis of Lemma 3.30.

Exercise 3.11. Suppose that \( p(t) \) is a non-zero polynomial of \( t \in \mathbb{R} \) with (possibly) complex coefficients. Show
\[
\lim_{n \to \infty} |p(n)|^{1/n} = 1.
\]

• End of Lecture 7, 10/12/2012.

Exercise 3.12. If \( a_n \geq 0 \), then \( \lim_{n \to \infty} a_n = 0 \) iff \( \limsup_{n \to \infty} a_n = 0 \).

Proposition 3.32. Suppose that \( \{a_n\}_{n=1}^{\infty} \) is a sequence of real numbers and let
\[
B := \{ y \in \mathbb{R} : a_n \geq y \text{ for i.o. } n \}.
\]
Then \( \sup B = \limsup_{n \to \infty} a_n \) with the convention that \( \sup B = -\infty \) if \( B = \emptyset \).

Proof. If \( \{a_n\}_{n=1}^{\infty} \) is not bounded from above, then \( B \) is not bounded from above and \( \sup B = \infty = \limsup_{n \to \infty} a_n \). If \( B = \emptyset \) so that \( \sup B = -\infty \), then for all \( y \in \mathbb{R} \) we must have \( a_n \leq y \) for a.a. \( n \). This then implies \( \limsup_{n \to \infty} a_n \leq y \) for all \( y \in \mathbb{R} \) from which we conclude that \( \limsup_{n \to \infty} a_n = -\infty \). So let us now assume that \( B \neq \emptyset \) and \( \{a_n\}_{n=1}^{\infty} \) is bounded in which case \( B \) is bounded from above. Let us set \( \beta := \sup B \in \mathbb{R} \) and \( \alpha := \limsup_{n \to \infty} a_n \).

If \( y > \beta \), then \( a_n < y \) for a.a. \( n \) from which it follows that \( a^n := \limsup_{n \to \infty} a_n \leq y \). We may now let \( y \downarrow \beta \) in order to see that \( a^n \leq \beta \).\(^7\) Now suppose that \( y < \beta \), then \( a_n \geq y \) for a.a. \( n \) and hence \( a^n \leq \limsup_{n \to \infty} a_n \geq y \). Letting \( y \uparrow \beta \) then shows \( a^n \geq \beta \). Thus we have shown \( a^n = \beta \).

Theorem 3.33. There is a subsequence \( \{a_{n_k}\}_{k=1}^{\infty} \) of \( \{a_n\}_{n=1}^{\infty} \) such that \( \lim_{k \to \infty} a_{n_k} = \limsup_{n \to \infty} a_n \). Similarly, there is a subsequence \( \{a_{n_k}\}_{k=1}^{\infty} \) of \( \{a_n\}_{n=1}^{\infty} \) such that \( \lim_{k \to \infty} a_{n_k} = \liminf_{n \to \infty} a_n \). Moreover, every convergent subsequence, \( \{b_k := a_{n_k}\}_{k=1}^{\infty} \) of \( \{a_n\}_{n=1}^{\infty} \) satisfies,
\[
\liminf_{n \to \infty} a_n \leq \liminf_{k \to \infty} b_k \leq \lim sup_{k \to \infty} b_k \leq \limsup_{n \to \infty} a_n.
\]

Proof. Let me prove the last assertion first. Suppose that \( b_k := a_{n_k} \) is some convergent subsequence of \( \{a_n\}_{n=1}^{\infty} \). Then we have,
\[
\inf_{n \geq n_k} a_n \leq b_k \leq \sup_{n \geq n_k} a_n \text{ for all } k \in \mathbb{N}.
\]

Passing to the limit in this equation then implies,
\[
\liminf_{n \to \infty} a_n = \liminf_{k \to \infty} a_n \leq \limsup_{k \to \infty} b_k \leq \limsup_{k \to \infty} b_k \leq \limsup_{n \to \infty} a_n = \limsup_{n \to \infty} a_n.
\]
We have used, \( \{\inf_{n \geq n_k} a_n\}_{k=1}^{\infty} \) and \( \{\sup_{n \geq n_k} a_n\}_{k=1}^{\infty} \) are subsequence of the convergent subsequences of \( \{\inf_{n \geq k} a_n\}_{k=1}^{\infty} \) and \( \{\sup_{n \geq k} a_n\}_{k=1}^{\infty} \) respectively and therefore converge to the same limits respectively, see Lemma 3.30.

Now let us prove the first assertions. I will cover the limsup case here as the liminf case is similar or can be deduced from the limsup case with the aid of Exercise 3.5. Let \( A := \limsup_{n \to \infty} a_n \). We will need to consider three cases, \( A \in \mathbb{R}, A = \infty, \text{ and } A = -\infty. \)

\(^7\) This can be done more formally by choosing a sequence \( \{y_k\}_{k=1}^{\infty} \) such that \( y_k \downarrow \alpha \) so that \( a^n \leq y_k \). Now let \( k \to \infty \) to conclude \( a^n \leq \liminf_{k \to \infty} y_k = \alpha. \)
iii) Finally suppose that $A \in \mathbb{R}$, then by Proposition 3.32 for all $k \in \mathbb{N}$ we have $A - \frac{1}{k} \leq a_n$ for infinitely many $n$. In particular we can choose $n_1 < n_2 < n_3 < \ldots$ inductively so that $A - \frac{1}{k} \leq a_{n_k}$ for all $k$. Since

$$A - \frac{1}{k} \leq a_{n_k} \leq \sup_{m \geq n_k} a_m,$$

and the limit as $k \to \infty$ of both extremes of this inequality are $A$, it follow from the sandwich inequality that $\lim_{k \to \infty} a_{n_k} = A$.

ii) If $A = \limsup_{n \to \infty} a_n = \infty$, then $\sup_{k \geq n} a_k = \infty$ for all $n \in \mathbb{N}$ which implies for all $M < \infty$ that $a_k \geq M$ for i.o. $k$. Working similarly to case i) we can choose $n_1 < n_2 < n_3 < \ldots$ so that $a_{n_k} \geq k$ for all $k$ and therefore $\lim_{k \to \infty} a_{n_k} = \infty$.

iii) Finally suppose that $A = \limsup_{n \to \infty} a_n = -\infty$ so that for all $M \in \mathbb{N}$ there exists $N \in \mathbb{N}$ such that $\sup_{k \geq n} a_k \leq M$ for all $n \geq N$, i.e. $a_n \leq -M$ for all $n \geq N$. In this case it follows that in fact $\lim_{n \to \infty} a_n = -\infty$ and we do not have to even choose as subsequence.

Corollary 3.34 (Bolzano–Weierstrass Property / Compactness). Every bounded sequence of real numbers, $\{a_n\}_{n=1}^\infty$, has a convergent in $\mathbb{R}$ subsequence, $\{a_{n_k}\}_{k=1}^\infty$. If we drop the bounded assumption then we may only assert that there is a subsequence which is convergent in $\mathbb{R}$.

**Proof.** Let $M < \infty$ such that $|a_n| \leq M$ for all $n \in \mathbb{N}$, i.e. $-M \leq a_n \leq M$ for all $n$. We may then conclude from Exercise 3.7 that,

$$-M \leq \limsup_{n \to \infty} a_n \leq M.$$ 

It now follows from Theorem 3.33 that there exists a subsequence, $\{a_{n_k}\}_{k=1}^\infty$, of $\{a_n\}_{n=1}^\infty$ such that

$$\lim_{k \to \infty} a_{n_k} = \limsup_{n \to \infty} a_n \in [-M, M] \subset \mathbb{R}.$$ 

Theorem 3.35 ($\mathbb{R}$ is Cauchy complete). If $\{a_n\}_{n=1}^\infty \subset \mathbb{R}$ is a Cauchy sequence, then $\lim_{n \to \infty} a_n$ exists in $\mathbb{R}$ and in fact,

$$\lim_{n \to \infty} a_n = \limsup_{n \to \infty} a_n = \liminf_{n \to \infty} a_n.$$ 

**Proof.** We will give two proofs of this important theorem. Each proof uses the fact that $\{a_n\}_{n=1}^\infty \subset \mathbb{R}$ is Cauchy implies $\{a_n\}_{n=1}^\infty$ is bounded. This is proved exactly in the same way as the solution to Exercise 1.9.

First proof. By Corollary 3.34 there is a subsequence, $\{a_{n_k}\}_{k=1}^\infty$, such that $\lim_{k \to \infty} a_{n_k} = L \in \mathbb{R}$. As in the proof of Exercise 1.7 it follows that $\lim_{n \to \infty} a_n$ exists and is equal to $L$.

Second proof. Let $a := \liminf_{n \to \infty} a_n$ and $b := \limsup_{n \to \infty} a_n$. It suffices to show $a = b$. As we always know that $a \leq b$ it will suffice to show $b \leq a$.

Given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$|a_m - a_n| \leq \varepsilon$$

for all $m, n \geq N$. In particular, for $m, n \geq k \geq N$ we have $a_m \leq a_n + \varepsilon$ and hence

$$b \leq \sup_{m \geq k} a_m \leq a_n + \varepsilon$$

for all $n \geq k$. From this inequality we may further conclude,

$$b \leq \inf_{n \geq k} a_n + \varepsilon \leq a + \varepsilon.$$ 

As $\varepsilon > 0$ is arbitrary, we have indeed shown $b \leq a$.

Exercise 3.13. Suppose that $\{a_n\}_{n=1}^\infty$ is a sequence of real numbers and let

$$A := \{y \in \mathbb{R} : a_n \geq y \text{ for a.a. } n\}.$$ 

Then $\sup A = \liminf_{n \to \infty} a_n$ with the convention that $\sup A = -\infty$ if $A = \emptyset$.

Exercise 3.14. Suppose that $\{a_n\}_{n=1}^\infty$ is a sequence of real numbers. Show $\limsup_{n \to \infty} a_n = a^* \in \mathbb{R}$ iff for all $\varepsilon > 0$,

$$a_n \leq a^* + \varepsilon \text{ for a.a. } n \text{ and } a^* - \varepsilon \leq a_n \text{ i.o. } n.$$ 

Similarly, show $\liminf_{n \to \infty} a_n = a_* \in \mathbb{R}$ iff for all $\varepsilon > 0$,

$$a_n \leq a_* + \varepsilon \text{ i.o. } n \text{ and } a_* - \varepsilon \leq a_n \text{ for a.a. } n.$$ 

Notice that this exercise gives another proof of item 2. of Proposition 3.28 in the case all limits are real valued.

Exercise 3.15 (Cauchy Complete $\implies$ L.U.B. Property). Suppose that $\mathbb{R}$ denotes any ordered field which is Cauchy complete. Show $\mathbb{R}$ has the least upper bound property and therefore is the field of real numbers.
3.3 Partitioning the Real Numbers

Notation 3.36 (Intervals) For \( a, b \in \mathbb{R} \) with \( a < b \) we define,

\[
(a, b) := \{ x \in \mathbb{R} : a < x < b \},
\]

\[
[a, b] := \{ x \in \mathbb{R} : a \leq x \leq b \},
\]

\[
(a, b) := \{ x \in \mathbb{R} : a < x \leq b \}, \quad \text{and}
\]

\[
[a, b) := \{ x \in \mathbb{R} : a \leq x < b \}.
\]

We also also a \( -\infty \) in the intervals, \((a, b)\) and \([a, b)\] and allows \( b = +\infty \) in the intervals \((a, b)\) and \([a, b)\).

Notation 3.37 (Pairwise disjoint unions) If \( X \) is a set and \( A_\alpha \subset X \) for \( \alpha \in I \), we write \( X = \sum_{\alpha \in I} A_\alpha \) to mean; \( X = \bigcup_{\alpha \in I} A_\alpha \) and \( A_\alpha \cap A_\beta \) for all \( \alpha \neq \beta \).

Exercise 3.16. Suppose that \( a, b, c, d \in \mathbb{R} \) such that \( a < b \leq c < d \). Show \((a, b) \cap (c, d) = \emptyset\) and \([a, b) \cap [c, d) = \emptyset\).

Lemma 3.38 (Well Ordering II). Suppose that \( S \) is a non-empty subset of \( \mathbb{Z} \) which is bounded from below, then \( \inf(S) \in S \), i.e. \( S \) has a (unique) minimizer.

Proof. As \( S \) is bounded from below, there exists \( k \in \mathbb{Z} \) such that \( k \leq s \) for all \( s \in S \). Therefore \( \hat{S} := \{ s - k + 1 : s \in S \} \subset \mathbb{N} \) and hence by the Well ordering principle, \( \min(\hat{S}) := m \in \mathbb{N} \) exists. That is \( m \leq s - k + 1 \) for all \( s \in S \) and there exists \( s_0 \in S \) such that \( m = s_0 - k + 1 \). These last statements are equivalent to saying,

\[
s_0 = m + k - 1 \leq s \quad \text{for all} \quad s \in S,
\]

which is to say \( s_0 = \min(S) \). ■

Proposition 3.39. Suppose that \( \{S_n\}_{n=-\infty}^\infty \subset \mathbb{R} \) such that \( S_n < S_{n+1} \) for all \( n \in \mathbb{Z} \), \( \lim_{n \to \infty} S_n = \infty \) and \( \lim_{n \to -\infty} S_n = -\infty \). Then

\[
\sum_{n \in \mathbb{Z}} (S_{n-1}, S_n] = \mathbb{R} = \sum_{n \in \mathbb{Z}} [S_n, S_{n+1}). \tag{3.11}
\]

Proof. The fact that \((S_n, S_{n+1}) = (S_m, S_{m+1}) = \emptyset \) follows from Exercise 3.16. For \( x \in \mathbb{R} \), let

\[
n_0 := \min(\{n \in \mathbb{Z} : x \leq S_n\})
\]

which exists since \( \{n \in \mathbb{Z} : x \leq S_n\} \) is non-empty as \( S_n \to \infty \) as \( n \to \infty \) and is bounded from below since \( S_n \to -\infty \) as \( n \to -\infty \). It then follows that \( x \leq S_{n_0} \) while \( x \not\leq S_{n_0-1} \), i.e. \( S_{n_0-1} < x \leq S_{n_0} \) and we have shown \( x \in (S_{n_0-1}, S_{n_0}] \) which completes the proof of the first equality in Eq. (3.11). The proof of the second equality is similar and so will be omitted. ■

Proposition 3.40. Suppose that \(-\infty < a < b < \infty \) and \( \{S_n\}_{n=0}^N \subset [a, b] \) such that \( a = S_0 < S_1 < \cdots < S_{N-1} < S_N = b \), then

\[
[a, b) = \sum_{n=1}^N [S_{n-1}, S_n).
\]

This result also holds if \( N = \infty \) provided we now assume \( S_n < S_{n+1} \) for all \( n \), \( a = S_0 \), and \( S_n \uparrow b \) as \( n \to \infty \).

Proof. This proof is very similar to the proof of Proposition 3.39 and so will be omitted. ■

3.4 The Decimal Representation of a Real Number

Lemma 3.41 (Geometric Series). Let \( \alpha \in \mathbb{R} \) or \( \alpha \in \mathbb{Q} \), \( m, n \in \mathbb{Z} \) and \( S := \sum_{k=n}^m \alpha^k \). Then

\[
S = \begin{cases} 
\frac{m-n+1}{\alpha - 1} & \text{if } \alpha = 1 \\
\frac{\alpha^{m+1} - \alpha^n}{\alpha - 1} & \text{if } \alpha \neq 1.
\end{cases}
\]

Proof. When \( \alpha = 1 \),

\[
S = \sum_{k=n}^m 1^k = m - n + 1.
\]

If \( \alpha \neq 1 \), then \( \alpha S = \alpha^{m+1} - \alpha^n \).

Solving for \( S \) gives

\[
S = \sum_{k=n}^m \alpha^k = \frac{\alpha^{m+1} - \alpha^n}{\alpha - 1} \quad \text{if } \alpha \neq 1. \tag{3.12}
\]

Taking \( \alpha = 10^{-1} \) in Eq. (3.12) implies

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

and in particular, for all \( M \geq n \),

\[
\lim_{m \to \infty} \sum_{k=n}^m 10^{-k} = \frac{1}{9} \cdot 10^{n-1} \geq \sum_{k=n}^M 10^{-k}.
\]
Definition 3.42 (Decimal Numbers). Let $\mathbb{D}$ denote those sequences $\alpha \in \{0, 1, 2, \ldots, 9\}^\infty$ 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 $n \in \mathbb{Z}$.

A decimal number is then an expression of the form

$$\alpha = \alpha_{-N}\alpha_{-N+1} \ldots \alpha_0.\alpha_1\alpha_2\ldots$$

For example

$$52 + \sqrt{2} \approx 53.41421356237309504880168872420969807856967187537694807\ldots$$

To every decimal number $\alpha \in \mathbb{D}$ is the sequence $a_n = a_n(\alpha)$ defined for $n \in \mathbb{N}$ by

$$a_n := \sum_{k=-\infty}^{n} \alpha_k 10^{-k}.$$ (a finite sum).

Since for $m > n$,

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

it follows that

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

which shows $\{a_n\}_{n=1}^{\infty}$ is a Cauchy sequence. Thus to every decimal number we may associate the real number

$$a(\alpha) := \lim_{n \to \infty} a_n.$$

Theorem 3.43. If $x \geq 0$ is a real number, there exists $\alpha \in \mathbb{D}$ such that $x = a(\alpha)$, i.e. all real numbers can be represented in decimal form.

Proof. If $x = 0$, we can take $\alpha_n = 0$ for all $n$ so that $0 = a(\alpha)$. So suppose that $x > 0$ and let $p := \min\{\{n \in \mathbb{N} : x < n\}\}$. Set $m = p - 1$, then $m \leq x < m + 1$. We then define $\alpha_k$ for $k \leq 0$ so that $m = \alpha_{-N} \ldots \alpha_0$. We now construct $\alpha_k$ for $k \geq 1$. For $k = 1$ we write

$$[m, m + 1) = \sum_{l=0}^{9} \left[ m + \frac{l}{10}, m + \frac{l+1}{10} \right)$$

and then choose $\alpha_1 = l$ if $x \in [m + \frac{l}{10}, m + \frac{l+1}{10})$. We then construct $\alpha_2$ using,

$$[m + \alpha_1 10, m + \frac{\alpha_1 + 1}{10}] \sum_{l=0}^{9} \left[ m + \frac{\alpha_1 + l}{10}, m + \frac{\alpha_1 + l + 1}{10} \right)$$

and set $\alpha_2 = l$ for $x \in [m + \frac{\alpha_1 10 + l}{10}, m + \frac{\alpha_1 + l + 1}{10})$. Continuing this way inductively we construct $\{\alpha_k\}_{k=1}^{\infty}$ such that

$$x \in [m + \sum_{j=1}^{k} \frac{\alpha_j}{10^j}, m + \sum_{j=1}^{k-1} \frac{\alpha_j}{10^j} + \frac{\alpha_k + 1}{10^k}).$$

It is now easy to see that $x = a(\alpha)$.

Remark 3.44. The representation of $x \geq 0$ as a decimal number may not be unique. For example,

$$0.999\ldots = \sum_{k=1}^{\infty} \frac{9}{10^k} = \frac{9}{10} \cdot \frac{1}{1 - \frac{1}{10}} = 1.000.$$

[Or note that

$$1 - 0.9\cdot 9 = 0.0\cdot 0\cdot 1 = 10^{-n} \rightarrow 0 \text{ as } n \rightarrow \infty.]$$

On the other hand if we agree to not allow a tail of repeated 9’s as an element of $\mathbb{D}$, then the representation would be unique.

3.5 Summary of Key Facts about Real Numbers

1. The real numbers, $\mathbb{R}$, is the unique (up to order preserving field isomorphism) ordered field with the least upper bound property or equivalently which is Cauchy complete.
2. Informally the real numbers are the rational numbers with the (irrational) hole filled in.
3. Monotone bounded sequence always converge in $\mathbb{R}$.
4. A sequence converges in $\mathbb{R}$ iff it is Cauchy.
5. Cauchy sequences are bounded.
6. $\mathbb{N}$ is unbounded from above in $\mathbb{R}$.
7. For all $\varepsilon > 0$ in $\mathbb{R}$ there exists $n \in \mathbb{N}$ such that $\frac{1}{n} < \varepsilon$.
8. $\mathbb{Q}$ and $\mathbb{R} \setminus \mathbb{Q}$ is dense in $\mathbb{R}$. In particular, between any two real numbers $a < b$, there are infinitely many rational and irrational numbers.
9. Decimal numbers map (almost 1-1) into the real numbers by taking the limit of the truncated decimal number.
10. If $a, b, \varepsilon \in \mathbb{R}$, then
a) $a \leq b$ by showing that $a \leq b + \varepsilon$ for all $\varepsilon > 0$.
b) $a = b$ by proving $a \leq b$ and $b \leq a$ or
c) $a = b$ by showing $|b - a| \leq \varepsilon$ for all $\varepsilon > 0$.

11. A number of standard limit theorems hold, see Theorem 3.13.

12. Unlike limits, lim sup and lim inf always exist. Moreover we have:
\[ \liminf_{n \to \infty} a_n \leq \limsup_{n \to \infty} a_n \]
with equality iff $\lim_{n \to \infty} a_n$ exists in which case
\[ \liminf_{n \to \infty} a_n = \lim_{n \to \infty} a_n = \limsup_{n \to \infty} a_n. \]

We may allow the values of $\pm \infty$ in these statements.

13. If $b_k := \{a_{n_k}\}_{k=1}^\infty$ is a convergent subsequence of $\{a_n\}$, then
\[ \liminf_{n \to \infty} a_n \leq \lim_{k \to \infty} b_k \leq \limsup_{n \to \infty} a_n \]
and we may choose $\{b_k\}$ so that $\lim_{k \to \infty} b_k = \limsup_{n \to \infty} a_n$ or
$\lim_{k \to \infty} b_k = \liminf_{n \to \infty} a_n$.

14. Bounded sequences of real numbers always have convergence subsequences.

15. If $S \subseteq \mathbb{R}$ and $A := \sup \{S\}$, then there exists $\{a_n\}_{n=1}^\infty \subseteq S$ such that
$\forall n, a_n \leq a_{n+1}$ for all $n$ and $\lim_{n \to \infty} a_n = \sup \{S\}$.

16. If $S \subseteq \mathbb{R}$ and $A := \inf \{S\}$, then there exists $\{a_n\}_{n=1}^\infty \subseteq S$ such that
$a_{n+1} \leq a_n$ for all $n$ and $\lim_{n \to \infty} a_n = \inf \{S\}$.

### 3.6 (Optional) Proofs of Theorem 3.6 and Theorem 3.3

In this section, we assume that $\mathbb{R}$ is as described in Theorem 3.6. The next exercise is relatively straightforward.

**Exercise 3.17.** Prove the following properties of $\mathbb{R}$.

1. Show addition and multiplication in Theorem 3.6 are well defined.
2. Show ($\mathbb{R}, +, \cdot$) satisfies the axioms of a field. **Hint:** for constructing multiplicative inverses, make use of Proposition 3.45 below to conclude if $\alpha := \{(a_n)_{n=1}^\infty\} \in \mathbb{R}$ and $a \neq 0 = i(0)$, then there exists $N \in \mathbb{N}$ such that $|a_n| \geq \frac{1}{N}$ for all $a_n$. By redefining the first few terms of $a_n$ if necessary, you may assume that $|a_n| \geq \frac{1}{N}$ for all $n$ and then take $\alpha^{-1} := \{(a_{n-1})_{n=1}^\infty\}$.
3. Show $i : \mathbb{Q} \to \mathbb{R}$ is injective homomorphism of fields.

To finish the proof of Theorem 3.6 we must show that $P$ is an ordering on $\mathbb{R}$ with the least upper bound property. This will be carried out in the remainder of this section.

**Proposition 3.45.** Suppose that $\alpha := \{(a_n)_{n=1}^\infty\}$ and $\beta := \{(b_n)_{n=1}^\infty\}$ are real numbers. Then precisely one of the following three cases can happen:

1. $\lim_{n \to \infty} (a_n - b_n) = 0$, i.e. $\alpha = \beta$,
2. there exists $\varepsilon = \frac{1}{N} > 0$ such that $a_n \geq b_n + \varepsilon$ for a.a. $n$ in which case $\alpha > \beta$, or
3. there exists $\varepsilon = \frac{1}{N} > 0$ such that $b_n \geq a_n + \varepsilon$ for a.a. $n$ in which case $\beta > \alpha$.

**Proof.** If case 1. does not hold then there exists $\delta > 0$ such that $|a_n - b_n| \geq \delta$ for infinitely many $n$. There are now two possibilities (which will turn out to me mutually exclusive):

- i) $a_n - b_n \leq \delta$ i.o. $n$,
- ii) $b_n - a_n \leq \delta$ i.o. $n$.

Since $\{a_n\}$ and $\{b_n\}$ are Cauchy sequences, there exists $N \in \mathbb{N}$ such that
\[ |a_n - a_m| \geq \delta/3 \] and $|b_n - b_m| \geq \delta/3$ for all $m, n \geq N$.

If case i) holds, we may choose an $m \geq N$ such that $a_m - b_m \geq \delta$ and so for $n \geq N$ we find,
\[ \delta \leq a_m - b_m = a_m - a_n + a_n - b_n + b_n - b_m \]
\[ \leq |a_m - a_n| + a_n - b_n + |b_n - b_m| \]
\[ = \delta/3 + a_n - b_n + \delta/3 \]
from which it follows that $a_n - b_n \geq \varepsilon := \delta/3$ for all $n \geq N$ and we are in case 2. Similarly if case ii) holds then we are in fact in case 3. of the proposition.

**Corollary 3.46.** Suppose that $\alpha := \{(a_n)_{n=1}^\infty\}$ and $\beta := \{(b_n)_{n=1}^\infty\}$ are real numbers, then $\alpha \geq \beta$ if and only if for all $N \in \mathbb{N}$,
\[ a_n - b_n \geq -\frac{1}{N} \text{ for a.a. } n. \tag{3.13} \]

Alternatively put, $\alpha \geq \beta$ if and only if for all $N \in \mathbb{N}$,
\[ b_n \leq a_n + \frac{1}{N} \text{ for a.a. } n. \]

**Proof.** If $\alpha = \beta$, then $\lim_{n \to \infty} (a_n - b_n) = 0$ and therefore Eq. (3.13) holds. If $\alpha > \beta$, then in fact $a_n - b_n \geq \varepsilon > 0 > -1/N$ for a.a. $n$.

Conversely, if $\alpha < \beta$, then there exists $\varepsilon > 0$ such that $b_n \geq a_n + \varepsilon$ for a.a. $n$. Thus if Eq. (3.13) were to also hold we could conclude for each $N \in \mathbb{N}$ that
\[ a_n \geq b_n - \frac{1}{N} \geq a_n + \varepsilon - \frac{1}{N} \text{ for a.a. } n. \]

This leads to a contradiction as soon as we choose $N$ so large as to make $1/N < \varepsilon$. Thus if Eq. (3.13) holds we must have $\alpha \geq \beta$.
Proposition 3.47. Suppose that $\lambda \in \mathbb{R}$, $\{a_n\}_{n=1}^{\infty}$ be a Cauchy sequence in $\mathbb{Q}$, and $\alpha := \{a_n\}_{n=1}^{\infty}$. If $\lambda \leq i(a_k)$ for all $k$ then $\lambda \leq \alpha$. Similarly if $i(a_k) \leq \lambda$ for all $k$ then $\alpha \leq \lambda$.

Proof. Let $\lambda = \{\lambda_n\}_{n=1}^{\infty}$ and suppose that $\lambda \leq i(a_n)$ for all $n$. For sake of contradiction, suppose that $\lambda > \alpha$, i.e. there exists an $N \in \mathbb{N}$ such that $\lambda_n \geq a_n + \frac{1}{N}$ for a.a. $n$. The assumption that $\lambda \leq i(a_k)$ implies that $\lambda_n \leq a_k + \frac{1}{N}$ for a.a. $n$. Because $\{a_k\}$ is Cauchy, we may conclude there exists $M \in \mathbb{N}$ such that

$$\lambda_n \leq a_k + \frac{1}{2N} \quad \text{for all } n, k \geq M.$$ 

By making $M$ even larger if necessary, we may assume that $\lambda_n \geq a_n + \frac{1}{N}$ for all $n \geq M$ as well. From these two inequalities with $k = n \geq M$ we learn

$$a_n + \frac{1}{N} \leq \lambda_n \leq a_n + \frac{1}{2N} \implies \frac{1}{2N} \geq \frac{1}{N},$$

and we have reached the desired contradiction. The fact that $i(a_k) \leq \lambda$ for all $k$ implies $\lambda \leq \alpha \leq \lambda$ is proved similarly. Alternatively if $i(a_k) \leq -\lambda$ then $-\lambda \leq i(-a_k)$ which implies $-\lambda \leq -\alpha$, i.e. $\alpha \leq \lambda$.

With these results in hand, let us now show that $\mathbb{R}$ as defined in Theorem 3.6 has the least upper bound property.

Proof of the least upper bound property. So suppose that $A \subset \mathbb{R}$ is a non empty set which is bounded from above. For each $m \in \mathbb{N}$, let $k_m \in \mathbb{Z}$ be the smallest integer such that $i(a_m) := i\left(\frac{a_m}{k_m}\right)$ is an upper bound for $A$. Since, for all $n \geq m$, $a_m - 2^{-m} \leq a_n \leq a_m$, we may conclude that

$$|a_n - a_m| \leq 2^{-\min(n,m)} \quad \text{as } n, m \to \infty.$$ 

This shows $\{a_n\}_{n=1}^{\infty}$ is Cauchy and hence we defined an element $\alpha := \{a_n\}_{n=1}^{\infty} \in \mathbb{R}$. We now will show $\alpha = \sup A$.

If $\lambda \in A$, then $\lambda \leq i(a_n)$ for all $n$ and so by Proposition 3.47 we conclude that $\lambda \leq \alpha$, i.e. $\alpha$ is an upper bound for $A$. Now suppose that $\beta$ is another upper bound for $A$. As $i(a_n - 2^{-n})$ is not an upper bound for $A$ there exists $\lambda \in A$ such that

$$i(a_n - 2^{-n}) < \lambda \leq \beta.$$ 

So by another application of Proposition 3.47 we learn that

$$\alpha = \{a_n\}_{n=1}^{\infty} = \{a_n - 2^{-n}\}_{n=1}^{\infty} \leq \beta.$$ 

This shows that $\alpha$ is in fact the least upper bound for $A$.

Theorem 3.48 (Real numbers are unique). Suppose that $F$ and $G$ are two complete ordered fields. Then there is a unique order preserving isomorphism, $\varphi : F \to G$.

(Sketch). Suppose that $\varphi : F \to G$ is an order preserving homomorphism. The usual arguments show that any homomorphism, $\varphi : F \to G$ must satisfy $\varphi(q1_F) = q1_G$. We know that $\{q \cdot 1_F : q \in \mathbb{Q}\}$ and $\{q \cdot 1_G : q \in \mathbb{Q}\}$ are dense copies of $\mathbb{Q}$ inside of $F$ and $G$ respectively. Now for general $a \in F$ choose $q_n, p_n \in \mathbb{Q}$ that $q_n1_F \uparrow a$ and $p_n1_F \downarrow a$. Since $\varphi$ is order preserving we must have $q_n1_G = \varphi(q_n1_F)$ is increasing and $p_n1_G = \varphi(p_n1_F)$ is decreasing. Moreover, since $p_n - q_n \to 0$ we must have $\lim_{n \to \infty} \varphi(q_n1_F) = \lim_{n \to \infty} \varphi(p_n1_F)$. Since $\varphi(q_n1_F) \leq \varphi(a) = \varphi(p_n1_F)$ for all $n$ it then follows that $\varphi(a) = \lim_{n \to \infty} q_n1_G = \lim_{n \to \infty} p_n1_G$ and we have shown $\varphi$ is uniquely determined.

For the converse, if $q_n \in \mathbb{Q}$ we know that

$$|q_n1_F - q_m1_F| = |q_n - q_m|1_F \quad \text{and}$$

$$|q_n1_G - q_m1_G| = |q_n - q_m|1_G.$$ 

Thus if $\{q_n1_F\}_{n=1}^{\infty}$ is convergent in $F$ iff $\{q_n1_G\}_{n=1}^{\infty}$ is convergent in $G$. Thus for any $a \in F$ we choose $q_n \in \mathbb{Q}$ such that $q_n1_F \to a$ and then define $\varphi(a) := \lim_{n \to \infty} q_n1_G$. One now checks that this formula is well defined (independent of the choice of $\{q_n\} \subset \mathbb{Q}$ such that $q_n1_F \to a$) and defines an order preserving isomorphism. For example, if $a \leq b$ we may choose $\{q_n\} \subset \mathbb{Q}$ and $\{p_n\} \subset \mathbb{Q}$ such that $q_n1_F \uparrow a$ and $p_n1_F \downarrow b$. Then $q_n1_G \leq p_n1_G$ for all $n$ and letting $n \to \infty$ shows,

$$\varphi(a) = \lim_{n \to \infty} q_n1_G \leq \lim_{n \to \infty} p_n1_G = \varphi(b).$$

The other properties of $\varphi$ are proved similarly.

3.7 Supremum and Infinum of sets

Definition 3.49. Given a set $A \subset X$, let

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

be the indicator function of $A$.

Lemma 3.50 (Properties of $\inf$ and $\sup$). We have:

1. $(\cup_{n=1}^{\infty} A_n)^c = \cap_{n=1}^{\infty} A_n^c$,
2. $\{A_n \text{ i.o.}\}^c = \{A_n^c \text{ a.a.}\}$,
3. $\limsup_{n \to \infty} A_n = \{x \in X : \sum_{n=1}^{\infty} 1_{A_n}(x) = \infty\}$,
4. $\liminf_{n \to \infty} A_n = \{x \in X : \sum_{n=1}^{\infty} 1_{A_n}(x) < \infty\}$,
5. $\sup_{k \geq n} A_k(x) = 1_{\cup_{k \geq n} A_k} = 1_{\sup_{k \geq n} A_k}$,
6. $\inf_{k \geq n} A_k(x) = 1_{\cap_{k \geq n} A_k} = 1_{\inf_{k \geq n} A_k}$.
7. \( \limsup_{n \to \infty} A_n = \limsup_{n \to \infty} 1_{A_n} \), and
8. \( \liminf_{n \to \infty} A_n = \liminf_{n \to \infty} 1_{A_n} \).

**Proof.** These results follow fairly directly from the definitions and so the proof is left to the reader. (The reader should definitely provide a proof for herself.)
Complex Numbers

**Definition 4.1 (Complex Numbers).** Let \( \mathbb{C} = \mathbb{R}^2 \) equipped with multiplication rule
\[
(a, b)(c, d) \equiv (ac - bd, bc + ad)
\]
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 \( z = x + yi \) which in the future will be written simply as \( z = x + iy \). If \( z = x + iy \), let \( Re \) \( z = x \) and Im \( z = y \).

Writing \( z = a + ib \) and \( w = c + id \), the multiplication rule in Eq. (4.1) becomes
\[
(a + ib)(c + id) \equiv (ac - bd) + i(bc + ad)
\]
and in particular \( 1^2 = 1 \) and \( i^2 = -1 \).

**Proposition 4.2.** The complex numbers \( \mathbb{C} \) with the above multiplication rule satisfies the usual definitions of a field – see Definition 2.1. For example \( 2 \) and in particular \( 1 \) becomes
\[
(x, 0) = (0, 1)
\]
and
\[
1 = (1, 0)
\]
Moreover \( \mathbb{C} \) contains \( \mathbb{R} \) as sub-field under the identification
\[
\mathbb{R} \ni a \rightarrow a1 + 0i = (a, 0) \in \mathbb{C}.
\]

**Proof.** Suppose \( z = a + ib \neq 0 \), we wish to find \( w = c + id \) such that \( zw = 1 \) and this happens by Eq. (4.2) iff
\[
\begin{align*}
ac - bd &= 1 \quad \text{and} \\
bc + ad &= 0.
\end{align*}
\]
Solving these equations as follows
\[
\begin{align*}
a(4.4) + b(4.5) & \implies (a^2 + b^2) c = a \implies Re w = c = \frac{a}{a^2 + b^2} \\
b(4.4) + a(4.5) & \implies (a^2 + b^2) d = -b \implies Im w = d = -\frac{b}{a^2 + b^2}.
\end{align*}
\]
gives implies the result in Eq. (4.3).

Probably the most painful thing to check directly is the associative law, namely that \( [z_1 z_2] z_3 = z_1 [z_2 z_3] \) for all \( z_1, z_2, z_3 \in \mathbb{C} \). This is equivalent to showing for all \( a, b, u, v, x, y \in \mathbb{R} \) that
\[
[(a + ib) (u + iv)] (x + iy) = (a + ib) [(u + iv) (x + iy)].
\]
We do this by working out both sides as follows;
\[
\begin{align*}
LHS &= [(au - bv) + i (av + bu)] (x + iy) \\
&= (au - bv) x - (av + bu) y + i [(av + bu) x + (au - bv) y]; \\
RHS &= (a + ib) [(ux - vy) + i (uy + vx)] \\
&= a (ux - vy) - b (uy + vx) + i [b (ux - vy) + a (uy + vx)].
\end{align*}
\]
The reader should now easily see that both of these expressions are in fact equal. The remaining axioms of a field are checked similarly.

- End of Lecture 9, 10/17/2012.
- Test 1 took place of lecture 10, 10/22/2012.

**Notation 4.3** We will write \( 1/z \) for \( z^{-1} \) and \( w/z \) to mean \( z^{-1} \cdot w \).

**Notation 4.4 (Conjugation and Modulus)** If \( z = a + ib \) with \( a, b \in \mathbb{R} \) let \( \bar{z} = a - ib \) and
\[
|z|^2 \equiv \bar{z} z = a^2 + b^2.
\]
Notice that
\[
Re z = \frac{1}{2} (z + \bar{z}) \text{ and } Im z = \frac{1}{2i} (z - \bar{z}).
\]

**Proposition 4.5.** Complex conjugation and the modulus operators satisfy:
\[
\begin{align*}
1. \quad & \bar{\bar{z}} = z, \\
2. \quad & \bar{zw} = \bar{z} \bar{w} \text{ and } \bar{z} + \bar{w} = \bar{z + w}; \\
3. \quad & |\bar{z}| = |z|
\end{align*}
\]
4. \(|zw| = |z||w|\) and in particular \(|z^n| = |z|^n\) for all \(n \in \mathbb{N}\).
5. \(|\text{Re } z| \leq |z|\) and \(|\text{Im } z| \leq |z|\).
6. \(|z+w| \leq |z| + |w|\).
7. \(z = 0\) iff \(|z| = 0\).
8. If \(z \neq 0\) then

\[
\frac{1}{|z|^2} \overline{z} = \frac{1}{|z|^2} |z|^2 = 1 \implies z^{-1} = \frac{\text{Re } z}{|z|^2} - i \frac{\text{Im } z}{|z|^2}.
\]

9. \(|z^{-1}| = |z|^{-1}\) and more generally \(|z^n| = |z|^n\) for all \(n \in \mathbb{Z}\).

**Proof.** 1. and 3. are geometrically obvious as well as easily verified.

2. Say \(z = a + ib\) and \(w = c + id\), then \(\overline{zw}\) is the same as \(zw\) with \(b\) replaced by \(-b\) and \(d\) replaced by \(-d\), and looking at Eq. (4.2) we see that

\[
\overline{zw} = (ac - bd) - i(bc + ad) = \overline{zw}.
\]

4. \(|zw|^2 = zw\overline{zw} = |z|^2 |w|^2\) as real numbers and hence \(|zw| = |z||w|\).
5. Geometrically obvious or also follows from

\[
|z| = \sqrt{|\text{Re } z|^2 + |\text{Im } z|^2}.
\]

6. This is the triangle inequality which may be understood geometrically or by the computation

\[
|z + w|^2 = (z + w)(\overline{z} + \overline{w}) = |z|^2 + |w|^2 + \overline{w}z + w\overline{z} = |z|^2 + |w|^2 + \overline{wz} + wz = |z|^2 + |w|^2 + 2 \text{Re}(w\overline{z}) \leq |z|^2 + |w|^2 + 2|z||w| = (|z| + |w|)^2.
\]

7. Obvious.

8. Follows from Eq. (4.3). Alternatively if \(\rho = \rho + i0 > 0\) is a real number then \(\overline{\rho^{-1}} = \rho^{-1} + i0\) as is easily verified since \(\mathbb{R}\) is a sub-field of \(\mathbb{C}\). Thus since \(\overline{\rho} = |\rho|^2\) we find

\[
\frac{1}{|\rho|^2} = \frac{1}{|\rho|^2} |\rho|^2 = 1 \implies \rho^{-1} = \frac{\text{Re } \rho}{|\rho|^2} - i \frac{\text{Im } \rho}{|\rho|^2}.
\]

9. \(|z^{-1}| = |z|^{-1}|\) and more generally \(|z^n| = |z|^n|\) for all \(n \in \mathbb{Z}\).

**Corollary 4.6.** If \(w, z \in \mathbb{C}\), then

\[
||z| - |w|| \leq |z - w|.
\]

**Proof.** Just copy the proof of Lemma 1.6.

**Lemma 4.7.** For complex numbers \(u, v, w, z \in \mathbb{C}\) with \(v \neq 0 \neq z\), we have

\[
\frac{1}{1} \frac{1}{uv} = \frac{1}{uv}, \text{ i.e. } u^{-1}v^{-1} = (uv)^{-1}
\]

\[
\frac{u}{v} \frac{w}{z} = \frac{uw}{vz} \text{ and }
\]

\[
\frac{u}{v} + \frac{w}{z} = \frac{uw + vz}{vz}.
\]

[These statements hold in any field.]
Theorem 4.10. If \( A \) is a sequence of numbers, then \( \lim_{n \to \infty} A = \infty \).

Theorem 4.9. \( \{z_n\}_{n=1}^{\infty} \subseteq \mathbb{C} \) 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 \( z = \lim_{n \to \infty} z_n \).

Proof. By assumption there exists \( M < \) such that \( |z_n| \leq M \) for all \( n \in \mathbb{N} \). Writing \( z_n = a_n + ib_n \) with \( a_n, b_n \in \mathbb{R} \) we may conclude that \( |a_n|, |b_n| \leq M \). According to Corollary 3.34 there exists an increasing function \( N \ni k \to n_k \in \mathbb{N} \) such that \( \lim_{k \to \infty} a_{n_k} = A \) exists. Similarly, we can apply Corollary 3.34 again to find an increasing function \( N \ni l \to k_l \in \mathbb{N} \) such that \( \lim_{l \to \infty} b_{n_{k_l}} = B \) exists. We now let \( w_l := z_{n_{k_l}} \) for \( l \in \mathbb{N} \). Then \( \{w_l\}_{l=1}^{\infty} \) is a subsequence of \( \{z_n\}_{n=1}^{\infty} \) which is convergent to \( A + iB \in \mathbb{C} \). Indeed, \( |w_l - (A + iB)| = |a_{n_{k_l}} - A + i(b_{n_{k_l}} - B)| \leq |a_{n_{k_l}} - A| + |b_{n_{k_l}} - B| \to 0 \) as \( l \to \infty \).

Notation 4.12 (Euclidean Spaces). Let \( \mathbb{C}^n := \{a := (a_1, \ldots, a_n) : a_i \in \mathbb{C}\} \) and \( \mathbb{R}^n := \{a := (a_1, \ldots, a_n) : a_i \in \mathbb{R}\} \subset \mathbb{C}^n \). For \( a, b \in \mathbb{C}^n \) we let \( \bar{a} = (\bar{a}_1, \ldots, \bar{a}_n) \),

\[
\bar{a} \cdot b := a_1 b_1 + \cdots + a_n b_n = \sum_{i=1}^{n} a_i b_i, \quad \text{and} \quad \|a\| = \|a\|_2 = \sqrt{\bar{a} \cdot a} = \sqrt{\sum_{i=1}^{n} |a_i|^2}.
\]

Exercise 4.1. If \( a, b \in \mathbb{C}^n \) and \( \lambda \in \mathbb{C} \), then \( a \cdot b = b \cdot a \), \( \lambda a \cdot b = a \cdot \lambda b = \lambda (a \cdot b) \) and \( \|\lambda a\| = |\lambda||a| = |\lambda| \|a\| \).

If we further assume that \( c \in \mathbb{C}^n \), then \( (a + b) \cdot c = a \cdot c + b \cdot c \).

4.1 A Matrix Perspective (Optional)

Here is a way to understand some of the basic properties of \( \mathbb{C} \) using your knowledge of linear algebra. Let \( M_z : \mathbb{C} \to \mathbb{C} \) denote multiplication by \( z = a + ib \).

We now identify \( \mathbb{C} \) with \( \mathbb{R}^2 \) by \( \mathbb{C} \ni c + id \cong \left( \begin{array}{c} c \\ d \end{array} \right) \in \mathbb{R}^2 \).

Using this identification, the product formula

\[
zw = (ac - bd) + i(bc + ad),
\]

becomes
\[ M_z w = \begin{pmatrix} ac - bd \\ bc + ad \end{pmatrix} = \begin{pmatrix} a - b \\ b \\ a \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} \]

so that
\[ M_z = \begin{pmatrix} a - b \\ b \\ a \end{pmatrix} = aI + bJ \]

where
\[ J := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \]

We now have the following simple observations;

1. \( J^2 = -I \) and \( J^t = -J \),
2. \( M_z M_w = M_w M_z \) because \( J \) and \( I \) commute,
3. we have
\[ M_z M_w = (aI + bJ)(cI + dJ) = (ac - bd)I + (ad + bc)J = M_{zw}, \]
4. the associativity of complex multiplication follows from the associativity properties of matrix multiplication,
5. \( M_z^t = aI - bJ = M_{\bar{z}} \) and in particular
6. \( M_{\bar{w}} = (M_z M_w)^t = M_w^t M_z^t = M_w M_{\bar{z}} = M_{\bar{w} \bar{z}}, \)
7. \( M_{\bar{w}} M_z = M_{\bar{z} \bar{z}} = M_{|z|^2} = \det(M_z). \)
8. \( |wz| = \det(M_{wz}) = \det(M_w M_z) = \det(M_w) \det(M_z) = |w| |z|, \)
9. \( M_z \) is invertible iff \( \det(M_z) \neq 0 \) which happens iff \( |z|^2 \neq 0 \) and in this case we know from basic linear algebra that
\[ M_z^{-1} = \frac{1}{a^2 + b^2} \begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} -b \\ a \end{pmatrix} = \frac{1}{|z|^2} M_z^t = M_{\frac{1}{|z|^2} \bar{z}}, \]
10. With this notation we have \( M_z M_w = M_{zw} \) and since \( I \) and \( J \) commute it follows that \( zw = wz \). Moreover, since matrix multiplication is associative so is complex multiplication. Also notice that \( M_z \) is invertible iff \( \det(M_z) = a^2 + b^2 = |z|^2 \neq 0 \) in which case
\[ M_z^{-1} = \frac{1}{|z|^2} \begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} -b \\ a \end{pmatrix} = M_{\frac{1}{|z|^2} \bar{z}}, \]
as we have already seen above.
Set Operations, Functions, and Counting

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, and \( \mathbb{C} \) the complex numbers. We will also use \( \mathbb{F} \) to stand for either of the fields \( \mathbb{R} \) or \( \mathbb{C} \).

5.1 Set Operations and Functions

**Notation 5.1** Given two sets \( X \) and \( Y \), let \( Y^X \) denote the collection of all functions \( f : X \rightarrow Y \). 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 5.2** More generally if \( \{X_\alpha : \alpha \in \alpha \} \) is a collection of non-empty sets, let \( X_\alpha = \prod_{\alpha \in \alpha} X_\alpha \) and \( \pi_\alpha : X_\alpha \rightarrow X_\alpha \) be the canonical projection map defined by \( \pi_\alpha (x) = x_\alpha \). If \( X_\alpha = X \) for some fixed space \( X \), then we will write \( \prod_{\alpha \in \alpha} X_\alpha \) as \( X^\alpha \) rather than \( X_\alpha \).

Recall that an element \( x \in X_\alpha \) is a “choice function,” i.e. an assignment \( x_\alpha := x(\alpha) \in X_\alpha \) for each \( \alpha \in \alpha \). The axiom of choice states that \( X_\alpha \neq \emptyset \) provided that \( X_\alpha \neq \emptyset \) for each \( \alpha \in \alpha \).

**Notation 5.3** 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

\[
\bigcup_{\alpha \in I} A_\alpha := \{x \in X : \exists \alpha \in I \; x \in A_\alpha\}
\]

and

\[
\bigcap_{\alpha \in I} A_\alpha := \{x \in X : x \in A_\alpha \forall \alpha \in I\}.
\]

**Example 5.4.** Let \( A, B, \) and \( C \) be subsets of \( X \). Then

\[
A \cap (B \cup C) = [A \cap B] \cup [A \cap C].
\]

Indeed, \( x \in A \cap (B \cup C) \iff x \in A \) and \( x \in B \cup C \iff x \in A \) and \( x \in B \) or \( x \in A \) and \( x \in C \iff x \in A \cap B \) or \( x \in A \cap C \iff x \in [A \cap B] \cup [A \cap C] \).

**Notation 5.5** We will also write \( \bigsqcup_{\alpha \in I} A_\alpha \) for \( \bigcup_{\alpha \in I} A_\alpha \) 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 \( \bigcup \) is closely related to \( \exists \) and \( \bigcap \) is closely related to \( \forall \). For example let \( \{A_n\}_{n=1}^\infty \) be a sequence of subsets from \( X \) and define

\[
\{A_n \text{ i.o.} \} := \{x \in X : \# \{n \in \mathbb{N} : x \in A_n\} = \infty\}
\]

and

\[
\{A_n \text{ a.a.} \} := \{x \in X : x \in A_n \forall n \text{ sufficiently large}\}.
\]

(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 \( x \in \{A_n \text{ i.o.} \} \iff \forall N \in \mathbb{N} \exists n \geq N \exists x \in A_n \)

and this may be expressed as

\[
\{A_n \text{ i.o.} \} = \bigcap_{n=1}^\infty \bigcup_{n \geq N} A_n.
\]

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

\[
\{A_n \text{ i.o.} \} = \bigcap_{n=1}^\infty \bigcup_{n \geq N} A_n.
\]
Let \( N \in \mathbb{N} \) \( \exists \forall n \geq N, \ x \in A_n \)

which may be written as

\[
\{ A_n \text{ a.a.} \} = \bigcup_{n=1}^{\infty} \cap_{n \geq N} A_n.
\]

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

**Definition 5.6.** If \( f : X \to Y \) is a function and \( B \subset Y \), then

\[
f^{-1}(B) := \{ x \in X : f(x) \in B \}.
\]

If \( A \subset X \) we also write,

\[
f(A) := \{ f(x) : x \in A \} \subset Y.
\]

**Example 5.7.** If \( f : X \to Y \) is a function and \( B \subset Y \), then \( f^{-1}(B^c) = [f^{-1}(B)]^c \) or to be more precise,

\[
f^{-1}(Y \setminus B) = X \setminus f^{-1}(B).
\]

To prove this notice that

\[
x \in f^{-1}(B^c) \iff f(x) \in B^c \iff f(x) \notin B \iff x \notin f^{-1}(B) \iff x \in [f^{-1}(B)]^c.
\]

On the other hand, if \( A \subset X \) it is not necessarily true that \( f(A^c) = [f(A)]^c \).

For example, suppose that \( f : \{1, 2\} \to \{1, 2\} \) is the defined by \( f(1) = f(2) = 1 \) and \( A = \{1\} \). Then \( f(A) = f(A^c) = \{1\} \) where \( [f(A)]^c = \{1\}^c = \{2\} \).

**Notation 5.8** If \( f : X \to Y \) is a function and \( \mathcal{E} \subset 2^Y \) let

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

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

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

**Definition 5.9.** Let \( \mathcal{E} \subset 2^X \) be a collection of sets, \( A \subset X \), \( i_A : A \to X \) be the inclusion map \( (i_A(x) = x \text{ for all } x \in A) \) and

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

### 5.1.1 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 5.1.** \( (\cap_{i \in I} A_i)^c = \cup_{i \in I} A_i^c \).

**Exercise 5.2.** Suppose that \( B \subset Y \), show that \( B \setminus (\cup_{i \in I} A_i) = \cap_{i \in I} (B \setminus A_i) \).

**Exercise 5.3.** \( f^{-1}(\cup_{i \in I} A_i) = \cup_{i \in I} f^{-1}(A_i) \).

**Exercise 5.4.** \( f^{-1}(\cap_{i \in I} A_i) = \cap_{i \in I} f^{-1}(A_i) \).

**Exercise 5.5.** Find a counterexample which shows that \( f(C \cap D) = f(C) \cap f(D) \) need not hold.

### 5.2 Cardinality

In this section, \( X \) and \( Y \) be sets.

**Definition 5.10 (Cardinality).** We say \( \text{card} (X) \leq \text{card} (Y) \) if there exists an injective map, \( f : X \to Y \) and \( \text{card} (Y) \geq \text{card} (X) \) if there exists a surjective map \( g : Y \to X \). We say \( \text{card} (X) = \text{card} (Y) \) (also denoted as \( X \sim Y \)) if there exists a bijections, \( f : X \to Y \).
5.3 Finite Sets

Notation 5.13 (Integer Intervals) For \( n \in \mathbb{N} \) we let
\[
J_n := \{1, 2, \ldots, n\} := \{k \in \mathbb{N} : k \leq n\}.
\]

Definition 5.14. We say a non-empty set, \( X \), is finite if \( \text{card}(X) = \text{card}(J_n) \) for some \( n \in \mathbb{N} \). We will also write \( \#(X) = n \) to indicate that \( \text{card}(X) = \text{card}(J_n) \). [It is shown in Theorem 5.17 below that \( \#(X) \) is well defined, i.e. it is not possible for \( \text{card}(X) = \text{card}(J_n) \) and \( \text{card}(X) = \text{card}(J_m) \) unless \( m = n \).]

Lemma 5.15. Suppose \( n \in \mathbb{N} \) and \( k \in J_{n+1} \), then \( \text{card}(J_{n+1} \setminus \{k\}) = \text{card}(J_n) \).

Proof. Let \( f : J_n \to J_{n+1} \setminus \{k\} \) be defined by
\[
f(x) = \begin{cases} 
x & \text{if } x < k \\
x + 1 & \text{if } x \geq k
\end{cases}
\]

Then \( f \) is the desired bijection. \(^2\)

Alternatively. If \( n = 1 \), then \( J_2 = \{1, 2\} \) and either \( J_2 \setminus \{k\} = J_1 \) or \( J_2 \setminus \{k\} = \{2\} \), either way \( \text{card}(J_2 \setminus \{k\}) = \text{card}(J_1) \). Now suppose that result holds for a given \( n \in \mathbb{N} \) and \( k \in J_{n+2} \). If \( k = n+1 \) we have \( J_{n+1} \setminus \{k\} = J_{n+1} \) so \( \text{card}(J_{n+1} \setminus \{k\}) = \text{card}(J_{n+1}) \) while if \( k \in J_{n+1} \subset J_{n+2} \), then \( J_{n+2} \setminus \{k\} = (J_{n+1} \setminus \{k\}) \cup \{n+2\} \subset J_n \cup \{n+2\} \cup J_n = J_{n+1} \).

Lemma 5.16. If \( m, n \in \mathbb{N} \) with \( n > m \), then every map, \( f : J_m \to J_n \), is not injective.

Proof. If \( f : J_m \to J_n \) were injective, then \( f|_{J_{m+1}} : J_{m+1} \to J_n \) would be injective as well. Therefore it suffices to show there is no injective map, \( f : J_m \to J_n \) for all \( m \in \mathbb{N} \). We prove this last assertion by induction on \( m \).

The case \( m = 1 \) is trivial as \( J_1 = \{1\} \) so the only function, \( f : J_2 \to J_1 \) is the function, \( f(1) = f(2) \) which is not injective.

Now suppose \( m \geq 1 \) and there were an injective map, \( f : J_{m+2} \to J_m \).

Letting \( k := f(m+2) \) we would have, \( f|_{J_{m+1}} : J_{m+1} \to J_{m+1} \setminus \{k\} \equiv J_m \), which would produce and injective map from \( J_{m+1} \) to \( J_m \). However this contradicts the induction hypothesis and thus completes the proof.

Theorem 5.17. If \( m, n \in \mathbb{N} \), then \( \text{card}(J_m) \leq \text{card}(J_n) \) iff \( m \leq n \). Moreover, \( \text{card}(J_n) = \text{card}(J_m) \) iff \( m = n \) and hence \( \text{card}(J_m) < \text{card}(J_n) \) iff \( m < n \).

Proof. As \( J_m \subset J_n \) if \( m \leq n \) and \( J_m = J_n \) if \( m = n \), it is only the forward implications that have any real content. If \( \text{card}(J_m) \leq \text{card}(J_n) \), there exists an injective map, \( g : J_m \to J_n \). According to Lemma 5.16 this can only happen if \( m \leq n \). If \( \text{card}(J_n) = \text{card}(J_m) \), then \( \text{card}(J_n) \leq \text{card}(J_m) \) and \( \text{card}(J_m) \leq \text{card}(J_n) \) and hence \( m \leq n \) and \( m \leq i \), i.e. \( m = n \).

Proposition 5.18. If \( X \) is a finite set with \( \#(X) = n \) and \( S \) is a non-empty subset of \( X \), then \( S \) is a finite set and \( \#(S) = k \leq n \). Moreover if \( \#(S) = n \), then \( S = X \).

Proof. It suffices to assume that \( X = J_n \) and \( S \subset J_n \). We now give two proofs of the result.

Proof 1. Let \( S_1 = S \) and \( f(1) := \min S \geq 1 \). If \( S_2 := S_1 \setminus \{f(1)\} \) is not empty, let \( f(2) := \min S_2 \geq 2 \). We then continue this construction inductively. So if \( f(k) = \min S_k \geq k \) has been constructed, then we define \( S_{k+1} := S_k \setminus \{f(k)\} \). If \( S_{k+1} \neq \emptyset \) we define \( f(k+1) := \min S_{k+1} \geq k+1 \). As \( f(k) \geq k \)

\(^2\) The inverse map, \( g : J_{n+1} \setminus \{k\} \to J_n \), to \( f \) is given by
\[
g(y) = \begin{cases} 
y & \text{if } y < k \\
y - 1 & \text{if } y > k
\end{cases}
\]

You should read \( \#(X) = n \), as \( X \) is a set with \( n \) elements.
for all k that f is defined, this process has to stop after at most n steps. Say it stops at k so that \( S_{k+1} = \emptyset \). Then \( f : J_k \to S \) is a bijection and therefore S is finite and \( \#(S) = k \leq n \). Moreover, the only way that \( k = n \) is if \( f(k) = k \) at each step of the construction so that \( f : J_n \to S \) is the identity map in this case, i.e. \( S = J_n \).

**Proof.** We prove this by induction on n. When \( n = 1 \) the only no-empty subset of S of J_1 is J_1 itself. Thus \( \#(S) = 1 \) and \( S = J_1 \). Now suppose that the result hold for some \( n \in \mathbb{N} \) and let \( S \subseteq J_{n+1} \). If \( n + 1 \not\in S \), then \( S \subseteq J_n \) and by the induction hypothesis we know \( \#(S) = k \leq n < n+1 \). So now suppose that \( n + 1 \in S \) and let \( S' := S \setminus \{n+1\} \subseteq J_n \). Then by the induction hypothesis, \( S' \) is a finite set and \( \#(S') = k \leq n \), i.e. there exists a bijection, \( f' : J_k \to S' \) and \( S' = J_k = k = n \). Therefore \( f : J_{k+1} \to S \) given by \( f = f' \) on \( J_k \) and \( f(k+1) = n+1 \) is a bijections from \( J_{k+1} \) to \( S \). Therefore \( \#(S) = k + 1 \leq n + 1 \) with equality iff \( S' = J_n \) which happens iff \( S = J_{n+1} \).

**Exercise 5.7.** Suppose that \( m, n \in \mathbb{N} \), show \( J_{m+n} = J_m \cup (m + J_n) \) and \( (m + J_n) \cap J_m = \emptyset \). Use this to conclude if \( X \) is a disjoint union of two non-empty finite sets, \( X_1 \) and \( X_2 \), then \( \#(X) = \#(X_1) + \#(X_2) \).

**Exercise 5.8.** Suppose that \( m, n \in \mathbb{N} \), show \( J_m \times J_n \sim J_{mn} \). Use this to conclude if \( X \) and \( Y \) are two non-empty sets, then \( \#(X \times Y) = \#(X) \cdot \#(Y) \).

### 5.4 Countable and Uncountable Sets

**Definition 5.22 (Countability).** A set \( X \) is said to be **countable** if \( X = \emptyset \) or if there exists a surjective map, \( f : \mathbb{N} \to X \). Otherwise \( X \) is said to be **uncountable**.

**Remark 5.23.** From Proposition 5.14, it follows that \( X \) is countable iff there exists an injective map, \( g : X \to \mathbb{N} \). This may be succinctly stated as: \( X \) is countable iff \( \#(X) \leq \#(\mathbb{N}) \). From a practical point of view as set \( X \) is countable iff the elements of \( X \) may be arranged into a linear list,

\[
X = \{x_1, x_2, x_3, \ldots \}.
\]

**Example 5.24.** The integers, \( \mathbb{Z} \), are countable. In fact \( \mathbb{N} \sim \mathbb{Z} \), for example define \( f : \mathbb{N} \to \mathbb{Z} \) by

\[
(f(1), f(2), f(3), f(4), f(5), f(6), f(7), \ldots ) := (0, -1, 2, -2, 3, -3, \ldots )
\]

**Remark 5.25 (Countability in a nutshell).** If \( f : \mathbb{N} \to X \) is surjective, then \( g(x) := \min f^{-1}(\{x\}) \) defines an injective map, \( g : X \to \mathbb{N} \). If \( f : X \to \mathbb{N} \) is injective, then \( f(n) := g^{-1}(n) \) for \( n \in g(X) =: S \) and \( f(n) = x_0 \in X \) for \( n \notin S \) defines a surjective map, \( f : \mathbb{N} \to X \). Moreover, if \( S \) is a subset of \( \mathbb{N} \) we may list its elements in increasing order so that either
In more detail, let \( T \) in the first case card (\( X \)) = card (\( J_k \)) while in the second card (\( S \)) = card (\( N \)).

Define \( f(j) := n_j \) to set up the bijections between \( J_k \) or \( N \) and \( S \).

The above arguments demonstrate that the following statements are equivalent;

1. \( X \) is countable, i.e. there exists a surjective map \( f : N \rightarrow X \).
2. card (\( X \)) ≤ card (\( N \)), i.e. there exists an injective map, \( g : X \rightarrow N \).
3. There exists \( S \subset N \) such that card (\( X \)) = card (\( S \)). Furthermore card (\( S \)) = card (\( J_k \)) for some \( k \in N \) if \( S \) is bounded and card (\( S \)) = card (\( N \)) if \( S \) is unbounded.
4. Either card (\( X \)) = card (\( N \)) for card (\( X \)) = card (\( J_k \)) for some \( k \in N \).

Formal proofs of these observations are given above and below.

**Lemma 5.26.** If \( S \subset N \) is an unbounded set, then card (\( S \)) = card (\( N \)).

**Proof.** The main idea is that any subset, \( S \subset N \), may be given as an finite or infinite list written in increasing order, i.e.

\[
S = \{n_1, n_2, n_3, \ldots \} \text{ with } n_1 < n_2 < n_3 < \ldots.
\]

If the list is finite, say \( S = \{n_1, \ldots, n_k\} \), then \( n_k \) is an upper bound for \( S \). So \( S \) will be unbounded iff only if the list is infinite in which case \( f : N \rightarrow S \) defined by \( f(k) = n_k \) defines a bijection.

**Formal proof.** Define \( f : N \rightarrow S \) via, let

\[
\begin{align*}
S_1 &= S \text{ and } f(1) := \min S_1, \\
S_2 &= S_1 \setminus \{f(1)\} \text{ and } f(2) := \min S_2, \\
S_3 &= S_2 \setminus \{f(2)\} \text{ and } f(3) := \min S_3 \\
& \vdots
\end{align*}
\]

In more detail, let \( T \) denote those \( n \in N \) such that there exists \( f : J_n \rightarrow S \) and \( \{S_k \subset S\}_{k=1}^{n} \) satisfying, \( S_1 = S \), \( f(k) = \min S_k \) and \( S_{k+1} = S_k \setminus \{f(k)\} \) for \( 1 \leq k < n \). If \( n \in T \), we may define \( S_{n+1} := S_n \setminus \{f(n)\} \) and \( f(n+1) := \min S_{n+1} \) in order to show \( n+1 \in T \). Thus \( T = N \) and we have constructed an injective map, \( f : N \rightarrow S \). Moreover \( \cap_{k \in N} S_k \subset N \setminus J_n \) for all \( n \) and therefore \( \cap_{k \in N} S_k = \emptyset \). Thus it follows that \( f \) is a bijection.

**End of Lecture 14, 10/31/2012.**

The following theorem summarizes most of what we need to know about counting and countability.

**Theorem 5.27.** The following properties hold;

1. \( N \times N \) is countable and in fact \( N \times N \sim N \), i.e. there exists a bijective map, \( h, \) from \( N \) to \( N \times N \).
2. If \( X \) and \( Y \) are countable, then \( X \times Y \) is countable.
3. If \( \{X_n\}_{n \in N} \) are countable sets then \( X := \bigcup_{n=1}^{\infty} X_n \) is a countable set.
4. If \( X \) is countable, then either there exists \( n \in N \) such that \( X \sim J_n \) or \( X \sim N \).
5. If \( S \subset N \) and \( S \sim J_n \) for some \( n \in N \) then \( S \) is bounded.
6. If \( X \) is a set and card \( J_n \leq \text{card} \ X \) for all \( n \in N \) then card \( N \leq \text{card} \ X \).
7. If \( A \subset X \) is a subset of a countable set \( X \) then \( A \) is countable.

**Proof.** We take each item in turn.

1. Put the elements of \( N \times N \) into an array of the form

\[
(1,1) \quad (1,2) \quad (1,3) \quad \ldots \\
(2,1) \quad (2,2) \quad (2,3) \quad \ldots \\
(3,1) \quad (3,2) \quad (3,3) \quad \ldots \\
\vdots \quad \vdots \quad \vdots \quad \ddots
\]

and then “count” these elements by counting the sets \( \{(i,j) : i+j = k\} \) one at a time. For example let \( h(1) = (1,1) \), \( h(2) = (2,1) \), \( h(3) = (1,2) \), \( h(4) = (3,1) \), \( h(5) = (2,2) \), \( h(6) = (1,3) \) and so on. In other words we put \( N \times N \) into the following list form,

\[
N \times N = \{(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3), \ldots \}.
\]

2. If \( f : N \rightarrow X \) and \( g : N \rightarrow Y \) are surjective functions, then the function \( (f \times g) \circ h : N \rightarrow X \times Y \) is surjective where \( (f \times g)(m,n) := (f(m), g(n)) \) for all \( (m, n) \in N \times N \).

3. By assumption there exists surjective maps, \( f_n : N \rightarrow X_n \), for each \( n \in N \). Let \( h(n) := (a(n), b(n)) \) be the bijection constructed for item 1. Then \( f : N \rightarrow X_n \) defined by \( f(n) := f_n(h(n)) \) is a surjective map.

4. To see this let \( f : N \rightarrow X \) be a surjective map and let \( g(x) := \min f^{-1}\{\{x\}\} \) for all \( x \in X \). Then \( g : X \rightarrow N \) is an injective map. Let \( S : = g(X) \), then \( g : X \rightarrow S \subset N \) is a bijection. So it remains to show \( S \sim N \) or \( S \sim J_n \) for some \( n \in N \). If \( S \) is unbounded, then \( S \sim N \) as we have already seen. So it suffices to consider the case where \( S \) is bounded. If \( S \) is bounded by \( 1 \) then \( S = \{1\} = J_1 \) and we are done. Now assume the result is true if \( S \) is bounded by \( n \in N \) and now suppose that \( S \) is bounded by \( n+1 \). If \( n+1 \not\in S \), then \( S \) is bounded by \( n \) and so by induction, \( S \sim J_k \) for some \( k \leq n < n+1 \). If \( n+1 \in S \), then from above, \( S \setminus \{n+1\} \sim J_k \) for some \( k \leq n \), i.e. there exists a bijection, \( f : J_k \rightarrow S \setminus \{n+1\} \). We then extend \( f \) to \( J_{k+1} \) by setting \( f(k+1) := n+1 \) which shows \( J_{k+1} \sim S \).
5. We again prove this by induction on \( n \). If \( n = 1 \), then \( S = \{ m \} \) for some \( m \in \mathbb{N} \) which is bounded. So suppose for some \( n \in \mathbb{N} \), every subset \( S \subset \mathbb{N} \) with \( S \sim J_n \) is bounded. Now suppose that \( S \subset \mathbb{N} \) with \( S \sim J_{n+1} \). Then \( f(J_n) \sim J_n \) and hence \( f(J_n) \) is bounded in \( \mathbb{N} \). Then \( \max f(J_n) \cap \{ f(n+1) \} \) is an upper bound for \( S \). This completes the inductive argument.

6. For each \( n \in \mathbb{N} \) there exists and injection, \( f_n : J_n \to X \). By replacing \( X \) by \( X_0 := \bigcup_{n \in \mathbb{N}} f_n(J_n) \) we may assume that \( X = \bigcup_{n \in \mathbb{N}} f_n(J_n) \). Thus there exists a surjective map, \( f : X \to X \) defined by \( f(x) := \min f^{-1}\{ \{ x \} \} \) for all \( x \in X \) and let \( S := g(X) \). To finish the proof we need only show that \( S \) is unbounded. If \( S \) were bounded, then we would find \( k \in \mathbb{N} \) such that \( J_k \sim S \). However this is impossible since \( \text{card}(J_n) \leq \text{card}(X) = \text{card}(J_k) \) would imply \( n \leq k \) even though \( n \) can be chosen arbitrarily in \( \mathbb{N} \).

7. If \( g : X \to \mathbb{N} \) is an injective map then \( g_{|A} : A \to \mathbb{N} \) is an injective map and therefore \( A \) is countable. \[ \blacksquare \]

Lemma 5.28. If \( X \) is a countable set which contains \( Y \subset X \) with \( Y \sim \mathbb{N} \), then \( X \sim \mathbb{N} \).

Proof. By assumption there is an injective map, \( g : X \to \mathbb{N} \) and a bijective map, \( f : \mathbb{N} \to Y \). It then follows that \( g \circ f : \mathbb{N} \to \mathbb{N} \) is injective from which it follows that \( g(X) \) is unbounded. [Indeed, \( (g \circ f)(J_n) \subset g(X) \) for all \( n \) implies \( \text{card}(J_n) \leq \text{card}(g(X)) \) for all \( n \) which implies \( g(X) \) is unbounded by Corollary 5.21] Therefore \( X \sim g(X) \sim \mathbb{N} \) by Lemma 5.26. \[ \blacksquare \]

Corollary 5.29. We have \( \text{card}(Q) = \text{card}(\mathbb{N}) \) and in fact, for any \( a < b \) in \( \mathbb{R} \), \( \text{card}(\mathbb{Q} \cap (a, b)) = \text{card}(\mathbb{N}) \).

Proof. First off \( \mathbb{Q} \) is a countable since \( \mathbb{Q} \) may be expressed as a countable union of countable sets;

\[ \mathbb{Q} = \bigcup_{m \in \mathbb{N}} \left\{ \frac{n}{m} : n \in \mathbb{Z} \right\}. \]

From this it follows that \( \mathbb{Q} \cap (a, b) \) is countable for all \( a < b \) in \( \mathbb{R} \). As these sets are not finite, they must have the cardinality of \( \mathbb{N} \). \[ \blacksquare \]

Theorem 5.30 (Uncountability results). 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. 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 \( i(A) := \{ x_n : n \in \mathbb{N} \} \subset X \) for all \( A \subset \mathbb{N} \). 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 seen that \( 2^\mathbb{N} \) is uncountable. \[ \blacksquare \]

Corollary 5.31. The set \( \{ 0, 1 \} := \{ a \in \mathbb{R} : 0 < a < 1 \} \) is uncountable while \( \mathbb{Q} \cap \{ 0, 1 \} \) is countable. More generally, for any \( a < b \) in \( \mathbb{R} \), \( \text{card}(\mathbb{Q} \cap (a, b)) = \text{card}(\mathbb{N}) \) while \( \text{card}(\mathbb{Q} \cap (a, b)) > \text{card}(\mathbb{N}) \).

Proof. From Section 3.4 the set \( \{ 0, 1, 2, \ldots, 8 \}^\mathbb{N} \) can be mapped injectively into \( \{ 0, 1 \} \) and therefore it follows from Theorem 5.30 that \( \{ 0, 1 \} \) is uncountable. For each \( m \in \mathbb{N} \), let \( A_m := \{ \frac{n}{m} : n \in \mathbb{N} \} \). Since \( \mathbb{Q} \cap \{ 0, 1 \} = \bigcup_{m=1}^{\infty} A_m \) and \( \#(X_m) < \infty \) for all \( m \), another application of Theorem 5.27 shows \( \mathbb{Q} \cap \{ 0, 1 \} \) is countable.

The fact that these results hold for any other finite interval follows from the fact that \( f(0, 1) \to (a, b) \) defined by \( f(t) := a + t(b - a) \) is a bijection. \[ \blacksquare \]

Definition 5.32. We say a non-empty set \( X \) is infinite if \( X \) is not a finite set.

Example 5.33. Any unbounded subset, \( S \subset \mathbb{N} \), is an infinite set according to Theorem 5.20.

Theorem 5.34. Let \( X \) be a non-empty set. The following are equivalent;

1. \( X \) is an infinite set.
2. \( \text{card}(J_n) \leq \text{card}(X) \) for all \( n \in \mathbb{N} \).
3. \( \text{card}(\mathbb{N}) \leq \text{card}(X) \).
4. \( \text{card}(X \setminus \{ x \}) = \text{card}(X) \) for some (or all) \( x \in X \).

Proof. 1. \( \Rightarrow \) 2. Suppose that \( X \) is an infinite set. We show by induction that \( \text{card}(J_n) \leq \text{card}(X) \) for all \( n \in \mathbb{N} \). Since \( X \) is not empty, there exists \( x \in X \) and we may define \( f : J_1 \to X \) by \( f(1) = x \) in order to learn \( \text{card}(J_1) \leq \text{card}(X) \). Suppose we have shown \( \text{card}(J_n) \leq \text{card}(X) \) for some \( n \in \mathbb{N} \), i.e. there exists and injective map, \( f : J_n \to X \). If \( f(J_n) = X \) it would follows that \( \text{card}(X) = \text{card}(J_n) \) and would violated the assumption that \( X \) is not a finite set. Thus there exists \( x \in X \setminus f(J_n) \) and we may define \( f' : J_{n+1} \to X \)
by $f|_{J_n} = f$ and $f'(n+1) = x$. Then $f' : J_{n+1} \to X$ is injective and hence $\text{card}(J_{n+1}) \leq \text{card}(X)$.

2. $\iff$ 3. This is the content of Theorem [B.19]

3. $\implies$ 4. Let $x_1 \in X$ and $f : \mathbb{N} \to X$ be an injective map such that $f(1) = x_1$. We now define a bijections, $\psi : X \to X \setminus \{x_1\}$ by

$$\psi(x) = \begin{cases} x & \text{if } x \notin f(\mathbb{N}) \\ f(i + 1) & \text{if } x = f(i) \in f(\mathbb{N}). \end{cases}$$

4. $\implies$ 1. We will prove the contrapositive. If $X$ is a finite and $x \in X$, we have seen that $\text{card}(X \setminus \{x\}) < \text{card}(X)$, namely $\#(X \setminus \{x\}) = \#(X) - 1$.

The next two theorems summarizes the properties of cardinalities that have been proven above.

**Theorem 5.35 (Cardinality/Counting Summary I).** Given a non-empty set $X$, then one and only one of the following statements holds:

1. There exists a unique $n \in \mathbb{N}$ such that $\text{card}(X) = \text{card}(J_n)$,
2. $\text{card}(X) = \text{card}(\mathbb{N})$,
3. $\text{card}(X) > \text{card}(\mathbb{N})$.

Cases 2. or 3. hold iff $\text{card}(J_n) \leq \text{card}(X)$ for all $n \in \mathbb{N}$ which happens iff $\text{card}(\mathbb{N}) \leq \text{card}(X)$.

If $X$ satisfies case 1. we say $X$ is a finite set. If $X$ satisfies case 2 we say $X$ is a countably infinite set and if $X$ satisfies case 3. we say $X$ is an uncountably infinite set.

**Theorem 5.36 (Cardinality/Counting Summary II).** Let $X$ and $Y$ be sets and $S$ be a subset of $\mathbb{N}$.

1. If $S \subset \mathbb{N}$ is an unbounded set, then $\text{card}(S) = \text{card}(\mathbb{N})$.
2. If $S \subset \mathbb{N}$ is a bounded set then $\text{card}(S) = \text{card}(J_n)$ for some $n \in \mathbb{N}$.
3. If $\{X_k\}_{k=1}^{\infty}$ are subsets of $X$ such that $\text{card}(X_k) \leq \text{card}(\mathbb{N})$, then $\text{card}(\bigcup_{k=1}^{\infty} X_k) \leq \text{card}(\mathbb{N})$.
4. If $X$ and $Y$ are sets such that $\text{card}(X) \leq \text{card}(\mathbb{N})$ and $\text{card}(Y) \leq \text{card}(\mathbb{N})$, then $\text{card}(X \times Y) \leq \text{card}(\mathbb{N})$.
5. $\text{card}(\mathbb{Q}) = \text{card}(\mathbb{N}) = \text{card}(\mathbb{Z})$.
6. For any $a < b$ in $\mathbb{R}$, $\text{card}(\mathbb{Q} \cap (a, b)) = \text{card}(\mathbb{N})$ while $\text{card}(\mathbb{Q}^c \cap (a, b)) > \text{card}(\mathbb{N})$.

**5.4.1 Exercises**

**Exercise 5.9.** Show that $\mathbb{Q}^n$ is countable for all $n \in \mathbb{N}$.
Part II

Normed and Metric Spaces
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 \( x = y \in X \), and
3. (Triangle inequality) \( d(x, z) \leq d(x, y) + d(y, z) \) for all \( x, y, z \in X \).

Example 6.2. Here are a few immediate examples of metric spaces;

1. Let \( X \) be any set and then define,
\[
    d(x, y) = \begin{cases} 
        0 & \text{if } x = y \\
        1 & \text{if } x \neq y
    \end{cases}
\]

2. Let \( X = \mathbb{R} \) with \( d(x, y) := |y - x| \). Notice that
\[
    d(x, z) = |x - y + y - z| \\
    \leq |x - y| + |y - z| = d(x, y) + d(y, z).
\]

3. Let \( X \) be any subset of \( \mathbb{C} \) and define \( d(w, z) := |z - w| \).

In general our typical example of a metric space will often be a generalization of the last example above, see Example 6.12.

6.1 Normed Spaces [Linear Algebra Meets Analysis]

6.1.1 Review of Vector Spaces and Subspaces

Definition 6.3 (Vector Space). A vector space is a non-empty set \( Z \) of objects, called vectors, equipped with an addition operation “+” and scalar (=\( \mathbb{R} \) or maybe \( \mathbb{C} \)) multiplication “·” satisfying all of the properties above: i.e. For all \( u, v, w \in Z \) and \( a, b \in \mathbb{R} \):

1. Associativity of addition: \( u + (v + w) = (u + v) + w \).
2. Commutativity of addition: \( v + w = w + v \).
3. Identity element of addition: There is an element \( 0 \in Z \) such that \( 0 + v = v \) for all \( v \).
4. Inverse elements of addition: \(-v + v = 0 \) for all \( v \in Z \). (In fact \(-v = (-1) \cdot v \).)
5. Distributivity of scalar multiplication with vector addition: \( a \cdot (v + w) = a \cdot v + a \cdot w \).
6. Distributivity of scalar multiplication with respect to field addition \( (a + b) \cdot v = a \cdot v + b \cdot v \).
7. Compatibility of scalar multiplication with the multiplication on \( \mathbb{R} \): \( a \cdot (b \cdot v) = (ab) \cdot v \).
8. Identity element of scalar multiplication \( 1 \cdot v = v \) for all \( v \in Z \).

Example 6.4. Here are two fundamental examples of vector spaces.

1. \( \mathbb{R} \) with usual vector addition and scalar multiplication is vector space over \( \mathbb{R} \).
2. \( \mathbb{C}^n \) with usual vector addition and scalar multiplication is vector space over \( \mathbb{C} \).

Notation 6.5 If \( T \) and \( X \) are sets, let \( X^T \) denote the collection of functions, \( f : T \to X \).

Example 6.6 (The Main Umbrella Example). Let \( T \) be a non-empty set and let \( Z := \mathbb{R}^T \). For \( f, g \in Z \) and \( \lambda \in \mathbb{R} \) we define \( f + g \) and \( \lambda \cdot f \) by
\[
(f + g)(t) = f(t) + g(t) \quad \text{(addition in } \mathbb{R} \text{) for all } t \in T \\
(\lambda \cdot f)(t) = \lambda f(t) \quad \text{(multiplication in } \mathbb{R} \text{) for all } t \in T.
\]

It can now be checked that \( Z \) is a vector space so that functions have now become vectors! Essentially all other examples of vector spaces we give will be related to an example of this form. The same observations show \( \mathbb{C}^T \) is a complex vector space.

Example 6.7. For example; \( \mathbb{R}^3 = \mathbb{R}^{\{1,2,3\}} \) and more generally \( \mathbb{R}^n = \mathbb{R}^{J_n} \) where \( J_n := \{1,2,\ldots,n\} \). In this setting we usually specify \( x \in \mathbb{R}^{J_n} \) by listing its values \( (x(1),\ldots,x(n)) \). To abbreviate notation a bit more we will usually write \( x(i) \) as \( x_i \) so that \( (x(1),\ldots,x(n)) \) becomes \( (x_1,\ldots,x_n) \).
Example 6.8. The vector space of $2 \times 2$ matrices:

$$M_{2 \times 2} = \{ A : A \text{ is a } 2 \times 2 \text{ - matrix } \} = \{ A : \{(1,1), (1,2), (2,1), (2,2)\} \rightarrow \mathbb{R} \}.$$  

This can be generalized.

Definition 6.9 (Subspace). Let $Z$ be a vector space. A non-empty subset, $H \subset Z$, is a subspace of $Z$ if $H$ is closed under addition and scalar multiplication. Note, if $H$ is a subspace and $v \in H$, then $0 = 0 \cdot v \in H$.

The vector space $\mathbb{R}^T$ and $\mathbb{C}^T$ are typically the “largest” vector spaces we will consider in this course.

Example 6.10. Here are three common subspaces of $\mathbb{R}^n$:

1. $H = \{ f \in Z : f \text{ is continuous} \}$.
2. $H = \{ f \in Z : f \text{ is continuously differentiable} \}$.
3. $H = \{ f \in Z : f \text{ is differentiable at } \pi \}$.

6.1.2 Normed Spaces

Definition 6.11. A norm on a vector space $Z$ is a function $\| \cdot \| : Z \rightarrow [0, \infty)$ such that

1. (Homogeneity) $\| \lambda f \| = |\lambda| \| f \|$ for all $\lambda \in \mathbb{R}$ and $f \in Z$.
2. (Triangle inequality) $\| f + g \| \leq \| f \| + \| g \|$ 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 or normed space for short.

Example 6.12. If $(Z, \| \cdot \|)$ is a normed space, then $d(x, y) := \| x - y \|$ is a metric on $Z$ and restricts to a metric on any subset of $Z$.

Example 6.13 (Normed Spaces). The following are normed spaces:

1. $Z = \mathbb{R}$ with $\| x \| = |x|$.
2. $Z = \mathbb{C}$ with $\| z \| = |z|$.
3. $Z = \mathbb{C}^n$ with

$$\| z \|_1 := \sum_{i=1}^{n} |z_i| \text{ for } z = (z_1, \ldots, z_n) \in Z.$$  

The triangle inequality is easily verified here since,

$$\| z + w \|_1 = \sum_{i=1}^{n} |z_i + w_i| \leq \sum_{i=1}^{n} (|z_i| + |w_i|) = \sum_{i=1}^{n} |z_i| + \sum_{i=1}^{n} |w_i| = \| z \|_1 + \| w \|_1.$$  

4. Let $X$ be a set and for any function $f : X \rightarrow \mathbb{C}$, let

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

Then $Z := \{ f : X \rightarrow \mathbb{C} : \| f \| < \infty \}$ is a vector space and $\| \cdot \|_u$ is a norm on $Z$.

Exercise 6.1. Verify the last item of Example 6.13. That is let $X$ be a set and for any function $f : X \rightarrow \mathbb{C}$, let

$$\| f \|_u := \sup_{x \in X} |f(x)|.$$  

Show $Z := \{ f : X \rightarrow \mathbb{C} : \| f \|_u < \infty \}$ is a vector space and $\| \cdot \|_u$ is a norm on $Z$.

Our next goal is to show that $\| \cdot \|$ defined in Eq. (4.7) defines a norm on $\mathbb{R}^n$ and $\mathbb{C}^n$. We will begin by proving the important Cauchy-Schwarz inequality.

Lemma 6.14. If $x, y \geq 0$ and $\rho > 0$, then

$$xy \leq \frac{1}{2} \left( \rho x^2 + \frac{1}{\rho} y^2 \right) \quad (6.1)$$  

with equality when $\rho = y/x$ in the case $x > 0$.

Proof. Since

$$0 \leq \left( \sqrt{\rho} x - \frac{y}{\sqrt{\rho}} \right)^2 = \rho x^2 + \frac{1}{\rho} y^2 - 2xy,$$

with equality iff $\sqrt{\rho} x = \frac{y}{\sqrt{\rho}}$, i.e. iff $\rho = y/x$, we see that

$$xy \leq \frac{1}{2} \left( \rho x^2 + \frac{1}{\rho} y^2 \right)$$  

with equality iff $\rho = y/x$.  

\[\square\]
Definition 6.15. The two norm, \( \| \cdot \|_2 \) on \( \mathbb{C}^n \) is the function defined by

\[
\| z \|_2 = \sqrt{\sum_{i=1}^{n} |z_i|^2} \text{ for all } z \in \mathbb{C}^n.
\]

Theorem 6.16 (Cauchy-Schwarz Inequality). For \( a, b \in \mathbb{C}^n \), \( |a \cdot b| \leq \| a \|_2 \cdot \| b \|_2 \).

Proof. The inequality holds true if \( a = 0 \) so we may now assume \( a \neq 0 \). Using Lemma 6.14 with \( x = |a_i| \) and \( y = |b_i| \) we find for any \( \rho > 0 \) that

\[
|a \cdot b| \leq \sum_{i=1}^{n} a_i b_i \leq \sum_{i=1}^{n} |a_i| |b_i| = \sum_{i=1}^{n} |a_i| \cdot |b_i| \\
\leq \sum_{i=1}^{n} \frac{1}{2} \left( \rho |a_i|^2 + \frac{1}{\rho} |b_i|^2 \right) = \frac{1}{2} \left[ \rho \| a \|_2^2 + \frac{1}{\rho} \| b \|_2^2 \right].
\]

Taking \( \rho = \| b \|_2 / \| a \|_2 \) then completes the proof.

Theorem 6.17 (Triangle Inequality). The function, \( \| \cdot \|_2 \), on \( \mathbb{C}^n \) is a norm and in particular for \( a, b \in \mathbb{C}^n \),

\[
\| a + b \|_2 \leq \| a \|_2 + \| b \|_2.
\]

Proof. The main point is to prove the triangle inequality. The proof is as follows:

\[
\| a + b \|_2^2 = (a + b) \cdot (\overline{a + b}) = (a + b) \cdot (\overline{a} + \overline{b}) \\
= (a + b) \cdot \overline{a} + (a + b) \cdot \overline{b} = \| a \|_2^2 + b \cdot \overline{a} + a \cdot \overline{b} + \| b \|_2^2 \\
= \| a \|_2^2 + \| b \|_2^2 + 2 \Re(a \cdot \overline{b}) \leq \| a \|_2^2 + \| b \|_2^2 + 2 |\Re(a \cdot \overline{b})| \\
\leq \| a \|_2^2 + \| b \|_2^2 + 2 \| a \|_2 \cdot \| b \|_2 \leq \| a \|_2^2 + \| b \|_2^2 + 2 \| a \|_2 \cdot \| b \|_2 \leq \| a \|_2^2 + \| b \|_2^2 + 2 \| a \|_2 \cdot \| b \|_2 = \| a \|_2^2 + \| b \|_2^2.
\]

The remaining properties of a norm are easily checked. For example, if \( \lambda \in \mathbb{C} \) and \( a \in \mathbb{C}^n \), then

\[
\| \lambda a \|_2 = \sqrt{\sum_{i=1}^{n} |\lambda a_i|^2} = \sqrt{\sum_{i=1}^{n} |\lambda|^2 |a_i|^2} \\
= \sqrt{|\lambda|^2 \sum_{i=1}^{n} |a_i|^2} = \sqrt{|\lambda|^2} \sqrt{\sum_{i=1}^{n} |a_i|^2} = |\lambda| \| a \|_2.
\]

Fact 6.18 For \( 0 < p < \infty \), let \( \| \cdot \|_p : \mathbb{C}^n \to [0, \infty) \) be defined by

\[
\| z \|_p = \left( \sum_{i=1}^{n} |z_i|^p \right)^{1/p}
\]

Then \( \| \cdot \|_p \) is a norm on \( \mathbb{C}^n \) for all \( 1 \leq p < \infty \) but is not a norm (the triangle inequality fails) for \( 0 < p < 1 \).

Definition 6.19 (Balls). Let \((X, d)\) be a metric space. For \( x \in X \) and \( r \geq 0 \) let

\[
B_x(r) := \{ y \in X : d(x, y) < r \} \quad \text{and} \quad C_x(r) := \{ y \in X : d(x, y) \leq r \}.
\]

We refer to \( B_x(r) \) and \( C_x(r) \) as the open and closed ball respectively about \( x \) with radius \( r \).

Example 6.20. Let us consider \( \| \cdot \|_p \) to \( \mathbb{R}^2 \). Figure 6.1 shows the boundary of the balls of radius 1 centered at 0 for some of the different \( p \) – norms.

![Fig. 6.1. Balls in \( \mathbb{R}^2 \) corresponding to \( p \in \{ \frac{1}{2}, 1, 2, 5, 20 \} \). The case \( p = \frac{1}{2} \) is not convex like the other balls which is an indication that the triangle inequality fails.](image)

Example 6.21. The next figures explain how to understand balls in \( C([0, 1], \mathbb{R}) \) equipped with the uniform norm.
Exercise 6.2 (Weighted 2–norms). Suppose that $\rho_i \in (0, \infty)$ for $1 \leq i \leq n$ and for $a, b \in \mathbb{C}^n$ let

$$a \ast b := \sum_{i=1}^{n} a_i b_i \rho_i \quad \text{and} \quad \|a\| := \sqrt{a \ast \bar{a}} = \sqrt{\sum_{i=1}^{n} |a_i|^2 \rho_i}.$$ 

Show, for all $a, b \in \mathbb{C}^n$ that $|a \ast b| \leq \|a\| \cdot \|b\|$ and that $\|\cdot\|$ is a norm on $\mathbb{C}^n$. \[\text{Hint: reduce to the case where } \rho_i = 1 \text{ for all } i.\]

For the next two exercise you will be using some concepts from calculus which we will developed in detail next quarter. For now, I assume you know what the Riemann integral is for continuous functions on $[0, 1]$ with values in $\mathbb{R}$. Let $Z$ denote the continuous functions on $[0, 1]$ with values in $\mathbb{R}$. The only properties that you need to know about the Riemann integral are;

1. The integral is linear, namely for all $f, g \in Z$ and $\lambda \in \mathbb{R}$,

$$\int_{0}^{1} (f(t) + \lambda g(t)) \, dt = \int_{0}^{1} f(t) \, dt + \lambda \int_{0}^{1} g(t) \, dt.$$ 

2. If $f, g \in Z$ and $f(t) \leq g(t)$ for all $t \in [0, 1]$, then

$$\int_{0}^{1} f(t) \, dt \leq \int_{0}^{1} g(t) \, dt.$$ 

3. For all $f \in Z$,

$$\left| \int_{0}^{1} f(t) \, dt \right| \leq \int_{0}^{1} |f(t)| \, dt.$$ 

In fact this item follows from items 1. and 2. Indeed, since $\pm f(t) \leq |f(t)|$ for all $t$, we find

$$\pm \int_{0}^{1} f(t) \, dt = \int_{0}^{1} \pm f(t) \, dt \leq \int_{0}^{1} |f(t)| \, dt \iff \int_{0}^{1} f(t) \, dt \leq \int_{0}^{1} |f(t)| \, dt.$$ 

The notion of continuity will be formally developed shortly.
Exercise 6.3. Let $Z$ denote the continuous function on $[0, 1]$ with values in $\mathbb{R}$ and for $f \in Z$ let

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

Show $\|\cdot\|_1$ satisfies,

1. (Homogeneity) $\|\lambda f\| = |\lambda| \|f\|$ for all $\lambda \in \mathbb{R}$ and $f \in Z$.
2. (Triangle inequality) $\|f + g\| \leq \|f\| + \|g\|$ for all $f, g \in Z$.

Remark 6.23 (An interpretation of $\|\cdot\|_1$). If $f(t)$ as the speed of a particle on the real line at time $t$, then $\|f\|_1$ represents the total distance (including retracing of its path) the particle travels over the time interval $[0, 1]$.

Exercise 6.4. Let $Z$ denote the continuous function on $[0, 1]$ with values in $\mathbb{R}$ and for $f \in Z$ let

$$\|f\|_2 := \sqrt{\int_0^1 |f(t)|^2 \, dt}.$$ 

Show;

1. for $f, g \in Z$ that

$$\int_0^1 f(t) g(t) \, dt \leq \|f\|_2 \cdot \|g\|_2,$$

2. Homogeneity) $\|\lambda f\|_2 = |\lambda| \|f\|_2$ for all $\lambda \in \mathbb{R}$ and $f \in Z$, and
3. (Triangle inequality) $\|f + g\|_2 \leq \|f\|_2 + \|g\|_2$ for all $f, g \in Z$.

Remark 6.23 (An interpretation of $\|\cdot\|_2$). Let us suppose that $f(t)$ is the voltage across a 1 Ohm resistor. By Ohms law the current through this resistor is $f(t)/1 = f(t)$ and the power dissipated by the resistor at time $t$ is (Voltage-Current) is $f(t)^2$. The work done over the time interval, $[0, 1]$ is then

$$\int_0^1 \text{Power}(t) \, dt = \int_0^1 f^2(t) \, dt = \|f\|_2^2.$$ 

On the other hand if we had a constant voltage of $\|f\|_2$ across the resistor over the time interval $[0, 1]$, the work done over the period would again be $\|f\|_2^2$. Thus $\|f\|_2$ is often referred to as the RMS voltage (root mean squared voltage) and represents the equivalent DC (Direct Current, i.e. constant voltage necessary to produce the same amount of work over the time interval $[0, 1]$.

Exercise 6.5. Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ such that $\|a\| \leq \|b\|$ whenever $0 \leq a_i \leq b_i$ for $1 \leq i \leq n$.\footnote{For example we could take $\|\cdot\|$ to be $\|\cdot\|_1$, or $\|\cdot\|_2$ on $\mathbb{R}^n$. Not all norms satisfy this assumption though. For example, take $\|(x, y)\| = |x| + |y|$ on $\mathbb{R}^2$, then $\|(x, y)\|$ is decreasing in $x$ when $0 \leq x < y$.} Further suppose that $(X_i, d_i)$ for $i = 1, \ldots, n$ is a finite collection of metric spaces and for $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $X := \prod_{i=1}^n X_i$, let

$$d(x, y) = \|(d_1(x_1, y_1), d_2(x_2, y_2), \ldots, d_n(x_n, y_n))\|.$$ 

Show $(X, d)$ is a metric space.

• End of Lecture 13, 10/28/2012. [We started Section 5.1 above as well this day.]

6.2 Sequences in Metric Spaces

Definition 6.24. 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 $\lim_{n \to \infty} x_n = x$ or $x_n \to x$ as $n \to \infty$.

Exercise 6.6. Show that $x$ in Definition 6.24 is necessarily unique.

Definition 6.25 (Cauchy sequences). A sequence $\{x_n\}_{n=1}^\infty$ in a metric space $(X, d)$ is Cauchy provided that $\lim_{n,m \to \infty} d(x_n, x_m) = 0$, i.e. for all $\varepsilon > 0$ there exists $N(\varepsilon) \in \mathbb{N}$ such that

$$d(x_n, x_m) \leq \varepsilon \text{ when } m, n \geq N(\varepsilon).$$

• End of Lecture 15, 11/2/2012.

Exercise 6.7. Show that convergent sequences in metric spaces 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. Of course the sequence is convergent in $\mathbb{R}$.

Exercise 6.8. If $\{x_n\}_{n=1}^\infty$ is a Cauchy sequence in a metric space $(X, d)$, $\lim_{n \to \infty} d(x_n, y)$ exists in $\mathbb{R}$ for all $y \in X$. In particular, $\{d(x_n, y)\}_{n=1}^\infty$ is a bounded sequence in $\mathbb{R}$ for all $y \in X$.

Definition 6.26. A metric space $(X, d)$ (or normed space $(X, \|\cdot\|)$) is complete if all Cauchy sequences are convergent sequences. A complete normed space is called a Banach space.

Lemma 6.27. Let $X$ be a non-empty set and

$$\|f\|_u := \sup_{x \in X} |f(x)| \text{ for all } f \in C^X.$$ 

Then the subspace, $Z := \{ f \in C^X : \|f\|_u < \infty \}$ is a Banach space, i.e. $(Z, \|\cdot\|_u)$ is a complete normed space.
Exercise 6.10 (Equivalence of 3 norms on $\mathbb{C}^n$). Let $\|\cdot\|_1$, $\|\cdot\|_u$, and $\|\cdot\|_2$ be the three norms on $\mathbb{C}^n$ given above. Show for all $z \in \mathbb{C}^n$ that
\[
\|z\|_u \leq \|z\|_1 \leq n \|z\|_u, \\
\|z\|_1 \leq \sqrt{n} \|z\|_2 \quad \text{(Hint: use Cauchy Schwarz.)} \\
\|z\|_2 \leq \sqrt{n} \|z\|_1 \leq \|z\|_1 \quad \text{and} \\
\|z\|_2 \leq \sqrt{n} \|z\|_u.
\]
It follows from these inequalities that $\|\cdot\|_1$, $\|\cdot\|_u$, and $\|\cdot\|_2$ are equivalent norms on $\mathbb{C}^n$.

Theorem 6.30 (Completeness of $\mathbb{C}^n$). Let $n \in \mathbb{N}$ and $\|\cdot\|$ denote any one of the norms, $\|\cdot\|_1$, $\|\cdot\|_2$, or $\|\cdot\|_u$ on $\mathbb{C}^n$. Then $(\mathbb{C}^n, \|\cdot\|)$ is complete.

Proof. By Exercise 6.10 all of these norms are equivalent to $\|\cdot\|_u$ and hence it suffices to show that $\|\cdot\|_u$ is a complete norm on $\mathbb{C}^n$. This is a special case of Lemma 6.27 with $X = \{1, 2, \ldots, n\}$.

Exercise 6.11. Let $X$ be a set and $(Y, \rho)$ be a complete metric space. Suppose that $f_n : X \to Y$ are functions such that
\[
\delta_{m,n} := \sup_{x \in X} d(f_n(x), f_m(x)) \to 0 \quad \text{as} \quad m, n \to \infty.
\]
Show there exists a (unique) function $f : X \to Y$ such that
\[
\lim_{n \to \infty} \sup_{x \in X} d(f_n(x), f(x)) = 0.
\]

Hint: mimic the proof of Lemma 6.27

Exercise 6.12. Let $Z$ denote the continuous functions on $[0, 1]$ with values in $\mathbb{R}$ and as above let
\[
\|f\|_1 := \int_0^1 |f(t)| \, dt, \\
\|f\|_2 := \sqrt{\int_0^1 |f(t)|^2 \, dt}, \quad \text{and} \quad \|f\|_u = \sup_{0 \leq t \leq 1} |f(t)|.
\]
Show for all $f \in Z$ that;
\[
\|f\|_1 \leq \|f\|_2 \quad \text{and} \quad \|f\|_2 \leq \|f\|_u.
\]

[Hint: for the first inequality use Cauchy Schwarz.] Also show there is no constant $C < \infty$ such that
\[
\|f\|_u \leq C \|f\|_2 \quad \text{for all} \quad f \in Z.
\]

[Hint: consider the sequence, $f_n(t) = t^n$.]
Example 6.31. Let $Z = C([0, 1], \mathbb{R})$ be the vector space of continuous functions on $[0, 1]$ with values in $\mathbb{R}$ and for $f \in Z$ let

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

Let us show that $(Z, \|\cdot\|_1)$ is not complete. To this end let

$$g(t) := \begin{cases} 2t & \text{if } t \leq 1/2 \\ 1 & \text{if } 1/2 \leq t \leq 1 \end{cases}$$

and then set $f_n(t) = g(t)^n$ or all $t \in [0, 1]$. Then

$$\lim_{n \to \infty} f_n(t) = h(t) = \begin{cases} 0 & \text{if } t < \frac{1}{2} \\ 1 & \text{if } t \geq \frac{1}{2} \end{cases}$$

which is discontinuous at $1/2$. Let us now observe that $\|h\|_1 = \frac{1}{2}$ and

$$\|f_n - h\|_1 = \int_0^{1/2} (2t)^n \, dt = \frac{1}{2(n+1)}$$

so that $f_n \to h$ in $\|\cdot\|_1$. If there were some $f \in Z$ so that $\|f - f_n\|_1 \to 0$ we would have to have $\|f - h\|_1 = 0$. If $\varepsilon := |f(t_0) - h(t_0)|$ for some $t_0 \neq \frac{1}{2}$, by continuity we would have $|f(t) - h(t)| \geq \varepsilon/2$ for $t$ near $t_0$ from which it would follow that $\|f - h\|_1 > 0$. Therefore we must have $f(t) = h(t)$ for all $t \neq t_0$. Since

there is not choice for $f(1/2)$ for which $f$ would be continuous at $1/2$. Hence we have shown that $f_n$ can not converge to any element, $f \in Z$.

In fact, $(Z, \|\cdot\|_1)$ is full of uncountably many “holes” and $h$ is the location of just one of these holes. In the third quarter of this course we will fill these holes.

6.3 General Limits and Continuity in Metric Spaces

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

Definition 6.32 (Limits of functions). If $x_0 \in X$ and $f : X \setminus \{x_0\} \to Y$ is a function, then we say $\lim_{x \to x_0} f(x) = y_0 \in Y$ iff for all $\varepsilon > 0$ there is a $\delta = \delta(\varepsilon, x_0) > 0$ such that

$$d(f(x), y_0) \leq \varepsilon \text{ provided that } 0 < \rho(x, x_0) \leq \delta(\varepsilon, x_0).$$

(6.4)

[In generally when we write $\lim_{x \to x_0} f(x)$ we do not need to assume that $f(x_0)$ is defined.]

Theorem 6.33 (Computing Limits Using Sequences). If $x_0 \in X$ and $f : X \setminus \{x_0\} \to Y$ is a function as above, then $\lim_{x \to x_0} f(x) = y_0 \in Y$ iff $\lim_{n \to \infty} f(x_n) = y_0$ for all sequences $\{x_n\}_{n=1}^{\infty} \subset X \setminus \{x_0\}$ such that $\lim_{n \to \infty} x_n = x_0$. 

$$\lim_{t \uparrow \frac{1}{2}} f(t) = \lim_{t \downarrow \frac{1}{2}} h(t) = 1$$

$$\lim_{t \uparrow \frac{1}{2}} f(t) = \lim_{t \downarrow \frac{1}{2}} f(t) = 0$$

If there were some $f \in Z$ so that $\lim_{n \to \infty} f(x_n) = y_0$ for all sequences $\{x_n\}_{n=1}^{\infty} \subset X \setminus \{x_0\}$ such that $\lim_{n \to \infty} x_n = x_0$. 

Fig. 6.4. Here is a plot of $f_1$, $f_2$, $f_3$, $f_5$, $f_{32}$, and the pointwise limit function $h$. 

\[\text{Fig. 6.4. Here is a plot of } f_1, f_2, f_3, f_5, f_{32}, \text{ and the pointwise limit function } h.\]
We say \( \lim_{x \to x_0} f(x) = y_0 \in Y \) and \( \{x_n\}_{n=1}^\infty \subset X \setminus \{x_0\} \) with \( \lim_{n \to \infty} x_n = x_0 \). Then all \( \varepsilon > 0 \) there is a \( \delta > 0 \) such that Eq. (6.4) holds.

Since \( x_n \to x_0 \) as \( n \to \infty \) it follows that \( 0 < \rho(x_n, x_0) \leq \delta \) for a.a. \( n \). As \( \varepsilon > 0 \) was arbitrary, it follows that \( \lim_{n \to \infty} d\left(f(x_n), y_0\right) = 0 \), i.e. that \( \lim_{n \to \infty} f(x_n) = y_0 \).

Conversely, if \( \lim_{x \to x_0} f(x) \neq y_0 \), then for all \( \varepsilon > 0 \) there exists \( \delta = \frac{\varepsilon}{n} > 0 \) such that for any \( \delta = \frac{1}{n} > 0 \) there exists \( x_n \in X \setminus \{x_0\} \) such that \( d(f(x_n), y_0) \geq \varepsilon \) while \( \rho(x, x_n) < \frac{\varepsilon}{n} \). We then have \( \lim_{n \to \infty} x_n = x \) while \( \lim_{n \to \infty} f(x_n) \neq y_0 \).

**Definition 6.34 (Continuity).** A function \( f : X \to Y \) is **continuous at** \( x \in X \) if for all \( \varepsilon > 0 \) there is a \( \delta > 0 \) such that

\[
d(f(x), f(x')) < \varepsilon \quad \text{provided that} \quad \rho(x, x') < \delta.
\]

(6.5)
The function \( f \) is said to be **continuous** if \( f \) is continuous at all points \( x \in X \). We will write \( C(X, Y) \) for the collection of continuous functions from \( X \) to \( Y \).

**Definition 6.35 (Sequential Continuity).** A function \( f : X \to Y \) is **continuous** at \( x \in X \) if

\[
\lim_{n \to \infty} f(x_n) = f(x) \quad \text{for all} \quad \{x_n\}_{n=1}^\infty \subset X \quad \text{with} \quad \lim_{n \to \infty} x_n = x.
\]

We say \( f \) is sequentially continuous on \( X \) if it is continuous at all points in \( X \).

**Example 6.37.** The functions \( f : \mathbb{C} \setminus \{0\} \to \mathbb{C} \) defined by \( f(z) = 1/z \) is continuous. Indeed, if \( \{z_n\}_{n=1}^\infty \subset \mathbb{C} \setminus \{0\} \) and \( \lim_{n \to \infty} z_n = z \in \mathbb{C} \setminus \{0\} \), then

\[
\lim_{n \to \infty} f(z_n) = \lim_{n \to \infty} \frac{1}{z_n} = \frac{1}{z} = f(z).
\]

**Example 6.38.** Let \( f : \mathbb{R} \to \mathbb{R} \) be the function defined by

\[
f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q}. \end{cases}
\]

The function \( f \) is discontinuous at all points in \( \mathbb{R} \). For example, if \( x_0 \in \mathbb{Q} \) we may choose \( x_n \in \mathbb{R} \setminus \mathbb{Q} \) such that \( \lim_{n \to \infty} x_n = x_0 \) while

\[
\lim_{n \to \infty} f(x_n) = \lim_{n \to \infty} 0 = 0 \neq 1 = f(x_0).
\]

Similarly if \( x_0 \in \mathbb{R} \setminus \mathbb{Q} \) we may choose \( x_n \in \mathbb{Q} \) such that \( \lim_{n \to \infty} x_n = x_0 \) while

\[
\lim_{n \to \infty} f(x_n) = \lim_{n \to \infty} 1 = 1 \neq 0 = f(x_0).
\]

**Exercise 6.13.** Consider \( \mathbb{N} \) as a metric space with \( d(m, n) := |m - n| \) and suppose that \( (Y, d) \) is a metric space. Show that every function, \( f : \mathbb{N} \to Y \) is continuous.

**Exercise 6.14.** Suppose that \( (X, d) \) is a metric space and \( f, g : X \to \mathbb{C} \) are two continuous functions on \( X \). Show;

1. \( f + g \) is continuous,
2. \( f \cdot g \) is continuous,
3. \( f/g \) is continuous provided \( g(x) \neq 0 \) for all \( x \in X \).

**Exercise 6.15.** Show the following functions from \( \mathbb{C} \) to \( \mathbb{C} \) are continuous.

1. \( f(z) = c \) for all \( z \in \mathbb{C} \) where \( c \in \mathbb{C} \) is a constant.
2. \( f(z) = |z| \).
3. \( f(z) = z \) if and \( f(z) = \bar{z} \).
4. \( f(z) = \text{Re}z \) and \( f(z) = \text{Im}z \).
5. \( f(z) = \sum_{m,n=0}^{N} a_{m,n} z^m \bar{z}^n \) where \( a_{m,n} \in \mathbb{C} \).

**Exercise 6.16.** Suppose now that \( (X, \rho) \), \( (Y, d) \), and \( (Z, \delta) \) are three metric spaces and \( f : X \to Y \) and \( g : Y \to Z \). Let \( x \in X \) and \( y = f(x) \in Y \), show \( g \circ f : X \to Z \) is continuous at \( x \) if \( f \) is continuous at \( x \) and \( g \) is continuous at \( y \). Recall that \( (g \circ f)(x) := g(f(x)) \) for all \( x \in X \). In particular this implies that if \( f \) is continuous on \( X \) and \( g \) is continuous on \( Y \), then \( f \circ g \) is continuous on \( X \).

**Corollary 6.36.** Continuity and sequential continuity are the same notions.

**Proof.** This follows rather directly from Theorem 6.33. □
Definition 6.40 (One sided limits). Suppose \((Y, d)\) is a metric space, \(-\infty < a < b < \infty\), and \(f : (a, b) \to Y\) is a function. For \(x_0 \in (a, b)\) we say
\[
\lim_{x \to x_0-} f(x) = y_0 \iff \forall \varepsilon > 0 \exists \delta = \delta(\varepsilon) > 0 \exists (f(x), y_0) \leq \varepsilon \text{ if } 0 < x - x_0 \leq \delta
\]
and
\[
\lim_{x \to x_0+} f(x) = y_0 \iff \forall \varepsilon > 0 \exists \delta = \delta(\varepsilon) > 0 \exists (f(x), y_0) \leq \varepsilon \text{ if } 0 < x - x_0 \leq \delta.
\]

Theorem 6.41 (One sided limit criteria). Suppose that \((Y, d)\) is a metric space, \((a, b) \subset \mathbb{R}, f : (a, b) \to Y\) is a function, and \(x_0 \in (a, b)\). Then the following are equivalent;
1. \(\lim_{x \to x_0-} f(x) = y_0\)
2. \(\lim_{n \to \infty} f(x_n) = y_0\) for all \(\{x_n\}_{n=1}^\infty \subset (a, b)\) such that \(\lim_{n \to \infty} x_n = x_0\).
3. \(\lim_{n \to \infty} f(x_n) = y_0\) for all \(\{x_n\}_{n=1}^\infty \subset (a, b)\) such that \(x_n \to x_0\), i.e. \(x_{n+1} \leq x_n\) and \(x_0 < x_n\) for all \(n\) and \(\lim_{n \to \infty} x_n = x_0\).

We also have the following equivalent statements;
1. \(\lim_{x \to x_0-} f(x) = y_0\)
2. \(\lim_{n \to \infty} f(x_n) = y_0\) for all \(\{x_n\}_{n=1}^\infty \subset (a, x_0)\) such that \(\lim_{n \to \infty} x_n = x_0\).
3. \(\lim_{n \to \infty} f(x_n) = y_0\) for all \(\{x_n\}_{n=1}^\infty \subset (a, x_0)\) such that \(x_n \uparrow x_0\), i.e. \(x_{n+1} \geq x_n\) and \(x_0 > x_n\) for all \(n\) and \(\lim_{n \to \infty} x_n = x_0\).

Moreover, \(\lim_{x \to x_0} f(x) = y_0\) iff \(\lim_{x \to x_0-} f(x) = \lim_{x \to x_0+} f(x) = y_0\).

Proof. (\(\Rightarrow \Rightarrow \)) If \(\lim_{x \to x_0-} f(x) = y_0\) then for all \(\varepsilon > 0\) there exists \(\delta = \delta(\varepsilon) > 0\) such that \(d(f(x), y_0) \leq \varepsilon\) if \(0 < x - x_0 \leq \delta\). Hence if \(\{x_n\}_{n=1}^\infty \subset (x_0, b)\) such that \(x_n \to x_0\) then \(0 < x_n - x_0 \leq \delta\) for a.a. \(n\) and hence \(d(f(x_n), y_0) \leq \varepsilon\) for a.a. \(n\). Since \(\varepsilon > 0\) was arbitrary, it follows that \(\lim_{n \to \infty} d(f(x_n), y_0) = 0\), i.e. \(\lim_{x \to x_0-} f(x) = y_0\). It is trivial that (2. \(\Rightarrow \))

(3. \(\Rightarrow \)) If \(\lim_{x \to x_0} f(x) \neq y_0\) there exists \(\varepsilon > 0\) such that for all \(\delta = \frac{1}{n} > 0\) there exists \(x_n \in (x_0, b)\) such that \(0 \leq x_n - x_0 \leq \frac{1}{n}\) while \(d(f(x_n), y_0) \geq \varepsilon\). Then take \(x_n = \min(x_1, \ldots, x_n)\). Then \(x_n \downarrow x\) while \(d(f(x_n), y_0) \geq \varepsilon\), i.e.,
\[
\lim_{n \to \infty} f(x_n) \neq y_0.
\]

Exercise 6.17 (Continuity of \(x^{1/m}\)). Show for each \(m \in \mathbb{N}\) that the function \(f(x) := x^{1/m}\) is continuous on \([0, \infty)\).

Exercise 6.18 (Differentiability of \(x^{1/m}\)). Show for each \(m \in \mathbb{N}\) that the function \(f(x) := x^{1/m}\) is differentiable on \((0, \infty)\) and that
\[
\frac{d}{dx} x^{1/m} = \frac{y^{1/m} - x^{1/m}}{y - x} = \frac{1}{m} x^{1/m-1}.
\]

Exercise 6.19 (Intermediate value theorem). Suppose that \(-\infty < a < b < \infty\) and \(f : [a, b] \to \mathbb{R}\) is a continuous function such that \(f(a) \leq f(b)\). Show for any \(y \in [f(a), f(b)]\), there exists a \(c \in [a, b]\) such that \(f(c) = y\) hint: Let \(S := \{t \in [a, b] : f(t) \leq y\}\) and let \(c := \sup(S)\).

Exercise 6.20 (Inverse Function Theorem I). Let \(f : [a, b] \to [c, d]\) be a strictly increasing (i.e. \(f(x_1) < f(x_2)\)) whenever \(x_1 < x_2\) continuous function such that \(f(a) = c\) and \(f(b) = d\). Then \(f\) is bijective and the inverse function, \(g := f^{-1} : [c, d] \to [a, b]\), is strictly increasing and is continuous.

Notations 6.43 Let \((X, \rho)\) and \((Y, d)\) be metric spaces and \(f : X \to Y\) be a function.

\[3\] The same result holds for \(y \in [f(b), f(a)]\) if \(f(b) \leq f(a)\) just replace \(f\) by \(-f\) in this case.
We say \( f \) is uniformly continuous, iff for all \( \varepsilon > 0 \) there exists \( \delta = \delta(\varepsilon) > 0 \) such that
\[
\forall x, x' \in X \text{ with } \rho(x, x') \leq \delta \implies d(f(x), f(x')) \leq \varepsilon.
\]

2. A function, \( f : X \to Y \), is said to be Lipschitz if there is a constant \( C < \infty \) such that
\[
d(f(x), f(x')) \leq C \rho(x, x') \text{ for all } x, x' \in X.
\]

Recall that a function \( f : X \to Y \) is continuous at \( x_0 \in X \) if for all \( \varepsilon > 0 \) there exists \( \delta = \delta(\varepsilon, x_0) > 0 \) such that
\[
\forall x \in X \text{ with } \rho(x, x_0) \leq \delta \implies d(f(x), f(x_0)) \leq \varepsilon.
\]
Thus we see that a function is uniformly provided we can take \( \delta(\varepsilon, x_0) \) to be independent of \( x_0 \). If \( f \) is Lipschitz and \( \varepsilon > 0 \), we may take \( \delta := \varepsilon/C \) in order to see that if
\[
\rho(x, x') \leq \delta \implies d(f(x), f(x')) \leq C \rho(x, x') \leq C \delta = \varepsilon
\]
which shows \( f \) is uniformly continuous.

**Example 6.44.** Any function, \( f : \mathbb{R} \to \mathbb{R} \) which is everywhere differentiable is Lipschitz iff \( K := \sup_{t \in \mathbb{R}} |f'(t)| < \infty \). Indeed if
\[
|f(y) - f(x)| \leq K |y - x|
\]
then
\[
|f'(x)| = \lim_{y \to x} \frac{|f(y) - f(x)|}{|y - x|} \leq K \text{ for all } x \in \mathbb{R}.
\]
Conversely, if \( K := \sup_{t \in \mathbb{R}} |f'(t)| < \infty \), then by the mean value Theorem \( 10.16 \) for all \( y > x \) there exists \( c \in (x, y) \) such that
\[
\left| \frac{f(y) - f(x)}{y - x} \right| = |f'(c)| \leq K.
\]

It turns out that every metric spaces with an infinite number of elements comes equipped with a large collection of Lipschitz functions.

**Lemma 6.45 (Distance to a Set).** For any non empty subset \( A \subset X \), let
\[
d_A(x) := \inf\{d(x, a) | a \in A\},
\]
then
\[
|d_A(x) - d_A(y)| \leq d(x, y) \quad \forall x, y \in X.
\]  
(6.6)
In particular, \( d_A : X \to [0, \infty) \) is continuous.

**Proof.** Let \( a \in A \) and \( x, y \in X \), then
\[
d_A(x) \leq d(x, a) \leq d(x, y) + d(y, a).
\]
Take the infimum over \( a \) in the above equation shows that
\[
d_A(x) \leq d(x, y) + d_A(y) \quad \forall x, y \in X.
\]
Therefore, \( d_A(x) - d_A(y) \leq d(x, y) \) and by interchanging \( x \) and \( y \) we also have that \( d_A(y) - d_A(x) \leq d(x, y) \) which implies Eq. (6.6).

**Corollary 6.46.** The function \( d \) satisfies,
\[
|d(x, y) - d(x', y')| \leq d(y, y') + d(x, x').
\]
Therefore \( d : X \times X \to [0, \infty) \) 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 \( y' \). In particular, if \( x_n \to x \) and \( y_n \to y \) then
\[
\lim_{n \to \infty} d(x_n, y_n) = d(x, y) = d\left(\lim_{n \to \infty} x_n, \lim_{n \to \infty} y_n\right).
\]

**Proof. First Proof.** By Lemma 6.45 for single point sets and the triangle inequality for the absolute value of real numbers,
\[
|d(x, y) - d(x', y')| \leq |d(x, y) - d(x, y')| + |d(x, y') - d(x', y')| \leq d(y, y') + d(x, x').
\]

**Second Proof.** By the triangle inequality,
\[
d(x, y) \leq d(x, x') + d(x', y) \leq d(x, x') + d(x', y') + d(y', y)
\]
from which it follows that
\[
d(x, y) - d(x', y') \leq d(x, x') + d(y', y).
\]
Interchanging \( x \) with \( x' \) and \( y \) with \( y' \) in this inequality shows,
\[
d(x', y') - d(x, y) \leq d(x, x') + d(y, y')
\]
and the result follows from the last two inequalities.

**Exercise 6.21 (Continuity of integration).** Let \( Z = C([0, 1], \mathbb{R}) \) be the continuous functions from \([0, 1]\) to \( \mathbb{R} \) and \( \|\cdot\|_u \) be the uniform norm, \( \|f\|_u := \sup_{0 \leq t \leq 1} |f(t)| \). Define \( K : Z \to Z \) by
\[
K(f)(x) := \int_0^x f(t) \, dt \text{ for all } x \in [0, 1].
\]
Show that $K$ is a Lipschitz function. In more detail, show
\[ \|K(f) - K(g)\|_u \leq \|f - g\|_u \text{ for all } f, g \in Z. \]

In this problem please take for granted the standard properties of the integral including

1. The function $x \to K(f)(x)$ is indeed continuous (in fact differentiable by the fundamental theorem of calculus).
2. $K : Z \to Z$ is a linear transformation.
3. If $f(t) \leq g(t)$ for all $t \in [0, 1]$, then $\int_0^x f(t) \, dt \leq \int_0^x g(t) \, dt$ for all $x \in X$.

Exercise 6.22 (Discontinuity of differentiation). Let $Z$ be the polynomial functions in $C([0,1], \mathbb{R})$, i.e. functions of the form $p(t) = \sum_{k=0}^n a_k t^k$ with $a_k \in \mathbb{R}$. As above we let $\|p\|_u := \sup_{0 \leq t \leq 1} |p(t)|$. Define $D : Z \to Z$ by $D(p) = p'$, i.e. if $p(t) = \sum_{k=0}^n a_k t^k$ then
\[ D(p)(t) = \sum_{k=1}^n k a_k t^{k-1}. \]

1. Show $D$ is discontinuous at 0 – where 0 represents the zero polynomial.
2. Show $D$ is discontinuous at all points $p \in Z$.

Exercise 6.23 (Continuity of integration II). Let $Z = C([0,1], \mathbb{R})$ and $K$ be as in Exercise 6.21 Further let

\[ \|f\|_2 := \sqrt{\int_0^1 |f(t)|^2 \, dt}. \]

Show
\[ \|K(f) - K(g)\|_2 \leq \frac{1}{\sqrt{2}} \|f - g\|_2 \text{ for all } f, g \in Z. \]

Definition 6.47 (Pointwise Convergence). Let $(X, d)$ and $(Y, \rho)$ be metric spaces and $f_n : X \to Y$ be functions for each $n \in \mathbb{N}$. We say that $f_n$ converges pointwise to $f : X \to Y$ provided $\lim_{n \to \infty} f_n(x) = f(x)$ for all $x \in X$, i.e. provided
\[ \lim_{n \to \infty} d(f(x), f_n(x)) = 0 \text{ for each } x \in X. \]

Definition 6.48 (Uniform Convergence). Let $(X, d)$ and $(Y, \rho)$ be metric spaces and $f_n : X \to Y$ be functions for each $n \in \mathbb{N}$. We say that $f_n$ converges uniformly to $f : X \to Y$ provided
\[ \delta_n := \sup_{x \in X} d(f(x), f_n(x)) \to 0 \text{ as } n \to \infty. \]

Fig. 6.5. Here is a sequence of continuous functions converging pointwise to a discontinuous function. This convergence is not uniform.

Hopefully it is clear to the reader the uniform convergence implies pointwise convergence. The next theorem is a basic fact about uniform convergence which does not hold in general for pointwise convergence. The proof of this theorem is a bit subtle but well worth mastering as the method will arise over and over again.

Theorem 6.49 (Uniform Convergence Preserves Continuity). Suppose that $(f_n)_{n=1}^\infty$ are continuous functions from $X$ to $Y$ and $f_n$ converges uniformly to $f : X \to Y$. If $f_n$ is continuous at $x \in X$ for all $n$ then $f$ is continuous at $x$ as well. In particular if $f_n$ is continuous on $X$ for all $n$ then $f$ is continuous on $X$ as well.

Proof. We will give three proofs of this important theorem. In these proofs we will let
\[ \delta_n := \sup_{x \in X} d(f(x), f_n(x)). \]

First Proof. Suppose that $f$ were discontinuous at some point $x_0 \in X$. Then there would exists $\varepsilon > 0$ and $x_k \in X \setminus \{x_0\}$ such that $\lim_{k \to \infty} x_k = x_0$ while $\rho(f(x_k), f(x_0)) \geq \varepsilon$ for all $x > 0$. Let $n \in \mathbb{N}$ and set $g := f_n$, then
\[ \varepsilon \leq \rho(f(x_k), f(x_0)) \leq \rho(f(x_k), g(x_k)) + \rho(g(x_k), g(x_0)) + \rho(g(x_0), f(x_0)) \leq \delta_n + \rho(g(x_k), g(x_0)) + \delta_n = 2\delta_n. \]

Letting $k \to \infty$ in this inequality implies, $\varepsilon \leq 2\delta_n$ and then letting $n \to \infty$ implies $\varepsilon = 0$ and we have reached the desired contradiction, see Figure 6.6.
Second Proof. We must show \( \lim_{k \to \infty} f(x_k) = f(x) \) whenever \( \{x_k\}_{k=1}^\infty \subset X \) is a convergent sequence such that \( x := \lim_k x_k \in X \). So assume we are given such a sequence \( \{x_k\}_{k=1}^\infty \). Then for any \( n \in \mathbb{N} \) we have,

\[
\rho(f,x_k) \leq \rho(f,x_n) + \rho(x_n,x_k) \leq \rho(f,x_n) + \rho(f,x) + \rho(x,x_k) + \delta_n.
\]

wherein we have used the continuity of \( f \). This suffices to show \( \lim\sup \rho(f,x_k) \).

Third Proof. Let \( x \in X \) and \( \varepsilon > 0 \) be given. Choose \( n \in \mathbb{N} \) so that \( \delta_n \leq \varepsilon \) and let \( g := f_n \). Since \( g \) is continuous there exists \( \delta > 0 \) such that \( \rho(g,x) \leq \varepsilon \) whenever \( d(x,x') \leq \delta \). So if \( d(x,x') \leq \delta \), then

\[
\rho(f(x),f(x')) \leq \rho(f(x),g(x)) + \rho(g(x),f(x')) \leq \rho(f(x),g(x)) + \rho(g(x),g(x')) + \rho(g(x'),f(x')) \leq \delta_n + \rho(g(x),g(x')) + \delta_n \leq \varepsilon + \varepsilon + 3\varepsilon = 3\varepsilon.
\]

As \( \varepsilon > 0 \) and \( x \in X \) were arbitrary, we have shown \( f \) is continuous on \( X \). 

\( \blacksquare \)

**Example 6.50 (Non-uniform convergence).** For an example of nonuniform convergence, suppose that \( g(x) = \max(1 - 4x^2, 0) \) and \( f_n(x) := g(x - 3n) \) for all \( n \), see Figure 6.7. Notice that for each \( x \in \mathbb{R}_+ \), \( f_n(x) = 0 \) for all \( n \) and therefore \( \lim_{n \to \infty} f_n(x) = 0 \). On the other hand, \( \|f_n\|_a = 1 \) for all \( n \) so \( f_n \) converges to \( 0 \) pointwise but not uniformly in \( x \). Nevertheless the limiting function is still continuous. This is not always the case as you will see in the next exercise.

**Exercise 6.24.** Let \( f_n : [0,1] \to \mathbb{R} \) be defined by \( f_n(x) = x^n \) for \( x \in [0,1] \). Show \( f(x) := \lim_{n \to \infty} f_n(x) \) exists for all \( x \in [0,1] \) and find \( f \) explicitly. Show that \( f_n \) does not converge to \( f \) uniformly.

### 6.4 Density and Separability

**Definition 6.51.** Let \( (X,d) \) be a metric space. We say \( A \subset X \) is **dense** in \( X \) if for all \( x \in X \), there exists \( \{x_n\}_{n=1}^\infty \subset A \) such that \( x = \lim_{n \to \infty} x_n \). [In words, all points in \( X \) are limit points of sequences in \( A \).] A metric space is said to be **separable** if it contains a countable dense subset, \( A \).

**Example 6.52.** The spaces \( \mathbb{R}^n \) and \( \mathbb{C}^n \) with their Euclidean metrics are separable. Indeed we can take \( D = \mathbb{Q}^n \) and \( D = (\mathbb{Q} + i\mathbb{Q})^n \) respectively for the
countable dense subsets. For example given $k \in \mathbb{N}$ and $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we may choose $q^k = (q_1^k, \ldots, q_n^k) \in D$ such that $|x_i - q_i^k| \leq \frac{1}{k}$ for $1 \leq i \leq n$. We then have,

$$d(x,q^k) = \left( \sum_{i=1}^{n} |x_i - q_i^k|^2 \right)^{1/2} \leq \left( \sum_{i=1}^{n} \left( \frac{1}{k} \right)^2 \right)^{1/2} = \sqrt{n} \cdot \frac{1}{k} \to 0 \quad \text{as} \quad k \to \infty.$$ 

The $\mathbb{C}^n$ case now follows from this case as $\mathbb{C}^n$ is really $\mathbb{R}^{2n}$ in disguise.

**Example 6.53.** Let $Y := \mathbb{R} \setminus \mathbb{Q}$ which we equip with the usual metric, $d(y,y') = |y - y'|$ for all $y, y' \in Y$. I now claim that $Y$ is separable. We can no longer use $\mathbb{Q}$ as the countable dense subset of $Y$ since $\mathbb{Q}$ is not contained in $Y$! On the other hand, for each $q \in \mathbb{Q}$ we may choose $y_n(q) \in Y$ such that $\lim_{n \to \infty} y_n(q) = q$. Then if $y \in Y$ and $\varepsilon = \frac{1}{n} > 0$ is given, we may choose $q \in \mathbb{Q}$ such that $|y - q| \leq \frac{1}{n}$. We then take $a_k := y_n(q)$ for some large $n$ so that $|a_k - q| \leq \frac{1}{n}$. It then follows that $|y - a_k| \leq \frac{1}{k} \to 0$ as $k \to \infty$ and this shows that $A := \bigcup_{q \in \mathbb{Q}} \{ y_n(q) : n \in \mathbb{N} \}$ is a dense subset of $Y$. Moreover $A$ is countable, why?

**Remark 6.54.** An equivalent way to say that $A \subset X$ is dense is to say $d_A \equiv 0$, i.e. $d_A(x) = 0$ for all $x \in X$. Indeed if $x \in X$, you should show that $d_A(x) = 0$ iff there exists $a_n \in A$ such that $d(x,a_n) \to 0$ as $n \to \infty$.

**Exercise 6.25.** Suppose that $(X,d)$ is a separable metric space and $Y$ is a non-empty subset of $X$ which is also a metric space by restricting $d$ to $Y$. Show $(Y,d)$ is separable. [Hint: suppose that $A \subset X$ be a countable dense subset of $X$. For each $a \in A$ choose $\{ y_n(a) \}_{n=1}^{\infty} \subset Y$ so that $d_Y(a) \leq d(a,y_n(a)) \leq d_X(a) + \frac{1}{n}$. Now show $A_Y := \bigcup_{a \in A} \{ y_n(a) : n \in \mathbb{N} \}$ is a countable dense subset of $Y$.]

**Exercise 6.26.** Let $n \in \mathbb{N}$. Show any non-empty subset $Y \subset \mathbb{C}^n$ equipped with the metric,

$$d(x,y) = \| y - x \|$$

is separable, where $\| \cdot \|$ is either $\| \cdot \|_1$, $\| \cdot \|_1'$, or $\| \cdot \|_2$.

**Exercise 6.27.** For $x, y \in \mathbb{R}$, let

$$d(x,y) := \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}.$$

Then $(\mathbb{R},d)$ is a non-separable metric space. (This metric space is also complete.)

**Exercise 6.28.** Suppose $(X,\rho)$ 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 $X$.
2. Now suppose that $(Y,d)$ is complete and $f : A \to Y$ is a function which is uniformly continuous (Notation 6.43). Recall this means for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$d(f(a), f(b)) \leq \varepsilon \quad \text{for all } a,b \in A \text{ with } \rho(a,b) \leq \delta. \quad (6.7)$$

Show there is a unique continuous function $F : X \to Y$ such that $F = f$ on $A$.

**Hint:** Define $F(x) = \lim_{n \to \infty} f(x_n)$ where $\{x_n\}_{n=1}^{\infty} \subset A$ is chosen to converge $x \in X$. You must show the limit exists and is independent of the choice of sequence $\{x_n\}_{n=1}^{\infty} \subset A$ which converges for $x$.

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 $\mathbb{R}$.

### 6.5 Test 2: Review Topics

1. Understand the basic properties of complex numbers.
2. **Countability.** Key facts are that countable union of countable sets is countable and the finite product of countable sets is countable.
3. Definitions of metric and normed spaces and their basic properties which in the end of the day typically follow from the triangle inequality.
4. Be aware of different norms, $\| \cdot \|_1$, $\| \cdot \|_1'$, and $\| \cdot \|_2$.
5. Understand the notion of: limits of sequences, Cauchy sequences, completeness, limits and continuity of functions.
6. Know what is meant by pointwise and uniform convergence. You should be able to compute pointwise limits and know how to test if the limit is uniform or not. A key theorem is the uniform limit of continuous functions is still continuous.
Series and Sums in Banach Spaces

Definition 7.1. Suppose \((X, \|\|)\) is a normed space and \(\{x_n\}_{n=1}^{\infty}\) is a sequence in \(X\). Then we say \(\sum_{n=1}^{\infty} x_n\) converges in \(X\) iff \(s := \lim_{N \to \infty} \sum_{n=1}^{N} x_n\) exists in \(X\) otherwise we say \(\sum_{n=1}^{\infty} x_n\) diverges. We often let \(S_N := \sum_{n=1}^{N} x_n\) and refer to \(\{S_N\}_{N=1}^{\infty} \subset X\) as the sequence of partial sums.

If \(X = \mathbb{R}\) and \(x_n \geq 0\), then \(\sum_{n=1}^{\infty} x_n\) diverges iff \(\lim_{N \to \infty} \sum_{n=1}^{N} x_n = \infty\) and so we will write \(\sum_{n=1}^{\infty} x_n = \infty\) in this case to indicate that \(\sum_{n=1}^{\infty} x_n\) diverges to infinity.

Theorem 7.2 (Comparison Theorem). Suppose that \(\{a_n\}_{n=1}^{\infty}\) and \(\{b_n\}_{n=1}^{\infty}\) are sequences in \([0, \infty)\). If \(a_n \leq b_n\) for all \(n\), then

\[
\sum_{n=1}^{\infty} a_n \leq \sum_{n=1}^{\infty} b_n
\]

where we allow for these sums to be infinite. Moreover if \(a_n \leq b_n\) for a.a. \(n\) then \(\sum_{n=1}^{\infty} b_n < \infty\) implies \(\sum_{n=1}^{\infty} a_n < \infty\) and if \(\sum_{n=1}^{\infty} a_n = \infty\) then \(\sum_{n=1}^{\infty} b_n = \infty\).

Proof. Let \(A_k := \sum_{n=1}^{k} a_n\) and \(B_k := \sum_{n=1}^{k} b_n\). Then a simple induction argument shows that \(A_k \leq B_k\) for all \(k\) and therefore

\[
\sum_{n=1}^{\infty} a_n = \lim_{k \to \infty} A_k \leq \lim_{k \to \infty} B_k = \sum_{n=1}^{\infty} b_n
\]

by the sandwich lemma.

Theorem 7.3 (Telescoping Series / Fundamental Theorem of Summation). Let \(\{f(n)\}_{n=1}^{\infty} \subset X\) be a sequence, then

\[
\sum_{n=1}^{N} [f(n + 1) - f(n)] = f(N + 1) - f(1) \quad \text{for all } N \in \mathbb{N} \quad (7.1)
\]

and \(\sum_{n=1}^{\infty} [f(n + 1) - f(n)]\) is convergent in \(X\) iff \(\lim_{N \to \infty} f(N)\) exists in \(X\) in which case,

\[
\sum_{n=1}^{\infty} (f(n + 1) - f(n)) = \lim_{N \to \infty} f(N) - f(1).
\]

We also have,

\[
\sum_{n=M}^{N} [f(n + 1) - f(n)] = f(N + 1) - f(M) \quad \text{for } N \geq M. \quad (7.2)
\]

Proof. When \(N = 3\) we have,

\[
\sum_{n=1}^{3} (f(n + 1) - f(n)) = (f(2) - f(1)) + (f(3) - f(2)) + (f(4) - f(3)) = f(4) - f(1).
\]

In general, Eqs. (7.1) and (7.2) are easily verified by a simple induction argument. The rest of the theorem is now evident.

Example 7.4 (Geometric Series). Suppose that \(f(n) = \alpha^n\) where \(\alpha \in \mathbb{C}\). Then

\[
f(n + 1) - f(n) = \alpha^{n+1} - \alpha^n = \alpha^n(\alpha - 1)
\]

and we find,

\[
(\alpha - 1) \sum_{n=1}^{N} \alpha^n = \alpha^{N+1} - \alpha.
\]

If \(\alpha \neq 1\) it follows that

\[
\sum_{n=1}^{N} \alpha^n = \frac{\alpha^{N+1} - \alpha}{\alpha - 1}
\]

and if \(|\alpha| < 1\), it follows that \(|\alpha^{N+1}| = |\alpha|^{N+1} \to 0\) as \(N \to \infty\) and therefore

\[
\sum_{n=1}^{\infty} \alpha^n = \frac{\alpha}{1 - \alpha}.
\]

Theorem 7.5 (Integral test). Let \(f : (0, \infty) \to \mathbb{R}\) be a \(C^1\) function such that \(f'(x) \geq 0\) and \(f'(x)\) is decreasing in \(x\) (i.e. \(f''(x) \leq 0\) if it exists) and \(f(\infty) := \lim_{x \to \infty} f(x)\), see Figure 7.7 (As \(f\) is an increasing function, \(f(\infty)\) exists in \((-\infty, \infty)\).) Then
\[ f(n + 1) - f(M) \leq \sum_{n=M}^{N} f'(n) \leq [f(n + 1) - f(M)] + f'(M) - f'(N + 1) \]

(7.3)

for all \( M \leq N \). Letting \( N \to \infty \) in these inequalities also gives,

\[ f(\infty) - f(M) \leq \sum_{n=M}^{\infty} f'(n) \leq f(\infty) - f(M) + f'(M) \]

(7.4)

and in particular, \( \sum_{n=1}^{\infty} f'(n) < \infty \) iff \( f(\infty) < \infty \).

Fig. 7.1. The items plotted here are \( \ln x \) and \( -1/\sqrt{x} \) both of which satisfy the hypothesis of the proposition.

**Proof.** By the mean value Theorem \[\text{[10.16]}\] there exists \( c_n \in (n - 1, n) \) such that \( f(n) - f(n - 1) = f'(c_n) \). Since \( f'(x) \) is decreasing, it follows that

\[ f'(n) \leq f'(c_n) \leq f'(n - 1) \]

for all \( n \), or equivalently\[\text{[10.16]}\]

\[ f'(n) \leq f(n) - f(n - 1) \leq f'(n - 1) \]

for all \( n \).

\[\text{[10.16]}\]

This inequality is equivalent to Eq. (7.3) because it gives,

\[ [f(n + 1) - f(M)] \leq \sum_{n=M}^{N+1} f'(n) \leq \sum_{n=M+1}^{N+1} f'(n) \leq [f(n + 1) - f(M)] + f'(M) - f'(N + 1). \]

**Corollary 7.6 (p-series).** Let \( p \in \mathbb{R} \), then

\[ \sum_{n=1}^{\infty} \frac{1}{n^p} = \begin{cases} \infty & \text{if } p \leq 1 \\ < \infty & \text{if } p > 1 \end{cases} \]

**Proof.** If \( p \leq 0 \), then \( \lim_{n \to \infty} \frac{1}{n^p} \neq 0 \) and hence the series diverges and we may assume that \( p > 0 \). For \( p > 1 \), let \( f(x) = -x^{-(p-1)} \) so that \( f'(x) = (p-1)x^{-p} \) which is decreasing in \( x \) and therefore by Theorem 7.5,

\[ f(\infty) - f(M) \leq \sum_{n=M}^{\infty} \frac{p-1}{n^p} \leq f(\infty) - f(M) + f'(M) = \frac{1}{M^{p-1}} + \frac{p-1}{M^p} < \infty \]

and so

\[ \frac{1}{(p-1)M^{p-1}} \leq \sum_{n=M}^{\infty} \frac{1}{n^p} \leq \frac{1}{(p-1)M^{p-1}} + \frac{1}{M^p}. \]

In particular for \( M = 1 \),

\[ \frac{1}{p-1} \leq \sum_{n=1}^{\infty} \frac{1}{n^p} \leq \frac{1}{p-1} + 1. \]

Since

\[ \frac{1}{p-1} \leq \sum_{n=1}^{\infty} \frac{1}{n^p} \leq \sum_{n=1}^{\infty} \frac{1}{n} \]
we may let $p \downarrow 1$ in order to show, $\infty \leq \sum_{n=1}^{\infty} \frac{1}{n}$. This complete the proof as for any $p < 1$ we have ,

$$\infty \leq \sum_{n=1}^{\infty} \frac{1}{n} \leq \sum_{n=1}^{\infty} \frac{1}{n^p}.$$ 

Exercise 7.1. Take $f(x) = \ln x$ in Theorem 7.5 in order to directly conclude that $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$.

Exercise 7.2. Let $0 \leq p < \infty$. Prove,

$$\sum_{n=2}^{\infty} \frac{1}{n (\ln n)^p} = \begin{cases} \infty & \text{if } 0 \leq p \leq 1 \\ < \infty & \text{if } p > 1 \end{cases}.$$ 

by applying Theorem 7.5 with the following functions;

1. for $0 \leq p < 1$ take $f(x) = (\ln x)^{1-p}$,
2. for $p = 1$ take $f(x) = \ln \ln x$, and
3. for $p > 1$ take $f(x) = - (\ln x)^{1-p}$.

Theorem 7.7. Let $(X, \| \cdot \|)$ be a Banach space and $\{x_k\}_{k=1}^{\infty} \subset X$ be a sequence. Then;

1. $\sum_{k=1}^{\infty} x_k$ converges iff

$$\left\| \sum_{k=m}^{n} x_k \right\| \to 0 \text{ as } n, m \to \infty \text{ with } n \geq m.$$ 

2. If $\sum_{k=1}^{\infty} x_k$ converges then $\lim_{k \to \infty} x_k = 0$ or alternatively if $\lim_{k \to \infty} x_k \neq 0$ then $\sum_{k=1}^{\infty} x_k$ diverges.

3. If $\sum_{k=1}^{\infty} x_k$ converges then $\lim_{N \to \infty} \sum_{k=N}^{\infty} x_k = 0$, i.e. the $N$ - tail, $\sum_{k=N}^{\infty} x_k$, of a convergent series, $\sum_{k=1}^{\infty} x_k$, go to zero as $N \to \infty$.

Remark: the only place we use $X$ is complete is in the implication (⇐) in item 1. All of the remaining statements hold for any normed space $X$.

Proof. Let $S_n := \sum_{k=1}^{n} x_k$ so that $\sum_{k=1}^{\infty} x_k$ converges iff $\lim_{n \to \infty} S_n$ exists iff $\{S_n\}_{n=1}^{\infty}$ is Cauchy since $X$ is a Banach space which gives item 1. since $S_n - S_m = \sum_{k=m+1}^{n} x_k$. For the second item apply the first with $n = m + 1$. For the third item let $S := \sum_{k=1}^{\infty} x_k$, then $\lim_{N \to \infty} S_N = S$ and so by very definition,

$$\sum_{k=N}^{\infty} x_k = S - S_{N+1} \to 0 \text{ as } N \to \infty.$$ 

Exercise 7.3. 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

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

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

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

Theorem 7.8 (Absolute Convergence Implies Convergence). Let $(X, \| \cdot \|)$ be a Banach space $\{x_k\}_{k=1}^{\infty} \subset X$ be a sequence. Then $\sum_{k=1}^{\infty} \|x_k\| < \infty$ implies $\sum_{k=1}^{\infty} x_k$ is convergent. [We say $\sum_{k=1}^{\infty} x_k$ is absolutely convergent if $\sum_{k=1}^{\infty} \|x_k\| < \infty$.]

Exercise 7.4. Prove Theorem 7.8. Namely if $(X, \| \cdot \|)$ is a Banach space and $\{x_k\}_{k=1}^{\infty} \subset X$ is a sequence, then $\sum_{k=1}^{\infty} \|x_k\| < \infty$ implies $\sum_{k=1}^{\infty} x_k$ is convergent.

Proposition 7.9 (Alternating Series Test). If $\{a_k\}_{k=1}^{\infty} \subset [0, \infty)$ is a non-increasing sequence (i.e. $a_k \geq a_{k+1}$ for all $k$) such that $\lim_{k \to \infty} a_k = 0$, then $s := \sum_{k=1}^{\infty} (-1)^{k+1} a_k$ is convergent. Moreover, for all $n \in \mathbb{N}$

$$\left| s - \sum_{k=1}^{n} (-1)^{k+1} a_k \right| = \sum_{k=n+1}^{\infty} (-1)^{k+1} a_k \leq a_{n+1}. \quad (7.5)$$
which imply Eq. (7.5).

**Alternatively,** here is another way to think about the error estimates:

\[ 0 \leq a_1 - a_2 = S_2 \leq S \leq S_1 \leq a_1 \]

so that \( 0 \leq S \leq a_1 \). Applying these results to \( \sum_{k=n+1}^{\infty} (-1)^{k+1} a_k \) shows,

\[ \left| s - \sum_{k=1}^{n} (-1)^{k+1} a_k \right| = \sum_{k=n+1}^{\infty} (-1)^{k+1} a_k \leq a_{n+1}. \]

**Example 7.10.** By the alternating series test,

\[ \sum_{n=1}^{\infty} (-1)^{n} \frac{1}{n^p} \]

is convergent for all \( p > 0 \). On the other hand, by Corollary 7.6, the series is absolutely convergent if \( p > 1 \).

**Theorem 7.11 (Uniform convergence and the Weierstrass M-test).** Suppose that \((X, \|\cdot\|)\) is a Banach space, \((Y, d)\) is a metric space, for each \( n \in \mathbb{N} \), \( f_n : Y \to X \) is a function, and there exists \( \{M_n\}_{n=1}^{\infty} \subset [0, \infty) \) satisfying:

\[ \sup_{y \in Y} \|f_n(y)\|_X \leq M_n \ orall n \in \mathbb{N} \text{ and } \sum_{n=1}^{\infty} M_n < \infty. \]

Then \( S_N(y) := \sum_{n=1}^{N} f_n(y) \) converges absolutely and uniformly to \( S(y) = \sum_{n=1}^{\infty} f_n(y) \). Moreover, if we further assume that \( \{f_n\}_{n=1}^{\infty} \subset C(Y, X) \) (i.e. \( f_n \) is continuous for all \( n \)), then the function \( S : Y \to X \) is also continuous.

**Proof.** For any \( y \in Y \),

\[ \sum_{n=1}^{\infty} \|f_n(y)\|_X \leq \sum_{n=1}^{\infty} M_n < \infty \]

and therefore \( S(y) = \sum_{n=1}^{\infty} f_n(y) \) converges absolutely. Moreover we have,

\[ \|S(y) - S_N(y)\| = \lim_{M \to \infty} \|S_M(y) - S_N(y)\| = \lim_{M \to \infty} \left\| \sum_{n=N+1}^{M} f_n(y) \right\| \]

\[ \leq \liminf_{M \to \infty} \sum_{n=N+1}^{M} \|f_n(y)\| = \sum_{n=N+1}^{\infty} \|f_n(y)\| \leq \sum_{n=N+1}^{\infty} M_n. \]
As the last member of this inequality does not depend on \( y \) we have,

\[
\sup_{y \in Y} ||S(y) - S_N(y)|| \leq \sum_{n=N+1}^{\infty} M_n \to 0 \text{ as } N \to \infty
\]

because tails of convergent series vanish, Theorem 7.7. The continuity of \( S \) now follows from the continuity of \( S_N \), the uniform convergence just proved, and Theorem 6.48.

**Theorem 7.12 (Root test).** Suppose that \( \{a_n\}_{n=1}^{\infty} \) is a sequence in \( \mathbb{C} \) and let \( \alpha := \limsup_{n \to \infty} |a_n|^{1/n} \). Then

1. If \( \alpha < 1 \) then \( \sum_{n=1}^{\infty} |a_n| < \infty \) and \( \sum_{n=1}^{\infty} a_n \) is absolutely convergent.
2. If \( \alpha > 1 \), then \( \limsup_{n \to \infty} |a_n| = \infty \) and \( \sum_{n=1}^{\infty} a_n \) diverges.
3. If \( \alpha = 1 \), the test fails, i.e. you must work harder!

**Proof.** We take each item in turn.

1. If \( \alpha < 1 \), let \( \beta \in (\alpha, 1) \), then \( |a_n|^{1/n} \leq \beta \) for a.a. \( n \) which implies that \( |a_n| \leq \beta^n \) for a.a. \( n \) and so the result follows by the comparison Theorem 7.2 as

\[
\sum_{n=1}^{\infty} \beta^n = \frac{\beta}{1 - \beta} < \infty.
\]

2. If \( \alpha > 1 \) and \( \beta \in (1, \alpha) \), then \( |a_n|^{1/n} \geq \beta \) i.o. \( n \) and hence \( |a_n| \geq \beta^n \) i.o. \( n \). As \( \beta^n \to \infty \) as \( n \to \infty \) it follows that \( \limsup_{n \to \infty} |a_n| = \infty. \)

3. From Corollary 7.6 we know that \( \sum_{n=1}^{\infty} \frac{1}{n} = \infty \) while \( \sum_{n=1}^{\infty} \frac{1}{n^p} < \infty. \)

However from Lemma 3.30

\[
\lim_{n \to \infty} \left( \frac{1}{n^p} \right)^{1/n} = 1 = \lim_{n \to \infty} \left( \frac{1}{n^p} \right)^{1/n}
\]

which shows the test has failed. In fact, if \( 0 < p < \infty \) and \( k \in \mathbb{N} \) such that \( k \geq p \), then \( \frac{1}{n^p} \leq \frac{1}{n^p} \leq 1 \) which implies

\[
\left( \frac{1}{n^k} \right)^{1/n} \leq \left( \frac{1}{n^p} \right)^{1/n} \leq (1)^{1/n} = 1.
\]

By Lemma 3.30 we know \( \lim_{n \to \infty} \left( \frac{1}{n^p} \right)^{1/n} = 1 \) and therefore by the sandwich lemma, \( \lim_{n \to \infty} \left( \frac{1}{n^p} \right)^{1/n} = 1. \)

**Exercise 7.5.** For every \( p \in \mathbb{N} \), show \( \sum_{n=0}^{\infty} (n^{n/p})^n \) is convergent iff \( z = 0. \)

---

**Theorem 7.13 (Ratio test).** Suppose that \( \{a_n\}_{n=1}^{\infty} \) is a sequence in \( \mathbb{C} \) such that \( a_n \neq 0 \) for a.a. \( n \). Then

1. If \( \alpha := \limsup_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right| < 1 \) then \( \sum_{n=1}^{\infty} |a_n| < \infty \) and \( \sum_{n=1}^{\infty} a_n \) is absolutely convergent.
2. If \( \frac{a_{n+1}}{a_n} \geq 1 \) for a.a. \( n \) then \( \lim_{n \to \infty} |a_n| > 0 \) and \( \sum_{n=1}^{\infty} a_n \) diverges.
3. If \( \liminf_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right| > 1 \) then \( \lim_{n \to \infty} |a_n| = \infty \) and \( \sum_{n=1}^{\infty} a_n \) diverges.
4. If \( \limsup_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right| = 1 \), the test fails, i.e. you must work harder!

**Proof.** We take each item in turn.

1. If \( \alpha < 1 \), let \( \beta \in (\alpha, 1) \), then \( \left| \frac{a_{n+1}}{a_n} \right| \leq \beta \) for a.a. \( n \), i.e. there exists \( N \in \mathbb{N} \) such that \( |a_{n+1}| \leq \beta |a_n| \) for all \( n \geq N \). A simple induction argument shows,

\[
|a_n| \leq |a_N| |a^{\alpha-N} = a^{-N} |a_N| \beta^n \text{ for } n \geq N.
\]

The result follows by the comparison Theorem 7.2 and the fact that

\[
\sum_{n=N}^{\infty} \beta^{-N} |a_N| \beta^n = |a_N| \frac{1}{1 - \beta} < \infty.
\]

2. Suppose there exists \( N \in \mathbb{N} \) such that \( \left| \frac{a_{n+1}}{a_n} \right| \geq 1 \) for all \( n \geq N \). This inequality says that \( |a_n| \) is non-decreasing for large \( n \) and therefore \( \lim_{n \to \infty} |a_n| \geq |a_m| \) for any \( m \geq N \). We may now choose \( m \geq N \) such that \( |a_m| \neq 0. \)

3. If \( \liminf_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right| > \beta > 1 \), then \( \left| \frac{a_{n+1}}{a_n} \right| \geq \beta \) for a.a. \( n \). Working as above there exists \( N \in \mathbb{N} \) such that \( |a_N| \neq 0 \) and \( |a_n| \geq |a_N| |a^{\beta-N} \text{ for all } n \geq N. \)

From this it follows that \( \lim_{n \to \infty} |a_n| = \infty. \)

4. From Corollary 7.6 we know that \( \sum_{n=1}^{\infty} \frac{1}{p} < \infty \) iff \( p > 1 \). However

\[
\lim_{n \to \infty} \left( \left( \frac{n+1}{n} \right)^p \right) = \lim_{n \to \infty} \left( \frac{n}{n+1} \right)^p = \left[ \lim_{n \to \infty} \left( \frac{n}{n+1} \right) \right]^p = 1^p = 1
\]

which shows the test has failed and does not resolve when \( \sum_{n=1}^{\infty} \frac{1}{n^p} < \infty. \)

In what follows let \( (Z, ||||) \) be a complex Banach space. For example, \( Z = \mathbb{C} \) and \( ||z|| = |z| \) is an important special case.
Remark 7.14. The root and ratio tests extend to series in Banach spaces \( \{a_n\}_{n=1}^{\infty} \subset Z \) with the only changes being that we replace \(|a_n|\) by \(\|a_n\|\) and \(\left|\frac{a_{n+1}}{a_n}\right|\) by \(\frac{\|a_{n+1}\|}{\|a_n\|}\) wherever these occur. The point is that we can apply the root and ratio test to the sequence of real numbers \(\{\|a_n\|\}_{n=1}^{\infty}\) in order to make conclusion about the absolute convergence or divergence of \(\sum_{n=1}^{\infty} a_n\).

Definition 7.15. Given \(z_0 \in \mathbb{C}\) and \(\{a_n\}_{n=0}^{\infty} \subset Z\), the series of the form

\[
\sum_{n=0}^{\infty} a_n (z - z_0)^n
\]  

(7.6)

is called a power series. If \(z_0 = 0\) we call it a Maclaurin series, i.e., a series of the form

\[
\sum_{n=0}^{\infty} a_n z^n
\]  

(7.7)

The \textit{radius of convergence} of either of these series is defined to be \(R := \frac{1}{\alpha} \in [0, \infty]\) where

\[
\alpha := \limsup_{n \to \infty} \|a_n\|^{1/n} \in [0, \infty].
\]

By definition, \(\frac{1}{\alpha} := \infty\) in this formula.

The next theorem shows that \(R\) is the critical radius governing the convergence of Eqs. (7.6) and (7.7).

Proposition 7.16. If \(R\) is the radius of convergence of a power series in Eq. (7.6) then:

1. If \(|z - z_0| < R\), the series converges.
2. If \(|z - z_0| > R\), the series diverges.
3. If \(|z - z_0| = R\), the series may or may not converge.

We may also characterize \(R\) as

\[
R := \sup \left\{ |z - z_0| : z \in \mathbb{C} \text{ and } \sum_{n=0}^{\infty} a_n (z - z_0)^n \text{ converges} \right\}
\]

(7.8)

Proof. Let

\[
\rho(z) := \limsup_{n \to \infty} \|a_n (z - z_0)^n\|^{1/n} = \left|\frac{z - z_0}{R}\right|
\]

By the root test, Theorem 7.12, we know that power series in Eq. (7.6) converges if \(\rho(z) < 1\) and diverges if \(\rho(z) > 1\) and the test fails if \(\rho(z) = 1\).

Corollary 7.17. If \(\mu = \lim_{n \to \infty} \frac{\|a_{n+1}\|}{\|a_n\|} \) exists, the radius \((R)\) of convergence of a power series in Eq. (7.6) is \(R = \frac{1}{\mu} = \lim_{n \to \infty} \frac{\|a_n\|}{\|a_{n+1}\|}\).

Proof. If \(\mu = \lim_{n \to \infty} \frac{\|a_{n+1}\|}{\|a_n\|} \), then

\[
\rho(z) := \lim_{n \to \infty} \frac{\|a_{n+1} (z - z_0)^n+1\|}{\|a_n (z - z_0)^n\|} = |z - z_0| \lim_{n \to \infty} \frac{\|a_{n+1}\|}{\|a_n\|} = \mu \cdot |z - z_0|.
\]

The Ratio test in Theorem 7.13 now implies \(\sum_{n=0}^{\infty} a_n (z - z_0)^n\) converges if \(|z - z_0| < \frac{1}{\mu}(\rho(z) < 1)\) and diverges if \(|z - z_0| > \frac{1}{\mu}(\rho(z) > 1)\), i.e., \(R = \frac{1}{\mu}\) from Eq. (7.8).

Theorem 7.18. If the radius of convergence \((R)\) of a power series \(\sum_{n=0}^{\infty} a_n (z - z_0)^n\) is positive \((R > 0)\), then the functions

\[
S : D(z_0, R) := \{z \in \mathbb{C} : |z - z_0| < R\} \to \mathbb{C}
\]

defined by \(S(z) := \sum_{n=0}^{\infty} a_n (z - z_0)^n\) is continuous and the series is uniformly convergent on \(D(z_0, R)\) for all \(R < R\).

Proof. For any \(\rho < R\) let \(M_n := \|a_n\| \rho^n\). Then \(\sum_{n=0}^{\infty} M_n < \infty\) by the root test or the definition of \(R\). So if \(z \in D(z_0, \rho)\) we find,

\[
\|a_n (z - z_0)^n\| \leq \|a_n\| \rho^n = M_n.
\]

It follows from the Weierstrass M-test, Theorem 7.13, that the series is uniformly convergent and \(S\) is continuous on \(D(z_0, \rho)\) for all \(R < R\). Since \(D(z_0, R) = \bigcup_{0 < \rho < R} D(z_0, \rho)\), we may conclude that \(S\) is continuous on \(D(z_0, R)\).

Exercise 7.6. Show that each of the following power series have an infinite radius of convergence and hence define continuous functions from \(\mathbb{C}\) to \(\mathbb{C}\).

1. \(\exp(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!}\),
2. \(\sin(z) := \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{(2n+1)!}\), and
3. \(\cos(z) := \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n}}{(2n)!}\).

Proposition 7.19. Suppose that \((X, \|\cdot\|)\) is a normed space and \(\{a_n\}_{n=1}^{\infty}, \{b_n\}_{n=1}^{\infty} \subset X\) such that \(A := \sum_{n=1}^{\infty} a_n\) and \(B := \sum_{n=1}^{\infty} b_n\) exist in \(X\). Then for all \(\lambda \in \mathbb{F} (\mathbb{F} = \mathbb{R} \text{ or } \mathbb{C})\) if \(X\) is a real or complex vector space respectively, we have the series, \(\sum_{n=1}^{\infty} (a_n + \lambda b_n)\), is convergent and

\[
\sum_{n=1}^{\infty} (a_n + \lambda b_n) = A + \lambda B.
\]
The easy proof of this proposition is left to the reader. Our next goal is to consider the multiplication of series. In this case we will need a multiplication on $X$ itself. We start with the case where $X = \mathbb{C}$ or $X = \mathbb{R}$. Let $\{a_n\}_{n=0}^{\infty}, \{b_n\}_{n=0}^{\infty} \subset \mathbb{C}$ and let us consider the power series,

$$A(z) := \sum_{n=0}^{\infty} a_n z^n \quad \text{and} \quad B(z) := \sum_{n=0}^{\infty} b_n z^n.$$  

Working formally,

$$A(z) B(z) = \left( \sum_{n=0}^{\infty} a_n z^n \right) \left( \sum_{m=0}^{\infty} b_m z^m \right) = \sum_{m,n=0}^{\infty} a_n b_m z^n z^m = \sum_{k=0}^{\infty} \sum_{m+n=k} a_n b_m z^k.$$

Thus we expect $A(z) B(z)$ to be represented by a power series

$$C(z) = \sum_{k=0}^{\infty} c_k z^k \quad \text{where} \quad c_k := \sum_{m+n=k} a_n b_m = \sum_{m=0}^{k} a_{k-m} b_m.$$  

**Theorem 7.20 (Multiplication of Series).** Suppose that $\{a_n\}_{n=0}^{\infty}, \{b_n\}_{n=0}^{\infty} \subset \mathbb{C}$ and $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$ are absolutely convergent series. Let $\{c_k\}_{k=0}^{\infty}$ be defined by

$$c_k := \sum_{m=0}^{k} a_{k-m} b_m. \quad (7.9)$$

Then $\sum_{k=0}^{\infty} c_k$ converges absolutely and

$$\sum_{k=0}^{\infty} c_k = \left( \sum_{n=0}^{\infty} a_n \right) \cdot \left( \sum_{n=0}^{\infty} b_n \right). \quad (7.10)$$

**Proof.** Let $J_N = \{0, 1, 2, \ldots, N\}$, $\Gamma_N := \{(m,n) \in J_N^2 : m + n > N\}$, and

$$R_N = \sum_{(m,n) \in \Gamma_N} |a_n b_m|.$$  

I now claim that $\lim_{N \to \infty} R_N = 0$. To see this let $A := \sum_{n=0}^{\infty} |a_n|$, $B := \sum_{m=0}^{\infty} |b_m|$ and observe $m + n > N$ implies either $m > N/2$ or $n > N/2$, see Figure 7.3. Hence we find,

$$R_N \leq \sum_{\frac{N}{2} < m \leq N} \sum_{n=0}^{N} |a_n b_m| + \sum_{\frac{N}{2} < n \leq N} \sum_{m=0}^{N} |a_n b_m|$$

$$\leq \sum_{\frac{N}{2} < m \leq N} |b_m| \sum_{n=0}^{N} |a_n| + \sum_{\frac{N}{2} < n \leq N} |a_n| \sum_{m=0}^{N} |b_m|$$

$$\leq \sum_{m > \frac{N}{2}} |b_m| \cdot A + \sum_{n > \frac{N}{2}} |a_n| \cdot B \to 0 \quad \text{as} \quad N \to \infty$$

because the tails of convergent series tend to zero.

We now have the identity,
Let Corollary 7.21.  

\[ \sum_{n=0}^{N} a_n \cdot \sum_{m=0}^{N} b_m = \sum_{(m,n)\in J_N \times J_N} a_n b_m \]

\[ = \sum_{m+n \leq N} a_n b_m + r_N = \sum_{k=0}^{N} c_k + r_N, \]

where \( r_N := \sum_{(m,n)\in \Gamma_N} a_n b_m. \) Since \(|r_N| \leq R_N \to 0\), it follows that

\[ \sum_{k=0}^{\infty} c_k = \lim_{N \to \infty} \sum_{k=0}^{N} c_k = \lim_{N \to \infty} \left[ \sum_{n=0}^{N} a_n \cdot \sum_{m=0}^{N} b_m - r_N \right] \]

\[ = \lim_{N \to \infty} \sum_{n=0}^{N} a_n \cdot \lim_{N \to \infty} \sum_{m=0}^{N} b_m = \left( \sum_{n=0}^{\infty} a_n \right) \cdot \left( \sum_{n=0}^{\infty} b_n \right). \]

Lastly observe that

\[ |c_k| \leq \sum_{m=0}^{k} |a_{k-m}| |b_m| =: \bar{c}_k \]

and so applying what we have just proved to \( \{a_n\}_{n=0}^{\infty} \) and \( \{b_n\}_{n=0}^{\infty} \) shows

\[ \sum_{k=0}^{\infty} |c_k| \leq \sum_{k=0}^{\infty} \bar{c}_k = \left( \sum_{n=0}^{\infty} |a_n| \right) \cdot \left( \sum_{n=0}^{\infty} |b_n| \right) < \infty. \]

Here is the short summary of the argument;

\[ \left| \sum_{(m,n)\in \Gamma_N} a_n \cdot b_m - \sum_{k=0}^{N} c_k \right| \leq \sum_{(m,n)\in \Gamma_N} |a_n \cdot b_m| \]

\[ \leq \sum_{n \leq m \leq N} |a_n \cdot b_m| + \sum_{n \leq m < N} |a_n \cdot b_m| \]

\[ \leq \sum_{m \geq \frac{N}{2}} |b_m| \cdot A + \sum_{n \geq \frac{N}{2}} |a_n| \cdot B \to 0 \text{ as } N \to \infty. \]

\[ \sum_{k=0}^{\infty} c_k z^k \text{ is convergent for } |z| < R \text{ and} \]

\[ \sum_{k=0}^{\infty} c_k z^k = \left( \sum_{n=0}^{\infty} a_n z^n \right) \left( \sum_{m=0}^{\infty} b_m z^m \right). \]

**Proof.** For \(|z| < R\), the root test shows

\[ \sum_{n=0}^{\infty} |a_n z^n| < \infty \text{ and } \sum_{m=0}^{\infty} |b_m z^m| < \infty. \]

The result now follows by applying Theorem 7.20 with \( a_n \to |a_n z^n| \) and \( b_n \to |b_n z^n| \).

**Example 7.22.** For \(|z| < 1\) we have

\[ (1 - z) \cdot \sum_{n=0}^{\infty} z^n = 1. \]

This shows that power series, \( \sum_{k=0}^{\infty} c_k z^k \), in Corollary 7.21 may in fact have radius of convergence larger than \( R \) in Eq. (7.11).

**Theorem 7.23.** The function \( \exp : \mathbb{C} \to \mathbb{C} \) satisfies the following properties:

1. \( \exp (0) = 1 \),
2. \( \exp (w) \exp (z) = \exp (w + z) \) for all \( w, z \in \mathbb{C} \),
3. \( \exp (w) \) is never 0 and \( \exp (w) \to 1 = \exp (-w) \) for all \( w \in \mathbb{C} \).

**Proof.** The first assertion is clear and item 3. easily follows from item 2. with \( z = -w \). We now prove item 2. by applying Theorem 7.20 with \( a_n := \frac{1}{n!} z^n \) and \( b_n := \frac{1}{n!} w^n \). In this case,

\[ c_k = \sum_{m=0}^{k} \frac{1}{m!} z^m \cdot \frac{1}{(k-m)!} w^{k-m} = \frac{1}{k!} \sum_{m=0}^{k} \left( \begin{array}{c} k \\ m \end{array} \right) z^m w^{k-m} = \frac{1}{k!} (z + w)^k \]

wherein we have used the Binomial theorem for the last equality, see Exercise 7.10. It now follows from Theorem 7.20 that

\[ \exp (w) \exp (z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \cdot \sum_{n=0}^{\infty} \frac{w^n}{n!} = \sum_{k=0}^{\infty} \frac{1}{k!} (z + w)^k = \exp (w + z). \]
Definition 7.24 (Euler’s Number). Euler’s number is defined by
\[ e := \exp(1) = \sum_{m=0}^{\infty} \frac{1^m}{m!} = \sum_{m=0}^{\infty} \frac{1}{m!}. \]

If \( m \in \mathbb{N} \), then
\[ e = \exp(1) = \exp \left( \frac{1}{m} \right) = \left[ \exp \left( \frac{1}{m} \right) \right]^m. \]

As \( \exp \left( \frac{1}{m} \right) > 0 \) we may conclude that \( \exp \left( \frac{1}{m} \right) = e^{1/m} \). Now if \( n \in \mathbb{N}_0 \) we find,
\[ e^{1/m} = \left[ \exp \left( \frac{1}{m} \right) \right]^n = \exp \left( \frac{n}{m} \right). \]

As
\[ e^{-1/m} = \frac{1}{e^{1/m}} = \frac{1}{\exp \left( \frac{n}{m} \right)} = \exp \left( -\frac{n}{m} \right), \]
we have in fact shown that \( e^q = \exp(q) \) for all \( q \in \mathbb{Q} \). Owing to this identity, we will often write \( e^z \) for \( \exp(z) \).

With the definitions in Exercise 7.6, we have
\[ \exp(iz) = \sum_{n=0}^{\infty} (iz)^n = \sum_{n=0}^{\infty} \frac{(iz)^{2n}}{(2n)!} + \sum_{n=0}^{\infty} \frac{(iz)^{2n+1}}{(2n+1)!} = \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n}}{(2n)!} + i \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{(2n+1)!} = \cos(z) + i \sin(z), \]
i.e.
\[ \exp(iz) = \cos(z) + i \sin(z) \text{ for all } z \in \mathbb{C}. \tag{7.12} \]
This is called Euler’s formula. If \( z = \theta \in \mathbb{R} \), it states that
\[ \text{Re } \exp(i\theta) = \cos \theta \text{ and } \text{Im } \exp(i\theta) = \sin \theta. \]
Thus if \( \alpha, \theta \in \mathbb{R} \), we have the following addition formulas;
\[ \cos(\theta + \alpha) = \text{Re } \exp(i(\theta + \alpha)) = \text{Re } [\exp(i\theta) \exp(i\alpha)] = \text{Re } [(\cos \theta + i \sin \theta) \cdot (\cos \alpha + i \sin \alpha)] = \cos \theta \cos \alpha - \sin \theta \sin \alpha \tag{7.13} \]
and
\[ \sin(\theta + \alpha) = \text{Im } \exp(i(\theta + \alpha)) = \text{Im } [\exp(i\theta) \exp(i\alpha)] = \cos \theta \sin \alpha + \sin \theta \cos \alpha. \]

We also define
\[ \sinh(z) = \frac{e^z - e^{-z}}{2} = \sum_{n=0}^{\infty} \frac{z^{2n+1}}{(2n+1)!} \quad \text{and} \quad \cosh(z) = \frac{e^z + e^{-z}}{2} = \sum_{n=0}^{\infty} \frac{z^{2n}}{(2n)!}. \]

The following identities are now easily verified;
\[ \cos(-z) = \cos(z) \quad \text{and} \quad \cosh(-z) = \cosh(z), \]
\[ \sin(-z) = -\sin(z) \quad \text{and} \quad \sinh(-z) = -\sinh(z), \]
\[ e^z = \cos(z) + \sin(z), \]
\[ e^z = \cos(z) + \sin(z), \]
\[ \sin(z) = \frac{e^{iz} - e^{-iz}}{2i} = \frac{1}{i} \sinh(z), \quad \text{and} \]
\[ \cos(z) = \frac{e^{iz} + e^{-iz}}{2} = \cosh(z). \]

Exercise 7.7 (Addition Formulas). Prove the addition formulas in Eqs. (7.13) and (7.14) extend to \( \alpha, \theta \in \mathbb{C} \), i.e. show for all \( w, z \in \mathbb{C} \) that
\[ \cos(w + z) = \cos(w) \cos(z) - \sin(w) \sin(z) \quad \text{and} \quad \sin(w + z) = \cos(w) \sin(z) + \cos(z) \sin(w). \]

Exercise 7.8. Suppose that \( p(t) \) and \( q(t) \) are non-zero polynomials of \( t \in \mathbb{R} \) with (possibly) complex coefficients. Further assume that \( q(n) \neq 0 \) for all \( n \in \mathbb{N}_0 \). Show for any sequence \( \{a_n\}_{n=0}^{\infty} \subset \mathbb{C} \) that the power series;
\[ \sum_{n=0}^{\infty} p(n) a_n z^n \quad \text{and} \quad \sum_{n=0}^{\infty} q(n) a_n z^n \]
have the same radius of convergence.

Theorem 7.25 (Hilbert Schmidt norm). Let \( Z = \mathbb{C}^{J_N \times J_N} \) denote the space of \( N \times N \) complex matrices, \( A = (A_{ij})_{i,j=1}^{N} \) with \( A_{ij} \in \mathbb{C} \). We let
\[ ||A||_2 := \left[ \sum_{i,j=1}^{N} |A_{ij}|^2 \right]^{1/2} \]
which is called the Hilbert Schmidt norm on \( Z \). This norm satisfies,
1. \( ||I||_2 = \sqrt{N} \) where \( I \) is the \( N \times N \) identity matrix.

2. \( ||AB||_2 \leq ||A||_2 \cdot ||B||_2 \) for all \( A, B \in Z \).

3. \( ||A^n||_2 \leq ||A||^{2n} \) for all \( n \in N \).

4. If \( A_n \to A \) and \( B_n \to B \) then \( A_nB_n \to AB \) and \( A_n + B_n \to A + B \) as \( n \to \infty \). Thus matrix multiplication and addition are continuous operations on \((Z, ||\cdot||_2)\).

5. If \((X, d)\) is a metric space and \( f : X \to Z \) and \( g : X \to Z \) are continuous functions then so is \( f \cdot g \) (order matters) and \( f + g \).

6. The functions \( f (A) = A^n \) is continuous on \( Z \) for all \( n \in N_0 \), where by convention \( A^0 = I \).

7. \((Z, ||\cdot||_2)\) is a Banach space.

**Proof.** Item 1 is clear. From the definition of matrix multiplication and the Cauchy Schwarz inequality we find

\[
\left| (AB)_{ij} \right|^2 = \left| \sum_{k=1}^{N} A_{ik}B_{kj} \right|^2 \leq \sum_{k=1}^{N} |A_{ik}|^2 \cdot \sum_{k=1}^{N} |B_{kj}|^2.
\]

Therefore,

\[
||AB||_2^2 = \sum_{i,j=1}^{n} \left| (AB)_{ij} \right|^2 \leq \sum_{i,j=1}^{n} \left( \sum_{k=1}^{N} |A_{ik}|^2 \cdot \sum_{k=1}^{N} |B_{kj}|^2 \right) = \sum_{i,k=1}^{N} |A_{ik}|^2 \cdot \sum_{j,k=1}^{N} |B_{kj}|^2 = ||A||^2_2 \cdot ||B||^2_2
\]

which proves item 2. Item 3 follows by an easy induction argument.

Item 4. Let \( \delta A_n := A_n - A \) and \( \delta B_n := B_n - B \) so that \( A_n = A + \delta A_n \) and \( B_n = B + \delta B_n \) where \( ||\delta A_n||_2 \to 0 \) and \( ||\delta B_n||_2 \to 0 \) as \( n \to \infty \). Then

\[
||A_nB_n - AB||_2 = ||(A + \delta A_n)(B + \delta B_n) - AB||_2 = \|\delta A_nB + A\delta B_n + \delta A_n\delta B_n\|_2 \leq \|\delta A_nB\|_2 + \|A\delta B_n\|_2 + \|\delta A_n\delta B_n\|_2 \to 0 \text{ as } n \to \infty.
\]

Similarly,

\[
||A_n + B_n - (A + B)||_2 = \|\delta A_n + \delta B_n\|_2 \to 0 \text{ as } n \to \infty.
\]

Item 5. Suppose that \( \{x_n\}_{n=1}^{\infty} \subset X \) and \( x_n \to x \) as \( n \to \infty \), then using item 4,

\[
\lim_{n \to \infty} (f \cdot g)(x_n) = \lim_{n \to \infty} [f(x_n) + g(x_n)] = f(x) + g(x) = (f + g)(x).
\]

This shows \( f \cdot g \) and \( f + g \) are continuous.

Item 6. follows from item 5 with \( X = Z \) along with an induction argument.

Item 7. follows from the fact that \((Z, ||\cdot||_2)\) is \( \mathbb{C}^{N^2} \) with the \( 2 \)-norm in a slight disguise.

**Definition 7.26 (Matrix Functions).** If \( f(z) = \sum_{n=0}^{\infty} c_n z^n \) for some \( c_n \in \mathbb{C} \) and \( A \in \mathbb{C}^{JN \times JN} \), then we define

\[
f(A) := \sum_{n=0}^{\infty} c_n A^n \text{ provided the sum is convergent.}
\]

For example,

\[
\sin(A) := \sum_{n=0}^{\infty} (-1)^{2n+1} A^{2n+1}/(2n+1)!
\]

**Theorem 7.27.** Suppose that \( \{c_n\}_{n=1}^{\infty} \) is a sequence in \( \mathbb{C} \) and \( R := (\limsup_{n \to \infty} |c_n|^{1/n})^{-1} \). If \( A \in \mathbb{C}^{JN \times JN} \) with \( ||A|| < R \), then

\[
f(A) := \sum_{n=0}^{\infty} c_n A^n
\]

is convergent and \( f : B_0(R) := \{ A : ||A|| < R \} \to \mathbb{C}^{JN \times JN} \) is continuous. Moreover, for any \( \rho < R \), the sum converges absolutely and uniformly on \( B_0(\rho) \).

**Proof.** Let \( \rho \in (||A||, R) \) and let \( M_n := |c_n| \rho^n \). Then \( \limsup_{n \to \infty} M_n^{1/n} = \rho/R < 1 \) and therefore by the Root test, \( \sum_{n=0}^{\infty} M_n < \infty \). Since, for \( n \geq 1 \),

\[
||c_n A^n|| = |c_n| ||A^n|| \leq |c_n| ||A||^n \leq M_n,
\]

the results are now again a consequence of the Weierstrass M-test in Theorem 7.11.

**Theorem 7.28 (Matrix Exponentials and etc.).** The series for \( e^A, \sin (A), \cos (A), \sinh (A), \) and \( \cosh (A) \) are all absolutely convergent define continuous functions from \( \mathbb{C}^{JN \times JN} \) to \( \mathbb{C}^{JN \times JN} \) in the Hilbert Schmidt norm.

**Proof.** Since the corresponding numerical series all of infinite radius of convergence the result follows from Theorem 7.27.
Exercise 7.9 (Inverting perturbations of the identity). For \( \|A\|_2 < 1 \) in \( \mathbb{C}^{J_N \times J_N} \), \( I - A \) is invertible and
\[
(I - A)^{-1} = \sum_{n=0}^{\infty} A^n
\]
where the sum is absolutely convergent. Moreover the function \( A \to (I - A)^{-1} \) is continuous on the ball, \( B := \{ A : \|A\|_2 < 1 \} \) and
\[
\left\| (I - A)^{-1} \right\|_2 \leq \sum_{n=0}^{\infty} \|A^n\|_2 \leq \frac{\|A\|_2}{1 - \|A\|_2} + \sqrt{N}.
\]

Exercise 7.10 (Binomial Theorem). Suppose that \( A, B \in \mathbb{C}^{J_N \times J_N} \) are commuting matrices, i.e. \( AB = BA \). For any \( n \in \mathbb{N} \), show by induction that
\[
\sum_{k=0}^{n} \binom{n}{k} A^k B^{n-k} = (A + B)^n.
\] (7.15)

Exercise 7.11 (Matrix Addition Formula). If \( A, B \in \mathbb{C}^{J_N \times J_N} \) are commuting matrices, show \( \exp (A + B) = \exp (A) \cdot \exp (B) \). Hint: mimic the proof of Theorem 7.23.

Example 7.29 (\( \exp(\exp(A) \cdot \exp(B)) \)). Let \( \theta \in \mathbb{R} \) and
\[
A = \begin{bmatrix} 0 & \theta \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 0 \\ -\theta & 0 \end{bmatrix}.
\]

Since \( A^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = B^2 \) it follows that
\[
e^{A} = I + A = \begin{bmatrix} 1 & \theta \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad e^{B} = I + B = \begin{bmatrix} 1 & 0 \\ -\theta & 1 \end{bmatrix}
\]
and in particular,
\[
e^{A}e^{B} = \begin{bmatrix} 1 & \theta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\theta & 1 \end{bmatrix} = \begin{bmatrix} 1 & \theta^2 \theta \\ 0 & 1 \end{bmatrix}.
\]

Let us now compute \( e^{A+B} \). First observe that
\[
A + B = \theta J \quad \text{where} \quad J := \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.
\]

Since
\[
j^2 = -I \quad \text{where} \quad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},
\]
it follows that \( j^{2n} = (-1)^n = (-1)^n I \) and \( j^{2n+1} = j^{2n}j = (-1)^n J \) for all \( n \in \mathbb{N} \). Therefore,
\[
e^{A+B} = e^{\theta J} = \sum_{n=0}^{\infty} \frac{\theta^n}{n!} J^n
\]
\[
= \sum_{n=0}^{\infty} \frac{1}{(2n)!} \theta^{2n} j^{2n} + \sum_{n=0}^{\infty} \frac{1}{(2n+1)!} \theta^{2n+1} j^{2n+1}
\]
\[
= \left( \sum_{n=0}^{\infty} \frac{1}{(2n)!} \theta^{2n} (-1)^n \right) J + \left( \sum_{n=0}^{\infty} \frac{1}{(2n+1)!} \theta^{2n+1} (-1)^n \right) J
\]
\[
= \cos \theta \cdot I + \sin \theta \cdot J = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}.
\]

From these formula it is evident (using facts that you know about \( \sin \theta \) and \( \cos \theta \)) that \( e^{A}e^{B} \neq e^{A+B} \) unless \( \theta = 0 \). There is no contradiction to Exercise 7.11 since
\[
[A, B] = \begin{bmatrix} 0 & \theta \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ \theta & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ -\theta & 0 \end{bmatrix} \begin{bmatrix} 0 & \theta \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} -\theta^2 & 0 \\ 0 & \theta^2 \end{bmatrix} = -\theta^2 \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}
\]
which is zero iff \( \theta = 0 \).

Exercise 7.12 (Bounded Operators). Let \( (X, \|\cdot\|) \) be a normed space and \( T : X \to X \) be a linear transformation and suppose \( T \) is continuous at \( 0 \in X \). Show:

1. There exists \( C < \infty \) such that \( \|T(x)\| \leq C \|x\| \) for all \( x \in X \). Hint: for sake of contradiction suppose that no such \( C < \infty \) exists. Then construct a sequence \( \{y_n\}_{n=1}^{\infty} \subset X \) such that \( \lim_{n \to \infty} y_n = 0 \) while \( \|T(y_n)\| = 1 \) for all \( n \).
2. Show \( T \) is continuous on all of \( X \).
More Sums and Sequences

**Warning:** this chapter is a bit rough and will be edited later when it becomes needed.

### 8.1 Rearrangements

[The stuff about sums of positive numbers could go much earlier in fact.]

**Definition 8.1.** If $A$ is a countable set and $a : A \to [0, \infty]$ is a function, let

$$
\sum_{x \in A} a(x) := \sup_{A_0 \subset f A} \sum_{x \in A_0} a(x).
$$

Notice that if $0 \leq a(x) \leq b(x)$, then $\sum_{x \in A} a(x) \leq \sum_{x \in A} b(x)$. Moreover if $B \subset A$, then

$$
\sum_{x \in B} a(x) = \sum_{x \in A} a(x) 1_B(x).
$$

**Theorem 8.2 (Monotone Convergence Theorem for Sums).** If $A$ is a countable set and $a_n : A \to [0, \infty]$ is an increasing sequence of functions, then

$$
\lim_{n \to \infty} \sum_{x \in A} a_n(x) := \sum_{x \in A} \lim_{n \to \infty} a_n(x).
$$

**Proof.** This is an easy consequence of the sup – sup theorem. Indeed, $\sum_{x \in A} a_n(x)$ is an increasing sequence of numbers, therefore,

$$
\lim_{n \to \infty} \sum_{x \in A} a_n(x) = \sup_n \sum_{x \in A} a_n(x) = \sup_{A_0 \subset f A} \sum_{x \in A_0} a_n(x)
\sup_{A_0 \subset f A} \sum_{x \in A_0} a_n(x)
\sup_{A_0 \subset f A} \sum_{x \in A_0} a_n(x)
\sup_{A_0 \subset f A} \sum_{x \in A_0} a_n(x)
\sup_{A_0 \subset f A} \sum_{x \in A_0} a_n(x)
\sup_{A_0 \subset f A} \sum_{x \in A_0} a_n(x)
\sup_{A_0 \subset f A} \sum_{x \in A_0} a_n(x).
$$

**Corollary 8.3.** Again suppose that $A$ is a countable set and $a : A \to [0, \infty]$ is a function. If $A_n \subset A$ and $A_n \uparrow A$, then

$$
\sum_{A} a(x) = \lim_{n \to \infty} \sum_{A_n} a(x).
$$

**Proof.** Apply Theorem 8.2 with $a_n(x) := a(x) 1_{A_n}(x)$.

**Corollary 8.4.** If $A = A \cup B$ with $A \cap B = \emptyset$, then

$$
\sum_{x \in A} a(x) = \sum_{x \in A} a(x) + \sum_{x \in B} a(x).
$$

**Proof.** Choose $A_n \subset f A$ such that $A_n \uparrow A$. Then $A_n \cap A \uparrow A$ and $A_n \cap B \uparrow B$ and hence,

$$
\sum_{x \in A} a(x) = \lim_{n \to \infty} \sum_{x \in A_n} a(x)
\lim_{n \to \infty} \sum_{x \in A_n \cap A} a(x)
\lim_{n \to \infty} \sum_{x \in A_n \cap B} a(x)
\sum_{x \in A} a(x) + \sum_{x \in B} a(x).
$$

**Corollary 8.5.** If $A$ is a countable set, $a : A \to [0, \infty]$ is a function, and $A = \sum_{n=1}^{\infty} A_n$, then

$$
\sum_{A} a(x) = \sum_{n=1}^{\infty} \sum_{A_n} a(x)
\sum_{n=1}^{\infty} \sum_{A_n} a(x)
\sum_{n=1}^{\infty} \sum_{A_n} a(x)
\sum_{n=1}^{\infty} \sum_{A_n} a(x)
\sum_{n=1}^{\infty} \sum_{A_n} a(x).
$$

**Proof.** We have

$$
\sum_{n=1}^{\infty} \sum_{A_n} a(x) = \lim_{N \to \infty} \sum_{n=1}^{N} \sum_{A_n} a(x) = \lim_{N \to \infty} \sum_{\cup_{n=1}^{N} A_n} a(x) = \sum_{x \in A} a(x).
$$
Corollary 8.6. If \( \{x_n\}_{n=1}^{\infty} \) is any enumeration of \( \Lambda \) then
\[
\sum_{x \in \Lambda} a(x) = \sum_{n=1}^{\infty} a(x_n).
\]

Proof. \( \sum_{n=1}^{\infty} \{x_n\} \) and therefore
\[
\sum_{x \in \Lambda} a(x) = \sum_{n=1}^{\infty} \sum_{x \in \{x_n\}} a(x) = \sum_{n=1}^{\infty} a(x_n).
\]

Corollary 8.7 (Tonelli Theorem for Sums). Suppose that \( \Lambda \) is a countable set and \( a_n : \Lambda \to [0, \infty) \) is a sequence of functions, then
\[
\sum_{n=1}^{\infty} \sum_{x \in \Lambda} a_n(x) = \sum_{x \in \Lambda} \sum_{n=1}^{\infty} a_n(x).
\]

Proof. This is a simple consequence of MCT,
\[
\sum_{n=1}^{\infty} \sum_{x \in \Lambda} a_n(x) = \lim_{N \to \infty} \sum_{n=1}^{N} \sum_{x \in \Lambda} a_n(x) = \lim_{N \to \infty} \sum_{x \in \Lambda} \sum_{n=1}^{N} a_n(x) = \sum_{x \in \Lambda} \lim_{N \to \infty} \sum_{n=1}^{N} a_n(x) = \sum_{x \in \Lambda} \sum_{n=1}^{\infty} a_n(x).
\]

Lemma 8.8 (Fatou’s Lemma). If \( \Lambda \) is a countable set and \( a_n : \Lambda \to [0, \infty) \) is a sequence of functions, then
\[
\sum_{x \in \Lambda} \liminf_{n \to \infty} a_n(x) \leq \liminf_{n \to \infty} \sum_{x \in \Lambda} a_n(x).
\]

Proof. By definition and MCT,
\[
\sum_{x \in \Lambda} \liminf_{n \to \infty} a_n(x) = \sum_{x \in \Lambda} \lim_{n \to \infty} \inf_{k \geq n} a_k(x)
= \lim_{n \to \infty} \sum_{x \in \Lambda} \inf_{k \geq n} a_k(x)
= \liminf_{n \to \infty} \sum_{x \in \Lambda} \inf_{k \geq n} a_k(x)
\leq \liminf_{n \to \infty} \sum_{x \in \Lambda} a_n(x).
\]

Theorem 8.9 (Dominated Convergence Theorem for Sums I). If \( \Lambda \) is a countable set and \( a_n : \Lambda \to [0, \infty) \) is a sequence of functions such that \( \sum_{x \in \Lambda} \sup_{n} a_n(x) < \infty \) and \( \lim_{n \to \infty} a_n(x) = 0 \) for all \( x \in \Lambda \), then
\[
\lim_{n \to \infty} \sum_{x \in \Lambda} a_n(x) = 0.
\]

Proof. For all \( \varepsilon > 0 \) there \( \Lambda_\varepsilon \subset \Lambda \) such that \( \sum_{x \not\in \Lambda_\varepsilon} \sup_{n} a_n(x) \leq \varepsilon \). Therefore,
\[
\sum_{x \in \Lambda} a_n(x) = \sum_{x \in \Lambda_\varepsilon} a_n(x) + \sum_{x \not\in \Lambda_\varepsilon} a_n(x)
\leq \sum_{x \in \Lambda_\varepsilon} a_n(x) + \varepsilon
\]
and therefore,
\[
\limsup_{n \to \infty} \sum_{x \in \Lambda} a_n(x) \leq \limsup_{n \to \infty} \sum_{x \in \Lambda_\varepsilon} a_n(x) + \varepsilon = \lim_{n \to \infty} \sum_{x \in \Lambda_\varepsilon} a_n(x) + \varepsilon = \varepsilon.
\]
Since \( \varepsilon > 0 \) is arbitrary, it follows that
\[
\limsup_{n \to \infty} \sum_{x \in \Lambda} a_n(x) = 0.
\]

Alternative Proof. Let \( \alpha(x) := \sup_{n} a_n(x) \), then \( \alpha(x) - a_n(x) \geq 0 \) for all \( x \in \Lambda \). Therefore by Fatou’s Lemma \[8.8\]
\[
\sum_{x \in \Lambda} \liminf_{n \to \infty} [\alpha(x) - a_n(x)] \leq \liminf_{n \to \infty} \sum_{x \in \Lambda} [\alpha(x) - a_n(x)]
= \sum_{x \in \Lambda} \alpha(x) + \liminf_{n \to \infty} \left[ -\sum_{x \in \Lambda} a_n(x) \right]
= \sum_{x \in \Lambda} \alpha(x) - \limsup_{n \to \infty} \left[ \sum_{x \in \Lambda} a_n(x) \right]
\]
while similarly,
\[
\sum_{x \in \Lambda} \liminf_{n \to \infty} [\alpha(x) - a_n(x)] = \sum_{x \in \Lambda} \left[ \alpha(x) + \liminf_{n \to \infty} [-a_n(x)] \right]
= \sum_{x \in \Lambda} \left[ \alpha(x) - \limsup_{n \to \infty} a_n(x) \right] = \sum_{x \in \Lambda} \alpha(x).
\]
This then implies that
\[
\sum_{x \in A} a(x) \leq \sum_{x \in A} a(x) - \limsup_{n \to \infty} \left[ \sum_{x \in A} a_n(x) \right]
\]
and hence that
\[
\limsup_{n \to \infty} \left[ \sum_{x \in A} a_n(x) \right] \leq 0
\]
as before.

Now suppose that \((Z, \|\cdot\|)\) is a Banach space and \(a : A \to Z\) is a function such that \(\sum_{x \in A} \|a(x)\| < \infty\).

**Theorem 8.10.** Let \(\ell^1(A, Z)\) be the space of functions \(a : A \to Z\) such that
\[
\|a\|_1 := \sum_{x \in A} \|a(x)\| < \infty.
\]
Then:
1. \((\ell^1(A, Z), \|\cdot\|_1)\) is a Banach space.
2. The set \(D = \{a : A \to Z : \#(\{x \in A : a(x) \neq 0\}) < \infty\}\) is dense subspace of \(\ell^1(A, Z)\).
3. For \(a \in D\) we may define
\[
\sum_{x \in A} a(x) = \sum_{x : a(x) \neq 0} a(x) \in Z.
\]
Since
\[
\left\| \sum_{x \in A} a(x) \right\| \leq \sum_{x \in A} \|a(x)\| = \|a\|_1,
\]
this linear operator has a unique continuous extension to a linear operator \(\Sigma : \ell^1(A, Z) \to Z\).

**Proof.** We take each item in turn.

1. Let \(\{a_n\}_{n=1}^{\infty}\) be a Cauchy sequence in \(\ell^1(A, Z)\). Since
\[
\|a_n(x) - a_m(x)\| \leq \|a_n - a_m\|_1 \to 0 \quad \text{as} \quad m, n \to \infty,
\]
it follows that \(\lim_{n \to \infty} a_n(x) =: a(x)\) exists for all \(x \in A\). Moreover, using Fatou's Lemma 8.8,
\[
\|a - a_n\|_1 = \sum_{x \in A} \|a(x) - a_n(x)\|
\]
\[
= \sum_{x \in A} \liminf_{m \to \infty} \|a_m(x) - a_n(x)\|
\]
\[
\leq \liminf_{m \to \infty} \sum_{x \in A} \|a_m(x) - a_n(x)\|
\]
\[
= \liminf_{m \to \infty} \|a_m - a_n\|_1 \to 0 \quad \text{as} \quad n \to \infty.
\]

2. Choose \(A_n \subset_f A\) such that \(A_n \uparrow A\) and set \(a_n(x) = a(x)1_{A_n}(x)\) for all \(n \in \mathbb{N}\) and \(x \in A\). Then
\[
\lim_{n \to \infty} \|a - a_n\|_1 = 0 \quad \text{by DCT}.
\]

3. This is a consequence of the BLT theorem.

**Theorem 8.11 (DCT II).** Suppose that \(\{a_n\}_{n=1}^{\infty} \subset \ell^1(A, Z)\) and \(A(x) := \sup_n \|a_n(x)\|\) satisfy:
1. \(a(x) := \lim_{n \to \infty} a_n(x)\) existing in \(Z\) for all \(x \in A\), \([\text{Conflict of notation}]\)
2. \(\sum_{x \in A} A(x) < \infty\),

then
\[
\lim_{n \to \infty} \|a - a_n\|_1 = 0 \quad \text{and} \quad \lim_{n \to \infty} \sum_{x \in A} a_n(x) = \sum_{x \in A} a(x).
\]

**Proof.** It suffices to prove the first assertion since the sum operation is continuous relative to the \(\|\cdot\|_1\) norm on \(\ell^1(A, Z)\). Since
\[
\|a(x)\| = \lim_{n \to \infty} \|a_n(x)\| \leq A(x)
\]
it follows that
\[
\|a(x) - a_n(x)\| \leq \|a(x)\| + \|a_n(x)\| \leq 2A(x).
\]

Since
\[
\lim_{n \to \infty} \|a(x) - a_n(x)\| = \|a(x) - \lim_{n \to \infty} a_n(x)\| = \|a(x) - a(x)\| = 0 \quad \text{for all} \quad x \in A,
\]
we may use the dominated convergence theorem for sums (Theorem 8.9) to conclude,
\[
\lim_{n \to \infty} \|a - a_n\|_1 = \lim_{n \to \infty} \sum_{x \in A} \|a(x) - a_n(x)\| = \sum_{x \in A} \lim_{n \to \infty} \|a(x) - a_n(x)\| = 0.
\]

**Corollary 8.12.** Suppose \((Z, \|\cdot\|)\) and \(a \in \ell^1(A, Z)\). If \(A_n \uparrow A\), then
\[
\lim_{n \to \infty} \sum_{x \in A_n} a(x) = \sum_{x \in A} a(x).
\]
Proof. Let \( a_n (x) := 1_{A_n} (x) \) for all \( n \). Then \( \|a_n (x)\| \leq \|a (x)\| \) and \( \sum_{x \in A} \|a (x)\| < \infty \). Since \( \lim_{n \to \infty} a_n (x) = a (x) \) for all \( x \in A \), it follows by DCT that

\[
\lim_{n \to \infty} \sum_{x \in A_n} a (x) = \lim_{n \to \infty} \sum_{x \in A} a_n (x) = \sum_{x \in A} \lim_{n \to \infty} a_n (x) = \sum_{x \in A} a (x).
\]

Corollary 8.13. Suppose \((Z, \|\|)\) is a Banach space, \( A \) is a countable set, and \( a \in \ell^1 (A, Z) \). If \( A = \bigcup_{n=1}^{\infty} A_n \), then

\[
\sum_{n=1}^{\infty} a (x) = \sum_{n=1}^{\infty} \sum_{x \in A_n} a (x).
\]

Proof. We have

\[
\sum_{n=1}^{\infty} a (x) = \lim_{N \to \infty} \sum_{n=1}^{N} \sum_{x \in A_n} a (x) = \lim_{N \to \infty} \sum_{x \in \bigcup_{n=1}^{N} A_n} a (x) = \sum_{x \in A} a (x).
\]

because \( \bigcup_{n=1}^{N} A_n \uparrow A \) as \( N \to \infty \). [Bruce: we need to prove the finite additivity here, namely that if \( A = A \cup B \), then

\[
\sum_{x \in A} a (x) = \sum_{x \in A} a (x) + \sum_{x \in B} a (x).
\]

But this is simple, since \( a = 1_A a + 1_B a \) and

\[
\sum_{x \in A} 1_A a = \sum_{x \in A} a
\]

where this last equality follow by choosing \( A_n \subset_f A \) such that \( A_n \uparrow A \) so that

\[
\sum_{x \in A} 1_A a = \lim_{n \to \infty} \sum_{x \in A_n} 1_A a = \lim_{n \to \infty} \sum_{x \in A_n \cap A} a = \sum_{x \in A} a.
\]

Corollary 8.14. If \( \sum_{n=1}^{\infty} \sum_{x \in A} \|a_n (x)\| < \infty \), then

\[
\sum_{n=1}^{\infty} \sum_{x \in A} a_n (x) = \sum_{x \in A} \lim_{n \to \infty} a_n (x) \in Z.
\]

Proof. First observe that

\[
\left\| \sum_{n=1}^{N} a_n (x) \right\| \leq \sum_{n=1}^{N} \|a_n (x)\| \leq \sum_{n=1}^{\infty} \|a_n (x)\| \in \ell^1 (A, Z)
\]

and \( \lim_{N \to \infty} \sum_{n=1}^{N} a_n (x) = \sum_{n=1}^{\infty} a_n (x) \). Therefore by DCT,

\[
\sum_{n=1}^{\infty} a_n (x) = \lim_{N \to \infty} \sum_{n=1}^{N} a_n (x) = \lim_{N \to \infty} \sum_{x \in A} \sum_{n=1}^{N} a_n (x)
\]

\[
= \sum_{x \in A} \lim_{N \to \infty} \sum_{n=1}^{N} a_n (x) = \sum_{x \in A} \sum_{n=1}^{\infty} a_n (x).
\]

Better proof. By assumption,

\[
\sum_{n=1}^{\infty} \|a_n\|_1 < \infty \implies \sum_{n=1}^{\infty} a_n \text{ converges in } \ell^1 (A, Z).
\]

Therefore,

\[
\sum_{n=1}^{\infty} \sum_{x \in A} a_n = \sum_{n=1}^{\infty} \sum_{x \in A} a_n.
\]

Exercise 8.1 (Differentiating past an infinite sum). Suppose that

\[
\sum_{n=0}^{\infty} \sup_{x} |f_n' (x)| < \infty \text{ and } \sum_{n=0}^{\infty} f_n (0) \text{ exists.}
\]

Then \( S (x) := \sum_{n=0}^{\infty} f_n (x) \) exists and

\[
S' (x) = \sum_{n=0}^{\infty} f_n' (x).
\]

Exercise 8.2 (Differentiating past a limit). Suppose \( \lim_{n \to \infty} f_n (x) = f (x) \) and \( f_n' \to g \) uniformly. Show \( f' (x) = g (x) \), i.e. we have in this case that

\[
\frac{d}{dx} \lim_{n \to \infty} f_n (x) = \lim_{n \to \infty} \frac{d}{dx} f_n (x).
\]

Exercise 8.3. Suppose that

\[
f (z) := \sum_{n=0}^{\infty} a_n z^n
\]

has radius of convergence \( R \). Show \( f' (z) = \sum_{n=0}^{\infty} n a_n z^{n-1} \) for all \( |z| < R \) and the radius of convergence of \( f' \) is still \( R \).
Theorem 8.15 (Changing the center of a power series). Suppose that
\[ A(z) = \sum_{n=0}^{\infty} \frac{a_n}{n!} z^n \] has radius of convergence \( R > 0 \). For \( |z| < R \) and \( |h| < R - |z| \) we have
\[ A(z + h) = \sum_{m=0}^{\infty} \frac{1}{m!} a_m (z) h^m \]
where
\[ a_m(z) := \sum_{n=m}^{\infty} \frac{a_n}{(n-m)!} z^{n-m} = A^{(m)}(z). \]

Proof. Working formally for the moment,
\[ A(z + h) = \sum_{n=0}^{\infty} \frac{a_n}{n!} (z + h)^n = \sum_{n=0}^{\infty} \frac{a_n}{n!} \sum_{m=0}^{n} \binom{n}{m} z^{n-m} h^m \]
\[ = \sum_{0 \leq m \leq n < \infty} \frac{a_n}{m! \cdot (n-m)!} z^{n-m} h^m \]
\[ = \sum_{m=0}^{\infty} \frac{1}{m!} \left[ \sum_{n=m}^{\infty} \frac{a_n}{(n-m)!} z^{n-m} \right] h^m \]
which gives the result provided the interchange of order of sums was permissible. To check this we consider,
\[ \sum_{0 \leq m \leq n < \infty} \left| \frac{a_n}{m! \cdot (n-m)!} z^{n-m} h^m \right| = \sum_{0 \leq m \leq n < \infty} \frac{|a_n|}{m! \cdot (n-m)!} |z|^{n-m} |h|^m \]
\[ = \sum_{n=0}^{\infty} \frac{|a_n|}{n!} (|z| + |h|)^n < \infty \]
as we know for all \( 0 \leq \rho < R \) that \( \sum_{n=0}^{\infty} \frac{|a_n|}{n!} \rho^n < \infty \). Hence we may apply Fubini’s theorem for sums in order to conclude the result. \( \blacksquare \)

Definition 8.16. Suppose that \( \{S_{m,n}\}_{m,n=1}^{\infty} \) is a sequence of complex numbers (or more generally elements of a metric space). We say \( \lim_{m,n \to \infty} S_{m,n} = L \) iff for all \( \epsilon > 0 \) there exists \( N \in \mathbb{N} \) such that
\[ |L - S_{m,n}| \leq \epsilon \quad \text{for all} \quad m \land n \geq N. \]
We say \( \lim_{m \lor n \to \infty} S_{m,n} = L \) exists iff for all \( \epsilon > 0 \) there exists \( N \in \mathbb{N} \) such that
\[ |L - S_{m,n}| \leq \epsilon \quad \text{for all} \quad m \lor n \geq N. \]
Clearly \( \lim_{m \lor n \to \infty} S_{m,n} = L \) implies \( \lim_{m \land n \to \infty} S_{m,n} = L \) but the converse is not true. We are mostly interested in finding sufficient conditions in order for iterated limits to be equal, i.e. for
\[ \lim_{m \to \infty} \lim_{n \to \infty} S_{m,n} = \lim_{n \to \infty} \lim_{m \to \infty} S_{m,n}. \]

Example 8.17 (Switching Limits is Dangerous I). If \( S_{m,n} = \frac{1}{1 + \frac{x}{n}} \), then \( \lim_{m \to \infty} S_{m,n} = 0 \) (but not uniformly in \( n \)) so that \( \lim_{n \to \infty} \lim_{m \to \infty} S_{m,n} = 0 \) while \( \lim_{n \to \infty} S_{m,n} = 1 \) and \( \lim_{m \to \infty} \lim_{n \to \infty} S_{m,n} = 1. \) In order to visualize better what is going on here let us make the change of variables, \( x = \frac{1}{m} \) and \( y = \frac{1}{n} \), i.e. let
\[ S(x, y) = S_{m,n} = \frac{1}{1 + \frac{x}{y}} = \frac{x}{x + y} \]
whose plot appears in Figure 8.17.

8.2 Double Sequences

In this chapter we will consider doubly indexed sequences, \( \{S_{m,n}\}_{m,n=1}^{\infty} \), of complex numbers\(^1\) To be more precise \( \{S_{m,n}\}_{m,n=1}^{\infty} \) is simply a function from \( \mathbb{N}^2 \) to \( \mathbb{C} \). In this chapter we are interested in the following limits;
\[ \lim_{m \to \infty} \lim_{n \to \infty} S_{m,n}, \quad \lim_{n \to \infty} \lim_{m \to \infty} S_{m,n}, \quad \lim_{m \to \infty} \lim_{n \to \infty} S_{m,n}, \quad \text{and} \quad \lim_{m \land n \to \infty} S_{m,n}, \]
where the last two limits are defined as follows.

\(^1\) Much of what we will say holds for sequences taking values in “complete metric spaces” to be covered later.
A plot of \( S_{m,n} = \frac{(-1)^{m+1}}{1 + \frac{m}{n}} \) in terms of the variables \( \frac{1}{m} \) and \( \frac{1}{n} \).

With this change of variables, \( m \to \infty \iff x \to 0 \) and \( n \to \infty \iff y \to 0 \) and \( m = \infty \) and \( n = \infty \) correspond to \( x = 0 \) and \( y = 0 \) respectively. It is now quite clearly that \( \lim_{x \to 0} A(x,y) = 0 \) for all \( y > 0 \) and \( \lim_{y \to 0} A(x,y) = 1 \) for all \( x > 1 \).

**Example 8.18 (Switching Limits is Dangerous II).** Let

\[
S_{m,n} := \frac{(-1)^{m+1}}{1 + \frac{m}{n}} = -\cos \left( \frac{\pi m}{1 + \frac{m}{n}} \right).
\]

Then \( \lim_{m \to \infty} S_{m,n} = 0 \) (but not uniformly in \( n \)) so that \( \lim_{n \to \infty} \lim_{m \to \infty} S_{m,n} = 0 \). We also have \( \lim_{n \to \infty} S_{m,n} = (-1)^m \) and \( \lim_{m \to \infty} \lim_{n \to \infty} S_{m,n} = \lim_{m \to \infty} (-1)^m \) does not exists. In order to visualize better what is going on here let us again make the change of variables, \( x = \frac{1}{m} \) and \( y = \frac{1}{n} \), i.e. let

\[
S(x,y) = S_{m,n} = -x \cos \left( \frac{\pi}{x+y} \right) \cos \left( \frac{\pi}{x+y} \right)
\]

whose plot appears in Figure 8.18.

**Example 8.19 (Switching Limits is Dangerous III).** If

\[
S_{m,n} := \frac{(-1)^n}{m} + \frac{(-1)^m}{n} = -\frac{\cos (n\pi)}{m} - \frac{\cos (m\pi)}{n}
\]

and we let \( x = \frac{1}{m} \) and \( y = \frac{1}{n} \). Then

\[
S(x,y) = S_{m,n} = -x \cos (\pi/x) - y \cos (\pi/y)
\]

whose plot appears in Figure 8.19.
A plot of $\frac{-\cos(nx)}{m} - \frac{-\cos(mx)}{n}$ in terms of the variables $\frac{1}{m}$ and $\frac{1}{n}$.

In this case, $\lim_{m\land n\to\infty} S_{m,n} = 0$ exists but $\lim_{m\to\infty} S_{m,n}$ and $\lim_{n\to\infty} S_{m,n}$ do not exist!

Remark 8.20. However, if $\lim_{m\land n\to\infty} S_{m,n}$ exists then we do always have,

$$\lim_{m\land n\to\infty} S_{m,n} = \lim_{m\to\infty} \limsup_{n\to\infty} S_{m,n} = \lim_{n\to\infty} \liminf_{m\to\infty} S_{m,n}.$$ 

On the other hand if $\lim_{m\lor n\to\infty} S_{m,n}$ exists then we will have

$$\lim_{m\lor n\to\infty} S_{m,n} = \lim_{m\to\infty} \lim_{n\to\infty} S_{m,n} = \lim_{n\to\infty} \lim_{m\to\infty} S_{m,n}.$$ 

One way to avoid these types of behaviors is to assume $S_{m,n} \geq 0$ and is increasing in each index.

Example 8.21 (Monotinicity is good). If

$$S_{m,n} := \frac{1}{1 + \frac{1}{m} + \frac{1}{n^2}}$$

and we let $x = \frac{1}{m}$ and $y = \frac{1}{n}$. Then

$$S(x, y) = S_{m,n} = \frac{1}{1 + y + x^2}$$

whose plot appears in Figure 8.21.

This example is covered by Theorem 8.22 below.

**Theorem 8.22 (Equality of monotone iterated limits).** Suppose that $S_{m,n} \geq 0$ for all $m,n \in \mathbb{N}$ and $S_{m+1,n} \geq S_{m,n}$ and $S_{m,n+1} \geq S_{m,n}$ for all $m,n \in \mathbb{N}$. Then

$$L := \lim_{n\to\infty} \lim_{m\to\infty} S_{m,n} = \lim_{m\to\infty} \lim_{n\to\infty} S_{m,n} = \sup_{(m,n)} S_{m,n}$$

where all limits exist but may take on the value infinity. Moreover $\lim_{m\land n\to\infty} S_{m,n} = L$.

**Proof.** Because $S_{m,n}$ is increasing in each of its variables, we find,

$$\lim_{n\to\infty} \lim_{m\to\infty} S_{m,n} = \sup_{m\in\mathbb{N}} \lim_{n\to\infty} S_{m,n} \text{ and } \lim_{m\to\infty} \lim_{n\to\infty} S_{m,n} = \sup_{m\in\mathbb{N}} \sup_{n\in\mathbb{N}} S_{m,n}$$

and therefore Eq. (8.2) follows from Corollary 8.2 of the sup-sup Theorem. So it only remains to show $\lim_{m\land n\to\infty} S_{m,n} = L$.

**Cases 1.** If $L = \infty$, then for any $N \in \mathbb{N}$, there exists $(m_N, n_N)$ such that $S_{m_N,n_N} \geq N$ and since $S_{m,n}$ is increasing in each of its variables it follows that

$$S_{m,n} \geq N \text{ for all } m \geq m_N \text{ and } n \geq n_N.$$
and it follows that \( \lim_{m,n \to \infty} S_{m,n} = \infty \).

Case 2. \( L < \infty \). In this case if \( \varepsilon > 0 \) is given, there exists \((m_\varepsilon, n_\varepsilon)\) such that 
\[ S_{m_\varepsilon,n_\varepsilon} \geq L - \varepsilon. \] Since \( S_{m,n} \) is increasing in each of its variables it follows that 
\[ L \geq S_{m,n} \geq L - \varepsilon \text{ for all } m \geq m_\varepsilon \text{ and } n \geq n_\varepsilon \]
and so \( |L - S_{m,n}| \leq \varepsilon \) for all \( m, n \geq m_\varepsilon \wedge n_\varepsilon \). So by definition, \( \lim_{m,n \to \infty} S_{m,n} = L \).

**Exercise 8.4.** If \( \{f_n(x)\}_{n=1}^{\infty} \) is a sequence of increasing continuous functions on \( \mathbb{R} \) such that \( f_n(x) \uparrow f(x) < \infty \) as \( n \to \infty \), then \( f(x) \) is continuous and increasing as well.

**Exercise 8.5.** Show \( \lim_{m,n \to \infty} S_{m,n} = L \) iff for all \( \varepsilon > 0 \) there exists \( M, N \in \mathbb{N} \) such
\[ |L - S_{m,n}| \leq \varepsilon \text{ for all } m \geq M \text{ and } n \geq N. \]

**Lemma 8.23.** Suppose that \( \{S_{m,n}\}_{m,n=1}^{\infty} \) is Cauchy in the sense that for all \( \varepsilon > 0 \) there exists \( M, N \in \mathbb{N} \) such that
\[ |S_{m,n} - S_{m',n'}| \leq \varepsilon \text{ for all } m, m' \geq M \text{ and } n, n' \geq N. \] Then \( \lim_{m,n \to \infty} S_{m,n} = L \) exists.

**Proof.** Let \( s_n := S_{n,n} \). Then the assumption shows that \( \{s_n\}_{n=1}^{\infty} \) is a Cauchy sequence and hence convergent. Let \( L := \lim_{n \to \infty} s_n \). Now take \( m' = n' \) and then let \( n' \to \infty \) in Eq. (8.3) in order to learn,
\[ |S_{m,n} - L| \leq \varepsilon \text{ for all } m \geq M \text{ and } n \geq N. \]
From this we conclude that \( \lim_{m,n \to \infty} S_{m,n} = L \).

Another way to ensure that the iterated limits are equal is to assume some uniformity in one of the limits as in the next key theorem. [This theorem will be used in one guise or another repeatedly throughout these notes.]

**Theorem 8.24.** Suppose that \( \{S_{m,n}\}_{m,n=1}^{\infty} \) is a sequence of complex numbers (or more generally elements of a complete metric space \((X, \rho)\)). Assume that
\[ S_{m,\infty} := \lim_{n \to \infty} S_{m,n} \text{ exists uniformly in } m \text{ and} \]
\[ S_{\infty,n} := \lim_{m \to \infty} S_{m,n} \text{ exists pointwise in } n. \]
Then \( \lim_{m \to \infty} \lim_{n \to \infty} S_{m,n} \), \( \lim_{n \to \infty} \lim_{m \to \infty} S_{m,n} \), and \( \lim_{m,n \to \infty} S_{m,n} \) all exist and are equal, i.e.
\[ L := \lim_{m \to \infty} \lim_{n \to \infty} S_{m,n} = \lim_{n \to \infty} \lim_{m \to \infty} S_{m,n} = \lim_{m,n \to \infty} S_{m,n}. \]
[In words, the existence of limits in both variables along with uniformity in one of the variables implies the iterated limits exists and agree.]

**Proof.** Let \( \varepsilon > 0 \) be given. Choose \( N \in \mathbb{N} \) such that
\[ \sup_m |S_{m,\infty} - S_{m,n}| \leq \varepsilon \text{ for all } n \geq N. \]
Now choose \( M \in \mathbb{N} \) such that
\[ |S_{\infty,N} - S_{m,n}| \leq \varepsilon \text{ for all } m \geq M. \]
Then for \( n \geq N \) and \( m \geq M \) we have,
\[ |S_{m,n} - S_{m,\infty} - S_{\infty,N}| \leq |S_{m,n} - S_{m,\infty}| + |S_{m,\infty} - S_{\infty,N}| \]
\[ \leq |S_{m,n} - S_{m,\infty}| + |S_{m,\infty} - S_{m,n}| + |S_{m,n} - S_{\infty,N}| \]
\[ \leq \epsilon + \epsilon + \epsilon = 3\epsilon. \]
Therefore it follows that
8.3 Iterated Limits

Theorem 8.26. Let \( \{a_{m,n}\}_{m,n=1}^\infty \) be a sequence of complex numbers and assume that

\[
\lim_{m \to \infty} a_{m,n} = A_n \quad \text{exists uniformly in } n \\
\lim_{n \to \infty} A_n = L \quad \text{exists}
\]

then \( \lim_{m,n \to \infty} a_{m,n} = L \) and in particular,

\[
\lim_{m \to \infty} \lim_{n \to \infty} a_{m,n} = \lim_{n \to \infty} \lim_{m \to \infty} a_{m,n}.
\]

Idea. The first assumption guarantees the rows of \( a_{m,*} \cong A \) for large \( m \). The second assertion say that \( A_n \cong L \) for large \( n \). Thus we must have \( a_{m,n} \cong A_n \cong L \) for large \( m \) and \( n \).

Theorem 8.27. Let \( \{a_{m,n}\}_{m,n=1}^\infty \) be a sequence of complex numbers and assume that

\[
\lim_{m \to \infty} a_{m,n} = a_n \quad \text{exists uniformly in } n \\
\lim_{n \to \infty} a_{m,n} = b_m \quad \text{exists} \\
\lim_{m \to \infty} b_m = B \quad \text{exists}.
\]

Then, \( \lim_{n \to \infty} a_n = B \). In other words under the above conditions,

\[
\lim_{m \to \infty} \lim_{n \to \infty} a_{m,n} = \lim_{n \to \infty} \lim_{m \to \infty} a_{m,n}.
\]

Moreover \( \lim_{m,n \to \infty} a_{m,n} = B \) as well.

Idea. As above we know that \( a_{m,*} \cong a \) for \( m \) large. We are also given that \( \lim_{n \to \infty} a_{m,n} \cong B \) for \( m \) large. Thus \( B \cong \lim_{n \to \infty} a_{m,n} \cong \lim_{n \to \infty} a_n \) for \( m \) large. Hence we may now apply the previous theorem.

[Details] We have

\[
|B - a_n| \leq |B - b_m| + |b_m - a_{m,n}| + \sup_k |a_{m,k} - a_k|
\]

and then taking \( \lim \sup_{n \to \infty} \) of this inequality shows,

\[
\lim_{n \to \infty} \sup_k |B - a_n| \leq |B - b_m| + \sup_k |a_{m,k} - a_k| \to 0 \quad \text{as } m \to \infty.
\]

We now prove the second assertion. For this we have,

\[
|B - a_{m,n}| \leq |B - a_n| + |a_n - a_{m,n}| \\
\leq |B - a_n| + \sup_k |a_k - a_{m,k}|.
\]

Thus given \( \varepsilon > 0 \) there exists \( M \in \mathbb{N} \) and \( N \in \mathbb{N} \) such that

\[
|B - a_{m,n}| \leq |B - a_n| + \sup_k |a_k - a_{m,k}| \leq \varepsilon + \varepsilon
\]

for \( n \geq N \) and \( m \geq M \). This proves the stronger statement that \( \lim_{m,n \to \infty} a_{m,n} = B \), i.e. \( a_{m,n} \) is near \( B \) as long as both \( m, n \) are sufficiently large.

Exercise 8.6. Suppose that \( f_n \to f \) uniformly and \( f_n \) is continuous for all \( n \), then \( f \) is continuous. [Use sequential notion of continuity here.]
8.4 Double Sums and Continuity of Sums

Here are a couple of very useful consequences of these theorems.

**Theorem 8.28 (Monotone convergence theorem for sums).** Let \( \{a_{k,n}\}_{k,n=1}^{\infty} \) be a sequence of non-negative numbers, assume that \( a_{k,n+1} \geq a_{k,n} \) for all \( k,n \). Then

\[
\lim_{n \to \infty} \sum_{k=1}^{\infty} a_{k,n} = \sum_{k=1}^{\infty} \lim_{n \to \infty} a_{k,n} = \lim_{m \to \infty} \sum_{k=1}^{m} a_{k,n}.
\]

**Proof.** Let \( S_{m,n} := \sum_{k=1}^{m} a_{k,n} \), then \( \{S_{m,n}\}_{m,n=1}^{\infty} \) satisfies the hypothesis of Theorem 8.22 and the conclusions now follow from that Theorem upon noting that

\[
\lim_{n \to \infty} \sum_{k=1}^{\infty} a_{k,n} = \lim_{n \to \infty} S_{n,m} \quad \text{and} \quad \lim_{n \to \infty} a_{k,n} = \lim_{m \to \infty} S_{m,n}.
\]

**Theorem 8.29 (Tonelli’s Theorem for sums).** Let \( \{a_{k,l}\}_{k,l=1}^{\infty} \) be any sequence of non-negative numbers, then

\[
\sum_{k=1}^{\infty} \sum_{l=1}^{\infty} a_{k,l} = \sum_{l=1}^{\infty} \sum_{k=1}^{\infty} a_{k,l}.
\]

**Proof.** Apply Theorem 8.22 with \( S_{m,n} := \sum_{k=1}^{m} \sum_{l=1}^{n} a_{k,l} \).

**Theorem 8.30 (Dominated convergence theorem for sums).** Let \( \{a_{k,n}\}_{k,n=1}^{\infty} \) be a sequence of complex numbers such that \( \lim_{n \to \infty} a_{k,n} = a_k \) exists for all \( n \) and there exists a summable dominating sequence, \( \{M_k\} \), such that \( |a_{k,n}| \leq M_k \) for all \( k \), \( n \in \mathbb{N} \). Then

\[
\lim_{n \to \infty} \sum_{k=1}^{\infty} a_{k,n} = \sum_{k=1}^{\infty} \lim_{n \to \infty} a_{k,n} = \lim_{m,n \to \infty} \sum_{k=1}^{m} a_{k,n}.
\]

**Proof.** Let \( S_{m,n} := \sum_{k=1}^{m} a_{k,n} \), then \( \{S_{m,n}\}_{m,n=1}^{\infty} \) satisfies the hypothesis of Theorem 8.27. Indeed,

\[
\left| S_{m,n} - \sum_{k=1}^{m} a_{k,n} \right| = \sum_{k=m+1}^{\infty} |a_{k,n}| \leq \sum_{k=m+1}^{\infty} M_k
\]

and hence,

\[
\sup_{n} \left| S_{m,n} - \sum_{k=1}^{m} a_{k,n} \right| \leq \sum_{k=m+1}^{\infty} M_k \to 0 \quad \text{as} \quad m \to \infty.
\]

Therefore \( S_{m,n} \to \sum_{k=1}^{\infty} a_{k,n} \) uniformly in \( n \) as \( m \to \infty \). Moreover,

\[
\lim_{n \to \infty} S_{m,n} = \sum_{k=1}^{\infty} \lim_{n \to \infty} a_{k,n} = \sum_{k=1}^{\infty} a_{k} \quad \text{as} \quad m \to \infty.
\]

Thus the hypothesis of Theorem 8.27 are satisfied and so we may conclude,

\[
\lim_{m \to \infty} \lim_{n \to \infty} S_{m,n} = \lim_{n \to \infty} \lim_{m \to \infty} S_{m,n} = \lim_{n \to \infty} S_{m,n} = \sum_{k=1}^{\infty} a_{k,}\]

which is exactly Eq. (8.4).

**Theorem 8.31 (Fubini’s Theorem for sums).** Let \( \{a_{k,l}\}_{k,l=1}^{\infty} \) be any sequence of complex numbers. If \( \sum_{k,l=1}^{\infty} |a_{k,l}| < \infty \), then

\[
\sum_{k=1}^{\infty} \sum_{l=1}^{\infty} a_{k,l} = \sum_{k,l=1}^{\infty} a_{k,l} = \sum_{l=1}^{\infty} \sum_{k=1}^{\infty} a_{k,l}.
\]

8.5 Sums of positive functions

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

**Definition 8.32.** Suppose that \( a : X \to [0, \infty] \) is a function and \( F \subset X \) is a subset, then

\[
\sum_{F} a = \sum_{x \in F} a(x) := \sup \left\{ \sum_{x \in \alpha} a(x) : \alpha \subset_f F \right\}.
\]

**Remark 8.33.** Suppose that \( X = \{1, 2, 3, \ldots\} \) and \( a : X \to [0, \infty] \), then

\[
\sum_{X} a = \sum_{n=1}^{\infty} a(n) := \lim_{N \to \infty} \sum_{n=1}^{N} a(n).
\]

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

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

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

Taking the supremum over $\alpha$ in the previous equation shows

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

**Remark 8.34.** Suppose $a : X \to [0, \infty]$ and $\sum_{\alpha} a < \infty$, then $\{x \in X : a(x) > 0\}$ is at most countable. To see this first notice that for any $\varepsilon > 0$, the set $\{x : a(x) \geq \varepsilon\}$ must be finite for otherwise $\sum_{\alpha} a = \infty$. Thus

$$\{x \in X : a(x) > 0\} = \bigcup_{k=1}^{\infty} \{x : a(x) \geq 1/k\}$$

which shows that $\{x \in X : a(x) > 0\}$ is a countable union of finite sets and thus countable by Lemma 8.33.

**Lemma 8.35.** Suppose that $a, b : X \to [0, \infty]$ are two functions, then

$$\sum_{X} (a + b) = \sum_{X} a + \sum_{X} b$$

and

$$\sum_{X} \lambda a = \lambda \sum_{X} a$$

for all $\lambda \geq 0$.

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

$$\sum_{\alpha} (a + b) = \sum_{\alpha} a + \sum_{\alpha} b \leq \sum_{X} a + \sum_{X} b$$

which after taking sups over $\alpha$ shows that

$$\sum_{X} (a + b) \leq \sum_{X} a + \sum_{X} b.$$

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

$$\sum_{\alpha} a + \sum_{\beta} b \leq \sum_{\alpha \cup \beta} a + \sum_{\alpha \cup \beta} b = \sum_{\alpha \cup \beta} (a + b) \leq \sum_{X} (a + b).$$

Taking sups over $\alpha$ and $\beta$ then shows that

$$\sum_{X} a + \sum_{X} b \leq \sum_{X} (a + b).$$

**Lemma 8.36.** Let $X$ and $Y$ be sets, $R \subset X \times Y$ and suppose that $a : R \to \tilde{\mathbb{R}}$ is a function. Let $x R := \{y \in Y : (x, y) \in R\}$ and $R_a := \{x \in X : (x, y) \in R\}$. Then

$$\sup_{(x, y) \in R} a(x, y) = \sup_{x \in X} \sup_{y \in x R} a(x, y) = \sup_{x \in X} a(x) \text{ and}$$

$$\inf_{(x, y) \in R} a(x, y) = \inf_{x \in X} \inf_{y \in x R} a(x, y) = \inf_{x \in X} a(x),$$

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

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

$$\sup_{x \in X} \sup_{y \in x R} a(x, y) = \sup_{x \in X} N_x \leq M. \quad (8.5)$$

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

$$a(x, y) \leq N_x \leq N_x = \sup \sup_{x \in X} a(x, y)$$

and therefore

$$M = \sup_{(x, y) \in R} a(x, y) \leq \sup_{x \in X} \sup_{y \in x R} a(x, y). \quad (8.6)$$

Equations (8.5) and (8.6) show that

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

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

**Theorem 8.37 (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 proofs.

**First proof.** Let $2^X := \{A \subset X : A \subset_f X\}$.
Second Proof. Let $S_n = \sum_X f_n$ and $S = \sum_X f$. Since $f_n \leq f_m \leq f$ for all $n \leq m$, it follows that $S_n \leq S_m \leq S$, which shows that $\lim_{n \to \infty} S_n$ exists and is less that $S$, i.e.

$$A := \lim_{n \to \infty} \sum_X f_n \leq \sum_X f. \quad (8.7)$$

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

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

Letting $n$ tend to infinity in this equation shows that

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

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

$$\sum_X f \leq A \leq \lim_{n \to \infty} \sum_X f_n \quad (8.8)$$

which combined with Eq. (8.7) proves the theorem.

**Lemma 8.38 (Fatou’s Lemma for Sums).** Suppose that $f_n : X \to [0, \infty]$ is a sequence of functions, then

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

**Proof.** Define $g_k := \inf_{n \geq k} f_n$ so that $g_k \uparrow \lim_{n \to \infty} f_n$ as $k \to \infty$. Since $g_k \leq f_n$ for all $n \geq k$,

$$\sum_X g_k \leq \sum_X f_n \text{ for all } n \geq k$$

and therefore

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

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

$$\sum_X \liminf_{n \to \infty} f_n = \sum_X \lim_{k \to \infty} g_k = \lim_{k \to \infty} \sum_X g_k \leq \liminf_{n \to \infty} \sum_X f_n.$$

**Remark 8.39.** If $A = \sum_X a < \infty$, then for all $\varepsilon > 0$ there exists $\alpha_\varepsilon \subset f X$ such that

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

for all $\alpha \subset f X$ containing $\alpha_\varepsilon$ or equivalently,

$$|A - \sum_\alpha a| \leq \varepsilon \quad (8.9)$$

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

### 8.6 Sums of complex functions

**Definition 8.40.** 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 $\alpha \subset f X$ containing $\alpha_\varepsilon$ we have

$$|A - \sum_\alpha a| \leq \varepsilon.$$
The following lemma is left as an exercise to the reader.

**Lemma 8.41.** 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 8.42 (Summable).** We call a function $a : X \to \mathbb{C}$ **summable** if

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

**Proposition 8.43.** 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

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

**Proof.** If $\sum_X |a| < \infty$, then $\sum_X (\text{Re } a)^\pm < \infty$ and $\sum_X (\text{Im } a)^\pm < \infty$ and hence by Remark [8.39] these sums exists in the sense of Definition [8.40]. Therefore by Lemma [8.41] $\sum_X a$ exists and

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

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

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

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

$$a^+(x) = \max(a(x), 0) \text{ and } a^-(x) = \max(-a(x), 0). \quad (8.10)$$

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 $X' := \{ x \in X : a(x) \geq 0 \}$, then we know that $\sum_{X'} a = \infty$ which means there are finite subsets $\alpha_n \subset X' \subset X$ such that $\sum_{\alpha_n} a \geq n$ for all $n$. Thus if $\alpha \subset f X$ is any finite set, it follows that

$$\lim_{n \to \infty} \sum_{\alpha_n} a = \infty,$$

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 $\rho \geq 0$ and $\theta \in \mathbb{R}$, then

$$\left| \sum_X a \right| = \rho = e^{-i\theta} \sum_X a = \sum_X e^{-i\theta} a \leq \sum_X |e^{-i\theta} a| \leq \sum_X |a|.$$

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

**Remark 8.44.** Suppose that $X = \mathbb{N}$ and $a : \mathbb{N} \to \mathbb{C}$ is a sequence, then it is not necessarily true that

$$\sum_{n=1}^{\infty} a(n) = \sum_{n \in \mathbb{N}} a(n). \quad (8.11)$$

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. (8.11) is valid.

**Theorem 8.45 (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

$$|f_n(x)| \leq g(x) \text{ for all } x \in X \text{ and } n \in \mathbb{N} \quad (8.12)$$

and that $g$ is summable. Then

$$\lim_{n \to \infty} \sum_{x \in X} f_n(x) = \sum_{x \in X} f(x). \quad (8.13)$$
Proof. Notice that \( |f| = \lim |f_n| \leq g \) 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,

\[
\sum_X (g \pm f) = \sum_X \lim \inf_{n \to \infty} (g \pm f_n) \leq \lim \inf_{n \to \infty} \sum_X (g \pm f_n)
\]

\[
= \sum_X g + \lim \inf_{n \to \infty} \left( \pm \sum_X f_n \right).
\]

Since \( \lim \inf_{n \to \infty} (\pm a_n) = -\lim \sup_{n \to \infty} a_n \), we have shown,

\[
\sum_X g \pm \sum_X f \leq \sum_X g + \left\{ \lim \inf_{n \to \infty} \sum_X f_n \right\} - \lim \sup_{n \to \infty} \sum_X f_n
\]

and therefore

\[
\lim \sup_{n \to \infty} \sum_X f_n \leq \sum_X f \leq \lim \inf_{n \to \infty} \sum_X f_n.
\]

This shows that \( \lim_{n \to \infty} \sum_X f_n \) exists and is equal to \( \sum_X f \).

Proof. (Second Proof.) Passing to the limit in Eq. (8.12) shows that \( |f| \leq g \) and in particular that \( f \) is summable. Given \( \varepsilon > 0 \), let \( \alpha \subset f X \) such that

\[
\sum_{X \setminus \alpha} g \leq \varepsilon.
\]

Then for \( \beta \subset f X \) such that \( \alpha \subset \beta \),

\[
\left| \sum_{\beta} f - \sum_{\beta} f_n \right| = \left| \sum_{\beta} (f - f_n) \right|
\]

\[
\leq \sum_{\beta} |f - f_n| = \sum_{\alpha} |f - f_n| + \sum_{\beta \setminus \alpha} |f - f_n|
\]

\[
\leq \sum_{\alpha} |f - f_n| + 2 \sum_{\beta \setminus \alpha} g
\]

\[
\leq \sum_{\alpha} |f - f_n| + 2\varepsilon.
\]

and hence that

\[
\left| \sum_{\beta} f - \sum_{\beta} f_n \right| \leq \sum_{\alpha} |f - f_n| + 2\varepsilon.
\]

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

\[
\sum_X f = \sum_X f_n + \varepsilon > \sum_X f_n.
\]

which then implies that

\[
\lim \sup_{n \to \infty} \sum_X f - \sum_X f_n \leq \lim \sup_{n \to \infty} \sum_{\alpha} |f - f_n| + 2\varepsilon
\]

\[
= 2\varepsilon.
\]

Because \( \varepsilon > 0 \) is arbitrary we conclude that

\[
\lim \sup_{n \to \infty} \left| \sum_X f - \sum_X f_n \right| = 0.
\]

which is the same as Eq. (8.13). □

Remark 8.46. Theorem 8.45 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, \( |f_n| \leq g_n \) and \( \sum_X g_n \to \sum_X g \) as \( n \to \infty \). Then \( f \) is summable and Eq. (8.13) still holds. For the proof we use Fatou’s Lemma to again conclude

\[
\sum_X (g \pm f) = \sum_X \lim \inf_{n \to \infty} (g_n \pm f_n) \leq \lim \inf_{n \to \infty} \sum_X (g_n \pm f_n)
\]

\[
= \sum_X g + \lim \inf_{n \to \infty} \left( \pm \sum_X f_n \right)
\]

and then proceed exactly as in the first proof of Theorem 8.45.

8.7 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 8.47. 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,

\[
\sum_F a = \sum a.
\]

Theorem 8.48 (Tonelli’s Theorem for Sums). Suppose that \( a : X \times Y \to [0, \infty] \), then

\[
\sum_{X \times Y} a = \sum_X \sum_Y a = \sum_X \sum_X a.
\]
Proof. It suffices to show, by symmetry, that
\[
\sum_{X \times Y} a = \sum_X \sum_Y a
\]
Let \( A \subset_f X \times Y \). Then for any \( \alpha \subset_f X \) and \( \beta \subset_f Y \) such that \( A \subset \alpha \times \beta \), we have
\[
\sum_a A a \leq \sum_{\alpha \times \beta} a = \sum_{\alpha} \sum_{\beta} a \leq \sum_{X} \sum_{Y} a \leq \sum_{X \times Y} a
\]
i.e. \( \sum a \leq \sum_X \sum_Y a \). Taking the sup over \( A \) in this last equation shows
\[
\sum_{X \times Y} a \leq \sum_X \sum a.
\]
For the reverse inequality, for each \( x \in X \) choose \( \beta_{x} \subset_f Y \) such that \( \beta_{x} \uparrow Y \) as \( n \uparrow \infty \) and
\[
\sum_{y \in Y} a(x, y) = \lim_{n \to \infty} \sum_{y \in \beta_{x}} a(x, y).
\]
If \( \alpha \subset_f X \) is a given finite subset of \( X \), then
\[
\sum_{y \in Y} a(x, y) = \lim_{n \to \infty} \sum_{y \in \beta_{x}} a(x, y) \text{ for all } x \in \alpha
\]
where \( \beta_{x} := \cup_{x \in \alpha} \beta_{x} \subset_f Y \). Hence
\[
\sum_{x \in \alpha \ y \in Y} a(x, y) = \sum_{x \in \alpha} \lim_{n \to \infty} \sum_{y \in \beta_{x}} a(x, y) = \lim_{n \to \infty} \sum_{x \in \alpha} \sum_{y \in \beta_{x}} a(x, y)
\]
\[
= \lim_{n \to \infty} \sum_{(x, y) \in \alpha \times \beta_{x}} a(x, y) \leq \sum_{X \times Y} a.
\]
Since \( \alpha \) is arbitrary, it follows that
\[
\sum_{x \in X} \sum_{y \in Y} a(x, y) = \sup_{\alpha \subset_f X} \sum_{x \in \alpha \ y \in Y} a(x, y) \leq \sum_{X \times Y} a
\]
which completes the proof. \( \blacksquare \)

**Theorem 8.49 (Fubini’s Theorem for Sums).** Now suppose that \( a : X \times Y \to \mathbb{C} \) is a summable function, i.e. by Theorem 8.48 any one of the following equivalent conditions hold:
1. \( \sum_{X \times Y} |a| < \infty \).
2. \( \sum_{X} \sum_{Y} |a| < \infty \) or
3. \( \sum_Y \sum_X |a| < \infty \).

Proof. If \( a : X \to \mathbb{R} \) is real valued the theorem follows by applying Theorem 8.48 to \( a^+ \) – 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 \( a \).

### 8.8 \( \ell^p \) – spaces, Minkowski and Hölder 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
\[
\| f \|_p := \left( \sum_{x \in X} |f(x)|^p \mu(x) \right)^{1/p}
\]
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 8.50.** 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 \( f \in \mathbb{F} \).
2. (Triangle inequality) \( \| f + g \| \leq \| f \| + \| g \| \) for all \( f, g \in \mathbb{F} \).
3. (Positive definite) \( \| f \| = 0 \) implies \( f = 0 \). \( \blacksquare \)

A function \( p : Z \to [0, \infty) \) satisfying properties 1. and 2. but not necessarily 3. above will be called a semi-norm on \( Z \).

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 8.51.** For \( p \in [1, \infty) \), \( (\ell^p(\mu), \| \cdot \|_p) \) is a normed vector space. \( \blacksquare \)

Proof. The only difficulty is the proof of the triangle inequality which is the content of Minkowski’s Inequality proved in Theorem 8.58 below.
Proposition 8.52. 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
\[
F(s) = \int_0^s f(s')ds' \quad \text{and} \quad G(t) = \int_0^t g(t')dt'.
\]
Then for all \( s, t \geq 0 \),
\[
st \leq F(s) + G(t)
\]
and equality holds iff \( t = f(s) \).

Proof. Let
\[
A_s := \{ (\sigma, \tau) : 0 \leq \tau \leq f(\sigma) \text{ for } 0 \leq \sigma \leq s \} \quad \text{and} \quad B_t := \{ (\sigma, \tau) : 0 \leq \sigma \leq g(\tau) \text{ for } 0 \leq \tau \leq t \}
\]
then as one sees from Figure 8.2, \([0, s] \times [0, t] \subset A_s \cup B_t \). (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
\[
st = m([0, s] \times [0, t]) \leq m(A_s) + m(B_t) = F(s) + G(t).
\]
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
\[
h(s) = \int_0^s (t - f(\sigma))d\sigma = \int_0^{g(t)} (t - f(\sigma))d\sigma + \int_{g(t)}^s (t - f(\sigma))d\sigma
\]
\[
\leq \int_0^{g(t)} (t - f(\sigma))d\sigma = h(g(t)).
\]
Combining this with \( h(0) = 0 \) we see that \( h(s) \) takes its maximum at some point \( s \in (0, g(t)) \) 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:
\[
\int_0^{g(t)} (t - f(\sigma))d\sigma = \int_0^t (t - f(g(\tau)))g'(\tau)d\tau = \int_0^t (t - \tau)g'(\tau)d\tau
\]
\[
= \int_0^t g(\tau)d\tau = G(t).
\]

Fig. 8.2. A picture proof of Proposition 8.52

Definition 8.53. 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:
\[
\frac{1}{p} + \frac{1}{q} = 1, \quad 1 + \frac{q}{p} = q, \quad p - \frac{p}{q} = 1 \text{ and } q(p-1) = p.
\] (8.14)

Lemma 8.54. Let \( p \in (1, \infty) \) and \( q := \frac{p}{p-1} \in (1, \infty) \) be the conjugate exponent. Then
\[
st \leq \frac{sp}{p} + \frac{tq}{q}
\] for all \( s, t \geq 0 \) (8.15)
with equality if and only if \( t^q = s^p \). (See Example ?? below for a generalization of the inequality in Eq. (8.15).)
Proof. Let $F(s) = \frac{s^p}{p}$ for $p > 1$. Then $f(s) = s^{p-1} = t$ and $g(t) = t^{\frac{1}{p}} = t^{q-1}$, wherein we have used $q - 1 = p/(p - 1) - 1 = 1/(p - 1)$. Therefore $G(t) = t^q/q$ and hence by Proposition 8.52

\[ \frac{st}{p} \leq \frac{sp}{p} + \frac{tq}{q} \]

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 8.52, here is a direct calculus proof. Fix $t > 0$ and let

\[ h(s) := st - \frac{sp}{p}. \]

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

\[ h\left(t^{\frac{1}{p-1}}\right) = t^{\frac{1}{p-1}}t - \frac{t^\frac{2}{p-1}}{p} = t^{\frac{1}{p-1}} - \frac{t^{\frac{2}{p-1}}}{p} = t^q \left(1 - \frac{1}{p}\right) = \frac{t^q}{q}, \]

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

\[ st - \frac{sp}{p} \leq \frac{tq}{q} \text{ with equality iff } t = s^{p-1}. \]

We also have the following closely related calculus result.

**Lemma 8.55.** Let $a, b > 0$ and $p, q \in [1, \infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$ and define

\[ f(\lambda) := \frac{a}{p} \lambda^p + \frac{b}{q} \lambda^{-q} \text{ for } \lambda > 0. \]

Then

\[ \min_{\lambda > 0} f(\lambda) = a^\frac{1}{p} b^\frac{1}{q}, \]

which in particular implies $\{\text{let } \lambda = 1 \text{ and replace } a \to a^p \text{ and } b \to b^p\}$ that

\[ ab \leq \frac{a^p}{p} + \frac{b^q}{q} \text{ for all } a, b \geq 0. \]

**Proof.** Since $f(\lambda) \to \infty$ if $\lambda \downarrow 0$ or $\lambda \uparrow \infty$, it follows that $f$ has a minimum in $(0, \infty)$. We can find this minimum using the first derivative test. So

\[ 0 = f'(\lambda) = a\lambda^{p-1} - b\lambda^{-q-1} \quad \Rightarrow \quad \frac{b}{a} = \frac{\lambda^{-q} - 1}{\lambda^{p-1}} = \lambda^{p-q}. \]

Evaluating $f$ at this $\lambda$ gives the desired minimum;

\[ f(\lambda) = \lambda^{-q} \left[ a \lambda^p + b \right] = \lambda^{-q} \left[ a \frac{b^p}{p} + \frac{b^q}{q} \right] = \lambda^{-q} \left[ a \frac{b^p}{p} + \frac{b^q}{q} \right] = \lambda^{-q} \left[ a \frac{b^p}{p} + \frac{b^q}{q} \right]. \]

where we have used

\[ \frac{q}{q+p} = \frac{q/(qp)}{1+\frac{1}{q}} = \frac{p}{q+p} = \frac{1}{q}. \]

**Theorem 8.56 (Hölder’s inequality).** Let $p, q \in [1, \infty]$ be conjugate exponents. For all $f, g : X \to \mathbb{F}$,

\[ \|fg\|_1 \leq \|f\|_p \cdot \|g\|_q. \quad (8.16) \]

If $p \in (1, \infty)$ and $f$ and $g$ are not identically zero, then equality holds in Eq. (8.16) iff

\[ \left(\frac{|f|}{\|f\|_p}\right)^p = \left(\frac{|g|}{\|g\|_q}\right)^q. \quad (8.17) \]

**Proof.** The proof of Eq. (8.16) for $p \in \{1, \infty\}$ is easy and will be left to the reader. The cases where $\|f\|_p = 0$ or $\|g\|_q = 0$ or $\infty$ are easily dealt with and are also left to the reader. So we will assume that $p \in (1, \infty) \text{ and } 0 < \|f\|_p, \|g\|_q < \infty$. Letting $s = |f(x)|/\|f\|_p$ and $t = |g(x)|/\|g\|_q$ in Lemma 8.54 implies

\[ \frac{|f(x)|}{\|f\|_p |g(x)|} \leq \frac{1}{p} |f(x)|^p + \frac{1}{q} |g(x)|^q \]

with equality iff $\frac{|f(x)|^p}{\|f\|_p^p |g(x)|^q} = \frac{|g(x)|^q}{\|g\|_q^q}$.

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

\[ \|fg\|_1 \leq \frac{1}{p} + \frac{1}{q} = 1 \]

with equality iff Eq. (8.18) holds for all $x \in X$, i.e. iff Eq. (8.17) holds.
Definition 8.57. For a complex number \( \lambda \in \mathbb{C} \), let

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

For \( \lambda, \mu \in \mathbb{C} \) we will write \( \text{sgn}(\lambda) \equiv \text{sgn}(\mu) \) if \( \text{sgn}(\lambda) = \text{sgn}(\mu) \) or \( \lambda \mu = 0 \).

Theorem 8.58 (Minkowski’s Inequality). If \( 1 \leq p \leq \infty \) and \( f, g \in l^p(\mu) \) then

\[
\|f + g\|_p \leq \|f\|_p + \|g\|_p.
\]

Moreover, assuming \( f \) and \( g \) are not identically zero, equality holds in Eq. \((8.19)\) iff

\[
\text{sgn}(f) \equiv \text{sgn}(g) \quad \text{when } p = 1 \quad \text{and}
\]

\[
f = cg \text{ for some } c > 0 \text{ when } p \in (1, \infty).
\]

Proof. For \( p = 1 \),

\[
\|f + g\|_1 = \sum_x |f + g| \mu = \sum_x (|f| \mu + |g| \mu) = \sum_x |f| \mu + \sum_x |g| \mu
\]

with equality iff

\[
|f| + |g| = |f + g| \iff \text{sgn}(f) \equiv \text{sgn}(g).
\]

For \( p = \infty \),

\[
\|f + g\|_\infty = \sup_x |f + g| \leq \sup_x (|f| + |g|)
\]

\[
\leq \sup_x |f| + \sup_x |g| = \|f\|_\infty + \|g\|_\infty.
\]

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

\[
\|f + g\|^p_p \leq 2^p (\|f\|^p + \|g\|^p_p) < \infty.
\]

Eq. \((8.19)\) is easily verified if \( \|f + g\|_p = 0 \), so we may assume \( \|f + g\|_p > 0 \). Multiplying the inequality,

\[
|f + g|^p = |f + g| |f + g|^{p-1} \leq |f| |f + g|^{p-1} + |g| |f + g|^{p-1}
\]

by \( \mu \), then summing on \( x \) and applying Hölder’s inequality on each term gives

\[
\sum_x |f + g|^p \mu \leq \sum_x |f| |f + g|^{p-1} \mu + \sum_x |g| |f + g|^{p-1} \mu
\]

\[
\leq (\|f\|_p + \|g\|_p) \left\| |f + g|^{p-1}\right\|_q.
\]

(8.21)

Since \( q(p-1) = p \), as in Eq. \((8.14)\),

\[
\|f + g|^{p-1}\|_q = \sum_x (|f + g|^{p-1})^q \mu = \sum_x |f + g|^p \mu = \|f + g\|_p^p.
\]

(8.22)

Combining Eqs. \((8.21)\) and \((8.22)\) shows

\[
\|f + g\|^p_p \leq (\|f\|^p + \|g\|^p_p) \|f + g\|_p^{p/q}
\]

(8.23)

and solving this equation for \( \|f + g\|_p \) (making use of Eq. \((8.14)\)) implies Eq. \((8.19)\). Now suppose that \( f \) and \( g \) are not identically zero and \( p \in (1, \infty) \). Equality holds in Eq. \((8.19)\) iff equality holds in Eq. \((8.23)\) iff equality holds in Eq. \((8.21)\) and Eq. \((8.20)\). The latter happens iff

\[
\text{sgn}(f) \equiv \text{sgn}(g) \quad \text{and}
\]

\[
\left( \frac{|f|}{\|f\|_p} \right)^p = \frac{|f + g|^p}{\|f + g\|^p_p} = \left( \frac{|g|}{\|g\|_p} \right)^p.
\]

(8.24)

wherein we have used

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

Finally Eq. \((8.24)\) is equivalent to \(|f| = c|g|\) with \( c = (\|f\|^p/\|g\|^p) > 0 \) and this equality along with \( \text{sgn}(f) \equiv \text{sgn}(g) \) implies \( f = cg \). \( \blacksquare \)

8.9 Exercises

Exercise 8.7. 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

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

(8.25)

Exercise 8.8. 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 \((8.25)\). (Hint: use the Cauchy criteria for convergence.)
8.9.1 Limit Problems

Exercise 8.9. Show \( \liminf_{n \to \infty} (-a_n) = - \limsup_{n \to \infty} a_n \).

Exercise 8.10. Suppose that \( \limsup_{n \to \infty} a_n = M \in \mathbb{R} \), 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 8.11. Show that

\[
\limsup_{n \to \infty} (a_n + b_n) \leq \limsup_{n \to \infty} a_n + \limsup_{n \to \infty} b_n \tag{8.26}
\]

provided that the right side of Eq. (8.26) 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 8.12. Suppose that \( a_n \geq 0 \) and \( b_n \geq 0 \) for all \( n \in \mathbb{N} \). Show

\[
\limsup_{n \to \infty} (a_n b_n) \leq \limsup_{n \to \infty} a_n \cdot \limsup_{n \to \infty} b_n, \tag{8.27}
\]

provided the right hand side of (8.27) is not of the form \( 0 \cdot \infty \) or \( \infty \cdot 0 \).

Exercise 8.13. Prove Lemma 8.41


8.9.2 Monotone and Dominated Convergence Theorem Problems

Exercise 8.15. Let \( M < \infty \), show there are polynomials \( p_n(t) \) and \( q_n(t) \) for \( n \in \mathbb{N} \) such that

\[
\lim_{n \to \infty} \sup_{0 \leq t \leq M} |\sqrt{t} - q_n(t)| = 0 \tag{8.28}
\]

and

\[
\lim_{n \to \infty} \sup_{|t| \leq M} |t| - p_n(t) = 0 \tag{8.29}
\]

using the following outline.

1. Let \( f(x) = \sqrt{1 - x} \) for \( |x| \leq 1 \) and use Taylor’s theorem with integral remainder (see Eq. (12.19) of Appendix ??), or analytic function theory if you know it, to show there are constants \( c_n > 0 \) for \( n \in \mathbb{N} \) such that

\[
\sqrt{1 - x} = 1 - \sum_{n=1}^{\infty} c_n x^n \text{ for all } |x| < 1. \tag{8.30}
\]

2. Let \( \tilde{q}_n(x) := 1 - \sum_{n=1}^{m} c_n x^n \). Use (8.30) to show \( \sum_{n=1}^{\infty} c_n = 1 \) and conclude from this that

\[
\lim_{m \to \infty} \sup_{|x| \leq 1} |\sqrt{1 - x} - \tilde{q}_m(x)| = 0. \tag{8.31}
\]

3. Conclude that \( q_n(t) := \sqrt{M} \tilde{q}_n(1 - t/M) \) and \( p_n(t) := q_n(t^2) \) for \( n \in \mathbb{N} \) are polynomials verifying Eqs. (8.28) and (8.29) respectively.

Notation 8.59 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 \( \delta \).

Exercise 8.16. 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 \( \lambda \in \mathbb{C} \) 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 8.17. 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 \( u \mapsto f(u, y) \) is continuous on \( U \).
2. There is a summable function \( g : Y \to [0, \infty) \) such that

\[
|f(u, y)| \leq g(y) \text{ for all } y \in Y \text{ and } u \in U.
\]

Show that

\[
F(u) := \sum_{y \in Y} f(u, y) \tag{8.32}
\]

is a continuous function for \( u \in U \).

Exercise 8.18. 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 \( u \mapsto f(u, y) \) is differentiable on \( J \),
2. There is a summable function \( g : Y \to [0, \infty) \) such that

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

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

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

\footnote{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 \( U \).}
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 $\sum_{y \in Y} |f(u, y)| < \infty$.

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.)

### 8.9.3 $\ell^p$ Exercises

**Exercise 8.19.** Generalize Proposition 8.52 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 $g = f^{-1}$, $b = f(0) \geq 0$,

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

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

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

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

$$st \leq e^s + (t \vee 1) \ln(t \vee 1) - (t \vee 1) \leq e^s + t \ln t - t,$$

where $s \vee t := \min(s, t)$. **Hint:** Refer to Figures 8.3 and 8.4.

**Exercise 8.20.** Using differential calculus, prove the following inequalities

1. For $y > 0$, let $g(x) := xy - e^x$ for $x \in \mathbb{R}$. Use calculus to compute the maximum of $g(x)$ and use this prove Young’s inequality:

$$xy \leq e^x + y \ln y - y \quad \text{for } x \in \mathbb{R} \text{ and } y > 0.$$

2. For $p > 1$ and $y \geq 0$, let $g(x) := xy - x^p/p$ for $x \geq 0$. Again use calculus to compute the maximum of $g(x)$ and show that your result gives the following inequality:

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q} \quad \text{for all } x, y \geq 0.$$

where $q = \frac{p}{p-1}$, i.e. $\frac{1}{q} = 1 - \frac{1}{p}$.

3. Suppose now that $u : [0, \infty) \to [0, \infty)$ is a $C^1$-function such that: $u(0) = 0$, $\lim_{x \to \infty} \frac{u'(x)}{x} = \infty$, and $u'(x) > 0$ for all $x > 0$. Show

$$xy \leq u(x) + v(y) \quad \text{for all } x, y \geq 0,$$

where $v(y) = y(u'(y)^{-1}(y) - u((u')^{-1}(y))$. **Hint:** consider the function, $g(x) := xy - u(x)$.
Topological Considerations

9.1 Closed and Open Sets

Let \((X, d)\) be a metric space.

**Definition 9.1.** 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 \(C_x(\delta) := \{y \in X : d(x, y) \leq \delta\}\).

![Fig. 9.1. Balls in \(\mathbb{R}^2\) corresponding to the 1–norm, 2–norm, 5–norm, and \(\frac{1}{2}\)–"norm."](image1)

![Fig. 9.2. The ball in \(C([0, 1], \mathbb{R})\) of radius 1/4 centered at \(f(x) = \sin(x^2)\) are all the continuous functions whose graphs lie between the green envelope.](image2)
Example 9.6 (Open Balls). Open balls in metric spaces are open sets. Indeed we have seen that open subsets of $X$ are closed. Theorem 9.7. Closed balls are closed. Indeed, we have seen that $A$ is a closed set. Notice that $A$ is closed.

**Lemma 9.3.** If $f : X \to \mathbb{R}$ is a continuous function and $k \in \mathbb{R}$, then the following sets are closed,

- $A := \{x \in X : f(x) \leq k\}$,
- $B := \{x \in X : f(x) = k\}$, and
- $C := \{x \in X : f(x) \geq k\}$.

**Proof.** The proof that $A$, $B$, and $C$ are closed all go the same way so let me just check that $A$ is closed. To this end, suppose that $\{x_n\}_{n=1}^{\infty}$ is a sequence in $A$ such that $x := \lim_{n \to \infty} x_n$ exists in $X$. Since $x_n \in A$, $f(x_n) \leq k$ and therefore, $k \geq \lim_{n \to \infty} f(x_n) = f(x)$ wherein the last equality we have used the definition of $f$ being continuous. By definition of $A$ it then follows that $x \in A$ and so we have checked that $A$ is closed.

**Example 9.4 (Closed Balls).** Closed balls are closed. Indeed, we have seen $f(y) := d(x, y)$ is continuous and therefore

$$C_x(\delta) := \{y \in X : d(x, y) \leq \delta\} = \{y \in X : f(y) \leq \delta\}$$

is a closed set. Notice that $\{x\} = C_x(0)$ is a closed set for all $x \in X$.

**Example 9.5.** The following subsets of $\mathbb{C}$ are closed;

1. $\{z \in \mathbb{C} : a \leq \text{Im} z \leq b\}$ for all $a \leq b$ in $\mathbb{R}$.
2. $\{z \in \mathbb{C} : a \leq \text{Re} z \leq b\}$ for all $a \leq b$ in $\mathbb{R}$.
3. $\{z \in \mathbb{C} : \text{Im} z = 0$ and $a \leq \text{Re} z \leq b\}$ for all $a \leq b$ in $\mathbb{R}$.

**Example 9.6 (Open Balls).** Open balls in metric spaces are open sets. Indeed let $f(y) := d(x, y)$, then

$$B_x(\delta)^c := X \setminus B_x(\delta) = \{y \in X : d(x, y) \geq \delta\} = \{y \in X : f(y) \geq \delta\}$$

which is closed since $f$ is continuous. Thus $B_x(\delta)$ is open.

**Theorem 9.7.** The closed subsets of $(X, d)$ have the following properties;

1. $X$ and $\emptyset$ are closed.
2. If $\{C_x\}_{x \in I}$ is a collection of closed subsets of $X$, then $\cap_{x \in I} C_x$ is closed in $X$.
3. If $A$ and $B$ are closed sets then $A \cup B$ is closed.

**Proof.** 3. Let $\{z_n\}_{n=1}^{\infty} \subset A \cup B$ such that $\lim_{n \to \infty} z_n = z$ exists. Then $z_n \in A$ i.o. or $z_n \in B$ i.o. For sake of definiteness say $z_n \in A$ i.o. in which case we may choose a subsequence, $w_k := z_n_k \in A$ for all $k$. Since $\lim_{k \to \infty} w_k = z$ and $A$ is closed it follows that $z \in A$ and hence $z \in A \cup B$. Thus we have shown $A \cup B$ is closed.

**Exercise 9.1.** Prove item 2. of Theorem 9.7. If $\{C_x\}_{x \in I}$ is a collection of closed subsets of $X$, then $\cap_{x \in I} C_x$ is closed in $X$.

**Exercise 9.2.** Give an example of a collection of closed subsets, $\{A_n\}_n$, of $\mathbb{R}$ such that $\cup_{n=1}^{\infty} A_n$ is not closed.

**Corollary 9.9.** Let $(X, d)$ be a metric space. Then the collection of open subsets, $\tau_d$, of $X$ satisfy;

1. $X$ and $\emptyset$ are in $\tau_d$.
2. $\tau_d$ is closed under taking arbitrary unions, i.e. if $\{U_x\}_{x \in I}$ is a collection of open sets then $\cup_{x \in I} U_x$ is open.
3. $\tau_d$ is closed under taking finite intersections, i.e. if $U$ and $V$ are open sets then $U \cap V$ is open as well.

**Exercise 9.3.** Prove Corollary 9.8.

**Exercise 9.4.** Let $U$ be a subset of a metric space $(X, d)$. Show the following are equivalent;

1. $U$ is open.
2. for all $z \in U$ there exists $\rho > 0$ such that $B_z(\rho) \subset U$.
3. $U$ can be written as a union of open balls.

**Exercise 9.5 (Redundant now).** 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.

**Exercise 9.6.** Let $(X, d)$ be a metric space and $\{x_1, \ldots, x_n\}$ be a finite subset of $X$. Show $X \setminus \{x_1, \ldots, x_n\}$ is an open subset.

**Exercise 9.7.** 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 $X$. 

Lemma 9.9. For any nonempty subset $A \subset X$, let $d_A(x) := \inf\{d(x,a) | a \in A\}$, then
\[ |d_A(x) - d_A(y)| \leq d(x,y) \quad \forall x, y \in X \tag{9.1} \]
and in particular if $x_n \to x$ in $X$ then $d_A(x_n) \to d_A(x)$ as $n \to \infty$. Moreover the set $F_{\varepsilon} := \{ x \in X | d_A(x) \geq \varepsilon \}$ is closed in $X$.

Proof. Let $a \in A$ and $x, y \in X$, then
\[ d_A(x) \leq d(x, a) \leq d(x, y) + d(y, a). \]
Take the infimum over $a$ in the above equation shows that
\[ d_A(x) \leq d(x, y) + d_A(y) \quad \forall x, y \in X. \]
Therefore, $d_A(x) - d_A(y) \leq d(x, y)$ and by interchanging $x$ and $y$ we also have that $d_A(y) - d_A(x) \leq d(x, y)$ which implies Eq. (9.1). If $x_n \to x \in X$, then by Eq. (9.1),
\[ |d_A(x) - d_A(x_n)| \leq d(x, x_n) \to 0 \quad \text{as } n \to \infty \]
so that $\lim_{n \to \infty} d_A(x_n) = d_A(x)$. Now suppose that $\{x_n\}_{n=1}^{\infty} \subset F_{\varepsilon}$ and $x_n \to x$ in $X$, then
\[ d_A(x) = \lim_{n \to \infty} d_A(x_n) \geq \varepsilon \]
since $d_A(x_n) \geq \varepsilon$ for all $n$. This shows that $x \in F_{\varepsilon}$ and hence $F_{\varepsilon}$ is closed.

Definition 9.10. A subset $A \subset X$ is a neighborhood of $x$ if there exists an open set $V \subset_0 X$ 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 $x \in A$.

Example 9.11. 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 $d(y_n, x) \leq \delta$ for all $n$ and using Corollary 6.46 it follows $d(y, x) \leq \delta$, i.e. $y \in C_x(\delta)$. A similar proof shows $B_x(\delta)^c$ is closed, see Exercise 9.5.

Lemma 9.12 (Approximating open sets from the inside by closed sets). Let $U \subset X$ be an open set and $F_{\varepsilon} := \{ x \in X | d_{U^c}(x) \geq \varepsilon \} \subset X$ be as in Lemma 9.9. Then $F_{\varepsilon} \uparrow U$ as $\varepsilon \downarrow 0$.

Proof. It is clear that $d_{U^c}(x) = 0$ for $x \in U^c$ so that $F_{\varepsilon} \subset U$ for each $\varepsilon > 0$ and hence $\cup_{\varepsilon > 0} F_{\varepsilon} \subset U$. Now suppose that $x \in U \subset_0 X$. By Exercise 9.5 there exists an $\varepsilon > 0$ such that $B_x(\varepsilon) \subset U$, i.e. $d(x, y) \geq \varepsilon$ for all $y \in U^c$. Hence $x \in F_{\varepsilon}$ and we have shown that $U \subset \cup_{\varepsilon > 0} F_{\varepsilon}$. Finally it is clear that $F_{\varepsilon} \subset F_{\varepsilon'}$ whenever $\varepsilon' \leq \varepsilon$.

Definition 9.13. Given a set $A$ contained in a metric space $X$, let $\bar{A} \subset X$ be the closure of $A$ defined by
\[ \bar{A} := \{ x \in X : \exists \{x_n\} \subset A \implies x = \lim_{n \to \infty} x_n \}. \]
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 $A$. A metric space is said to be separable if it contains a countable dense subset, $D$.

Lemma 9.14. For any $A \subset X$, then
\begin{enumerate}
\item $A = \bar{A}$ if $A$ is closed.
\item $\bar{A} = \{ x : d_A(x) = 0 \}$ and $\bar{A}$ is closed.
\item $\bar{A} = \{ x \in X : A \cap B_x(\rho) \neq \emptyset \forall \rho > 0 \}$.
\item $d_A(x) > 0$ for all $x \in A^c$ if $A$ is closed.
\end{enumerate}

Proof. 1. We always have $A \subset \bar{A}$. If $A$ is closed we can not leave $A$ by taking limits and hence $\bar{A} \subset A$, i.e. $A = \bar{A}$ if $A$ is closed.

2. Let $F := \{ x : d_A(x) = 0 \}$ which is a closed set since $d_A$ is continuous. If $x \in F$ (i.e. $d_A(x) = 0$), there exists $x_n \in A$ such that $d(x, x_n) \leq 1/n$ for all $n \in \mathbb{N}$. Thus $\lim_{n \to \infty} x_n = x$ and $x \in \bar{A}$. This shows $F \subset \bar{A}$. Conversely if $x \in \bar{A}$, there exists $\{x_n\} \subset A$ such that $\lim_{n \to \infty} x_n = x$ and so
\[ d_A(x) = d_A \left( \lim_{n \to \infty} x_n \right) = \lim_{n \to \infty} d_A(x_n) = \lim_{n \to \infty} 0 = 0 \]
which shows $x \in F$.

3. $\iff$ 3. Since $A \cap B_x(\rho) \neq \emptyset$ happens iff $d_A(x) < \rho$ we see that $A \cap B_x(\rho) \neq \emptyset \forall \rho > 0$ iff $d_A(x) < \rho$ for all $\rho > 0$, i.e. iff $d_A(x) = 0$.

4. If $A$ is closed then $A = \bar{A} = \{ x \in X : d_A(x) = 0 \}$ and therefore $A^c = \{ x \in X : d_A(x) > 0 \}$.

Exercise 9.8. If $A$ is a non-empty subset of $X$, then $d_A = d_{\bar{A}}$.

Exercise 9.9. Given $A \subset X$, show $\bar{A}$ is a closed set and in fact
\[ \bar{A} = \cap \{ F : A \subset F \subset X \text{ with } F \text{ closed} \}. \tag{9.2} \]
That is to say $\bar{A}$ is the smallest closed set containing $A$.

Definition 9.15. Let $(X,d)$ be a metric space and $A$ be a subset of $X$.

1. The closure of $A$ is the smallest closed set $\bar{A}$ containing $A$, i.e.
\[ \bar{A} := \cap \{ F : A \subset F \subset X \}. \]
(Because of Exercise 9.9 this is consistent with Definition 9.13 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.
\[
A^o = \{ V \in \tau : V \subset A \}.
\]
3. \( A \subset X \) is a **neighborhood of a point** \( x \in X \) if \( x \in A^o \).
4. The **accumulation points** of \( A \) is the set
\[
\text{acc}(A) = \{ x \in X : V \cap (A \setminus \{x\}) \neq \emptyset \text{ for all } V \in \tau_x \}.
\]
5. The **boundary** of \( A \) is the set \( \partial(A) := \bar{A} \setminus A^c \).
6. \( A \) is **dense** in \( X \) if \( A = X \) and \( X \) is said to be **separable** if there exists a countable dense subset of \( X \).

**Lemma 9.16.** Let \((X,d)\) be a metric space and \( A \) be a subset of \( X \), then
\[
A^o = \{ x \in X : B_x(\rho_x) \subset A \text{ for some } \rho_x > 0 \}.
\]

**Proof.** Let \( V := \{ x \in X : B_x(\rho) \subset A \text{ for some } \rho = \rho_x > 0 \} \). If \( B_x(\rho) \subset A \) and \( y \in B_x(\rho) \) and \( \delta = \rho - d(x,y) \), then \( B_y(\delta) \subset B_x(\rho) \subset A \) which shows that \( y \in V \), i.e. \( B_x(\rho) \subset V \). Thus we may write
\[
V = \{ B_x(\rho) : B_x(\rho) \subset A \}.\]

This shows that \( V \) is an open subset of \( A \). Moreover if \( W \) is another open subset of \( A \), then
\[
W = \{ B_x(\rho) : B_x(\rho) \subset W \} \subset \{ B_x(\rho) : B_x(\rho) \subset A \} = V
\]
so that \( V \) is the largest open subset contained in \( A \). This completes the proof. \( \blacksquare \)

**Exercise 9.10.** Let \((X,||\cdot||)\) be a normed space and \( d(x,y) := \|x-y\| \). Show;
1. \( C_x(\rho)^o = B_x(\rho) \),
2. \( B_x(\rho) = C_x(\rho) \),
3. \( \partial(C_x(\rho)) = \partial(B_x(\rho)) = \{ y \in X : \|y - x\| = \rho \} \).

**Example 9.17 (Words of Caution).** Let \((X,d)\) be a metric space. It is always true that \( B_{\varepsilon}(x) \subset C_{\varepsilon}(x) \) since \( C_{\varepsilon}(x) \) is a closed set containing \( B_{\varepsilon}(x) \). However, it is not always true that \( B_{\varepsilon}(x) = C_{\varepsilon}(x) \). For example let \( X = \{1,2\} \) and \( d(1,2) = 1 \), then \( B_1(1) = \{1\} \), \( B_1(\bar{1}) = \{1\} \) while \( C_1(1) = X \). For another counterexample, take
\[
X = \{(x,y) \in \mathbb{R}^2 : x = 0 \text{ or } x = 1\}
\]
with the usually Euclidean metric coming from the plane. Then
\[
B_{(0,0)}(1) = \{(0,y) \in \mathbb{R}^2 : |y| < 1 \}, \quad B_{(0,0)}(\bar{1}) = \{(0,y) \in \mathbb{R}^2 : |y| \leq 1 \}, \quad \text{while} \quad C_{(0,0)}(1) = B_{(0,0)}(\bar{1}) \cup \{(1,0)\}.
\]

**Exercise 9.11.** Suppose that \( A \) and \( B \) are subsets of a metric space, show \( \overline{A \cup B} = \overline{A} \cup \overline{B} \).

**Exercise 9.12.** Given an example showing that \( \bigcup_{n=1}^{\infty} A_n \) need not be equal to \( \bigcup_{n=1}^{\infty} \overline{A_n} \).

**Remark 9.18.** The relationships between the interior and the closure of a set are:
\[
(A^o)^c = \bigcap \{ V^c : V \subset \tau \text{ and } V \subset A \} = \bigcap \{ C : C \text{ is closed } \supset A^c \} = \overline{A^c}
\]
and similarly, \( \overline{(A^c)} = (A^o)^c \).

**Proposition 9.19.** If \( A \) is a subset of a metric space, \((X,d)\), then \( \partial(A) \) may be computed using either;
1. \( \partial(A) = \bar{A} \cap \overline{A^c} \),
2. \( \partial(A) = \{ x \in X : d_A(x) = 0 = d_{A^c}(x) \} \),
3. \( \partial(A) = \{ x \in X : B_x(\rho) \cap A \neq \emptyset \neq B_x(\rho) \cap A^c \text{ for all } \rho > 0 \} \), or
4. \( \partial(A) = \{ x \in X : \exists \{x_n\} \subset A \text{ and } \{y_n\} \subset A^c \Rightarrow \lim_{n \to \infty} x_n = x = \lim_{n \to \infty} y_n \} \).

**Proof.** From the definition of \( \partial(A) \) and the relation, \( \overline{(A^c)} = (A^o)^c \), we find
\[
\partial(A) = \bar{A} \setminus A^o = \bar{A} \cap (A^o)^c = \bar{A} \setminus \overline{A^c}. \tag{9.3}
\]

Item 2. now follow from item 1. and the fact that \( \bar{A} = \{ x \in X : d_A(x) = 0 \} \). I leave it to the reader to check that 2. \( \Rightarrow \) 3. \( \Rightarrow \) 4. \( \Rightarrow \) 1. \( \blacksquare \)

**Exercise 9.13.** If \( D \) is a dense subset of a metric space \((X,d)\) and \( E \subset X \) is a subset such that to every \( x \in D \) there exists \( \{x_n\}_{n=1}^{\infty} \subset E \) with \( x = \lim_{n \to \infty} x_n \), then \( E \) is also a dense subset of \( X \). If points in \( E \) well approximate every point in \( D \) and the points in \( D \) well approximate the points in \( X \), then the points in \( E \) also well approximate all points in \( X \).

**Exercise 9.14.** Suppose \((X,d)\) is a metric space which contains an uncountable subset \( A \subset X \) with the property that there exists \( \varepsilon > 0 \) such that \( d(a,b) \geq \varepsilon \) for all \( a,b \in A \) with \( a \neq b \). Show that \((X,d)\) is not separable.

**Exercise 9.15.** Let \( Y = BC(\mathbb{R},\mathbb{C}) \) be the Banach space of continuous bounded complex valued functions on \( \mathbb{R} \) equipped with the uniform norm, \( ||f||_u := \sup_{x \in \mathbb{R}} |f(x)| \). Further let \( C_0(\mathbb{R},\mathbb{C}) \) denote those \( f \in C(\mathbb{R},\mathbb{C}) \) such that vanish at infinity, i.e. \( \lim_{x \to \pm \infty} f(x) = 0 \). Also let \( C_c(\mathbb{R},\mathbb{C}) \) denote those \( f \in C(\mathbb{R},\mathbb{C}) \) with compact support, i.e. there exists \( N < \infty \) such that \( f(x) = 0 \) if \( |x| \geq N \). Show \( C_0(\mathbb{R},\mathbb{C}) \) is a closed subspace of \( Y \) and that
\[
C_c(\mathbb{R},\mathbb{C}) = C_0(\mathbb{R},\mathbb{C}).
\]


9.2 Continuity Revisited

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

**Definition 9.20.** A function \(f : X \to Y\) is continuous at \(x \in X\) if for all \(\varepsilon > 0\) there is a \(\delta > 0\) such that

\[
d(f(x), f(x')) < \varepsilon \quad \text{provided} \quad \rho(x, x') < \delta. \tag{9.4}
\]

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 9.21 (Local Continuity Lemma).** Suppose that \((X, \rho)\) 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 \(\{x_n\}_{n=1}^{\infty}\) of \(X\) 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 \(\Rightarrow\) 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. 9.4 holds. Therefore

\[
B_\delta(\delta) \subset f^{-1} \left( B_f(x) (\varepsilon) \right) \subset f^{-1}(A)
\]

showing \(f^{-1}(A)\) is a neighborhood of \(x\).

2 \(\Rightarrow\) 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 \(f^{-1}(B_f(x)(\varepsilon))\) is a neighborhood of \(x\) which must contain \(B_\varepsilon(\delta)\) for some \(\delta > 0\). Because \(x_n \to x\), it follows that \(x_n \in B_\varepsilon(\delta) \subset f^{-1}(B_f(x)(\varepsilon))\) 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. \(\Rightarrow\) 1. We will show not 1. \(\Rightarrow\) 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

\[
\rho(x_n, x) < \frac{\varepsilon}{2} \quad \text{yet} \quad d(f(x_n), f(x)) \geq \varepsilon.
\]

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 9.22 (Global Continuity Lemma).** Suppose that \((X, \rho)\) 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 \(f^{-1}(A^c) = [f^{-1}(A)]^c\), it is easily seen that 2. and 3. are equivalent. So because of Lemma 9.21 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 9.5 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 \(V \subset \rho\) 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.

**Exercise 9.16.** Use Example 9.28 and Lemma 9.22 to recover the results of Example 9.11.

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

**Lemma 9.23 (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

\[
f(x) = \frac{d_B(x)}{d_A(x) + d_B(x)} \quad \text{for} \quad x \in X\]

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 9.9 \(d_A\) 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 used 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 9.23 by taking \(A = F\) and \(B = V^c\).
9.3 Metric spaces as topological spaces (Not required Reading!)

Let \((X,d)\) be a metric space and let \(\tau = \tau_d\) denote the collection of open subsets of \(X\). (Recall \(V \subset X\) is open iff \(V^c\) is closed iff for all \(x \in V\) there exists an \(\varepsilon = \varepsilon_x > 0\) such that \(B(x,\varepsilon) \subset V\) iff \(V\) can be written as a (possibly uncountable) union of open balls.) Although we will stick with metric spaces in this chapter, it will be useful to introduce the definitions needed here in the more general context of a general “topological space,” i.e. a space equipped with a collection of “open sets.”

**Definition 9.24 (Topological Space).** Let \(X\) be a set. A topology on \(X\) is a collection of subsets (\(\tau\)) of \(X\) with the following properties:

1. \(\tau\) contains both the empty set (\(\emptyset\)) and \(X\).
2. \(\tau\) is closed under arbitrary unions.
3. \(\tau\) is closed under finite intersections.

The elements \(V \in \tau\) are called open subsets of \(X\). A subset \(F \subset X\) is said to be closed if \(F^c\) is open. I will write \(V \subset_X X\) to indicate that \(V \subset X\) and \(V \in \tau\) and similarly \(F \subset_X X\) will denote \(F \subset X\) and \(F\) is closed. Given \(x \in X\) we say that \(V \subset X\) is an open neighborhood of \(x\) if \(V \in \tau\) and \(x \in V\). Let \(\tau_x = \{V \in \tau : x \in V\}\) denote the collection of open neighborhoods of \(x\).

**Definition 9.25 (Continuity at a point in topological terms).** Let \((X,\tau_X)\) and \((Y,\tau_Y)\) be topological spaces. A function \(f : X \to Y\) is continuous at \(x \in X\) if

\[ f^{-1}(\tau_Y) := \{ f^{-1}(V) : V \in \tau_Y \} \subset \tau_X. \]

We will also say that \(f\) is \(\tau_X/\tau_Y\)-continuous or \((\tau_X,\tau_Y)\)-continuous. Let \(C(X,Y)\) denote the set of continuous functions from \(X\) to \(Y\).

**Exercise 9.17.** Show \(f : X \to Y\) is continuous (Definition ??) iff \(f\) is continuous at all points \(x \in X\).

**Exercise 9.18.** Show \(f : X \to Y\) is continuous iff \(f^{-1}(C)\) is closed in \(X\) for all closed subsets \(C\) of \(Y\).

**Definition 9.27.** 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.)

**Example 9.28.** The function \(d_A\) defined in Lemma 9.9 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 \(x \in X\).

9.4 Exercises

**Exercise 9.19.** 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 9.20.** Suppose that \(f : [0,\infty) \to [0,\infty)\) is a \(C^2\) function such that \(f(0) = 0\), \(f' > 0\) and \(f'' \leq 0\) and \((X,\rho)\) 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 \(d(a+b) \leq d(a) + d(b)\) for all \(a, b \in [0,\infty)\).)
Exercise 9.21. 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
\[
d(x, y) = \sum_{n=1}^{\infty} 2^{-n} \frac{d_n(x(n), y(n))}{1 + d_n(x(n), y(n))}.
\] (9.6)

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 \( n \).

9.5 Sequential Compactness

Suppose that \((X, d)\) and \((Y, \rho)\) are metric spaces.

Definition 9.29. As subset \( K \subset X \) is (sequentially) compact if every sequence \( \{z_n\}_{n=1}^{\infty} \subset K \) has a convergent subsequence, \( \{w_k := z_{n_k}\}_{k=1}^{\infty} \) such that \( \lim_{k \to \infty} w_k \in K \).

Example 9.30. Suppose that \( F \subset X \) is an unbounded set, i.e. for all \( n \in \mathbb{N} \) there exists \( z_n \in F \) such that \( d(z, z_n) \geq n \). The sequence \( \{z_n\}_{n=1}^{\infty} \) and all of its subsequences are unbounded and therefore not Cauchy in \( X \) and hence not convergent in \( X \). This shows that sequentially compact sets must be bounded.

Example 9.31. Suppose that \( F \subset X \) is not closed. Then there exists \( \{z_n\}_{n=1}^{\infty} \subset F \) such that \( z := \lim_{n \to \infty} z_n \notin F \). Moreover, although every subsequence of \( \{z_n\}_{n=1}^{\infty} \) is convergent, they all still converge to \( z \notin F \). This shows that a sequentially compact set must be closed.

Lemma 9.32 (Bolzano–Weierstrass property for \( \mathbb{C}^D \)). Let \( D \in \mathbb{N} \). Every bounded sequence, \( \{z(n)\}_{n=1}^{\infty} \subset \mathbb{C}^D \), has a convergent subsequence.

Proof. By assumption there exists \( M < \infty \) such that \( \|z(n)\| = d(z(n), 0) \leq M \) for all \( n \in \mathbb{N} \). Writing \( z(n) = (z_1(n), \ldots, z_D(n)) \in \mathbb{C}^D \). Since \( |z_i(n)| \leq \|z(n)\| \) it follows that \( \{z_i(n)\}_{n=1}^{\infty} \) is a bounded sequence in \( \mathbb{C} \). Hence by the Bolzano–Weierstrass property for \( \mathbb{C} \) we replace \( z(n) \) by a subsequence \( z(n_k) \) such that \( \lim_{k \to \infty} z(n_k) = z_1 \) exists. We may now replace the original \( z \) by this new subsequence and then find a further subsequence \( z(n_k) \) such that \( \lim_{k \to \infty} z_i(n_k) = z_i \) exists for \( i = 1, 2 \). We may continue this way inductively to find a subsequence such that \( \lim_{k \to \infty} z(n_k) = z_1 \) exists for all \( 1 \leq i \leq D \). It then follows that \( \lim_{k \to \infty} \|z - z(n_k)\| = 0 \) as desires where \( z := (z_1, \ldots, z_D) \).

Theorem 9.33 (Bolzano–Weierstrass / Heine–Borel theorem). As subset \( K \subset \mathbb{C}^D \) is sequentially compact iff it is closed and bounded.

Proof. In light of Examples 9.30 and 9.31 we are left to show that closed and bounded sets are sequentially compact. So let \( K \subset \mathbb{C}^D \) be a closed and bounded set and \( \{z_n\}_{n=1}^{\infty} \) be any sequence in \( K \). According to Lemma 9.32 \( \{z_n\}_{n=1}^{\infty} \) has a convergent subsequence, \( \{w_k := z_{n_k}\}_{k=1}^{\infty} \). Since \( w_k \in K \) for all \( k \) and \( K \) is closed it necessarily follows that \( \lim_{k \to \infty} w_k \in K \) which shows \( K \) is sequentially compact.

Example 9.34 (Warning!). It is not true that a closed and bounded subset of an arbitrary metric space \((X, d)\) is necessarily sequentially compact. For example let \( Z \) denote the vector space of continuous functions on \([0, 1]\) with values in \( \mathbb{R} \) and for \( f \in Z \) let \( \|f\| = \sup_{t \in [0, 1]} |f(t)| \). Then the set \( C := \{f_n\}_{n=0}^{\infty} \) where

\[
f_n(t) = 2^{(n+2)} \begin{cases} 
0 & \text{if } t \in [0, 2^{-(n+1)}] \cup [2^{-n}, 1] \\
n^{-2} & \text{if } 2^{-(n+1)} \leq t \leq 3 \cdot 2^{-(n+2)} \\
2^{-n} - t & \text{if } 3 \cdot 2^{-(n+2)} \leq t \leq 2^{-n}.
\end{cases}
\]

[So \( f_n(t) \) is a shark tooth over the interval \([2^{-(n+1)}, 2^{-n}]\).] Notice that \( \|f_n\| = 1 \) for all \( n \) so that \( C \) is bounded. Moreover \( \|f_n - f_m\| = 1 \) for all \( m \neq n \), therefore there are no convergent subsequence of \( C \). The reader should use this fact to see that \( C \) is closed and bounded but not sequentially compact!

Exercise 9.22 (Extreme value theorem). Let \( K \) be sequentially compact subset of \( X \) and \( f : K \to \mathbb{R} \) be a continuous function. Show \( -\infty < \)
Lemma 9.36. Suppose that \( p(z) = \sum_{k=0}^{n} a_k z^k \) is a polynomial on \( \mathbb{C} \) with \( a_n \neq 0 \) and \( n > 0 \). Then there exists \( z_0 \in \mathbb{C} \) such that \( p(z_0) = 0 \).

Proof. Since
\[
\lim_{|z| \to \infty} \frac{|p(z)|}{|z|^n} = \lim_{|z| \to \infty} \left| \sum_{k=0}^{n-1} \frac{a_k}{z^{n-k}} + a_n \right| = |a_n|,
\]
if \( R > 0 \) is sufficiently large, then
\[
|p(z)| \geq |a_n| |z|^n / 2 \geq |a_n| R^n / 2 \geq |a_0| = |p(0)| \quad \text{for} \quad |z| > R.
\]
Applying Exercise 9.23 with \( K = \{ z \in \mathbb{C} : |z| \leq R \} \) and \( x_0 = 0 \) there exists \( z_0 \in K \subset \mathbb{C} \) such that \( |p(z_0)| \leq |p(z)| \) for all \( z \in \mathbb{C} \). It now follows from Lemma 9.36 below that \( p(z_0) = 0 \).

Lemma 9.36. Suppose that \( p(z) = \sum_{k=0}^{n} a_k z^k \) is a polynomial on \( \mathbb{C} \) with \( a_n \neq 0 \) and \( n > 0 \). If \( |p(z)| \) has a minimum at \( z_0 \in \mathbb{C} \), then \( |p(z_0)| = 0 \).

Proof. For sake of contradiction, let us suppose that \( p(z_0) \neq 0 \) and set
\[
q(z) := p(z_0 + z) = \sum_{k=0}^{n} a_k (z + z_0)^k = \sum_{k=0}^{n} a_k z^k.
\]
Then \( 0 < |q(0)| = |a_0| \leq |q(z)| \) for all \( z \in \mathbb{C} \) and \( a_n = a_0 \neq 0 \). Let \( l \geq 1 \) be the first index such that \( a_l \neq 0 \) so that
\[
q(z) = a_0 + a_l z^l + \cdots + a_n z^n = a_0 \left[ 1 + \frac{a_l}{a_0} z^l + \cdots + \frac{a_n}{a_0} z^n \right].
\]
Evaluating this at \( z = r e^{i \theta} \) with \( \theta \) chosen so that \( \frac{a_l}{a_0} e^{i \theta} = -\frac{|a_l|}{|a_0|} = -\varepsilon \) implies for \( r > 0 \) small that
\[
|a_0| \leq q(r e^{i \theta}) = |a_0| |1 - \varepsilon r^l + c_l + r^{l+1} + \cdots + c_n r^n| \leq |a_0| |1 - \varepsilon r^l + \sum_{i=l+1}^{n} c_i r^i| \leq |a_0| |1 - \varepsilon r^l + Cr^{l+1}| = |a_0| |1 - r^l [\varepsilon - Cr]|
\]
where \( \{c_k\} \) and \( C \) are certain appropriate constants. Choosing \( r \in (0, \varepsilon/C) \) then gives the \( |a_0| < |a_0| \) which is absurd and we have reached the desired contradiction.

Remark 9.37. The fundamental theorem of algebra does not hold for polynomials in \( z \) and \( \bar{z} \) or in \( x \) and \( y \). For example consider the polynomial
\[
p(z, \bar{z}) = 1 + z \bar{z} = 1 + x^2 + y^2.
\]
The point is that Lemma 9.36 does not hold for these more general classes of polynomials.

Exercise 9.24 (Uniform Continuity). Let \( K \) be sequentially compact subset of \( X \) and \( f : K \to \mathbb{C} \) be a continuous function. Show that \( f \) is uniformly continuous, i.e. if \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that \( |f(z) - f(w)| < \varepsilon \) if \( w, z \in K \) with \( d(w, z) < \delta \). Hint: prove the contrapositive.

Exercise 9.25. If \( (X, d) \) is a metric space and \( K \subset X \) is sequentially compact. Show subset, \( C \subset K \), which is closed is sequentially compact as well.

Exercise 9.26. If \( K \subset \mathbb{R} \) is sequentially compact then \( \text{sup}(K) \in K \), i.e. \( \text{sup}(K) = \text{max}(K) \).

Exercise 9.27. Let \( (X, d) \) and \( (Y, \rho) \) be metric spaces, \( K \subset X \) be a sequentially compact set, and \( f : K \to Y \) be a continuous function. Show \( f(K) \) is sequentially compact in \( Y \). In particular, for \( C \subset K \) closed, we have \( f(C) \) is closed and in fact sequentially compact in \( Y \).

Exercise 9.28. Let \( f : [a, b] \to [c, d] \) be a strictly increasing continuous function such that \( f(a) = c \) and \( f(b) = d \) and \( g := f^{-1} : [c, d] \to [a, b] \) as in Exercise 6.20. Give one or better yet two alternative proofs that \( g \) is continuous based on sequentially compactness arguments.

Definition 9.38. Let \( Z \) be a vector space. We say that two norms, \( |\cdot| \) and \( \|\cdot\| \), on \( Z \) are equivalent if there exists constants \( \alpha, \beta \in (0, \infty) \) such that
\[
\|f\| \leq \alpha |f| \quad \text{and} \quad |f| \leq \beta \|f\| \quad \text{for all} \quad f \in Z.
\]

Theorem 9.39. Let \( Z \) be a finite dimensional vector space. Then any two norms \( |\cdot| \) and \( \|\| \) on \( Z \) are equivalent. (This is typically not true for norms on infinite dimensional spaces.)
Proof. Let \( \{ f_i \}_{i=1}^{n} \) be a basis for \( Z \) and define a new norm on \( Z \) 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
\[
\left| \sum_{i=1}^{n} a_i f_i \right| \leq \sum_{i=1}^{n} |a_i| \left| f_i \right| \leq \sqrt{\sum_{i=1}^{n} |a_i|^2} \left\| \sum_{i=1}^{n} a_i f_i \right\|_2 \leq M \left\| \sum_{i=1}^{n} a_i f_i \right\|_2
\]
where \( M = \sqrt{\sum_{i=1}^{n} |f_i|^2} \). Thus we have \( |f| \leq M \|f\|_2 \) for all \( f \in Z \) and this inequality shows that \(|\cdot|\) is continuous relative to \( \|\cdot\|_2 \). Since the normed space \((Z, \|\cdot\|_2)\) is homeomorphic and isomorphic to \( \mathbb{F}^n \) with the standard euclidean norm, the closed bounded set, \( S := \{ f \in Z : \|f\|_2 = 1 \} \subset Z \), is a sequentially compact subset of \( Z \) relative to \( \|\cdot\|_2 \). Therefore by Exercise 9.22 there exists \( f_0 \in S \) such that
\[
m = \inf \{ |f| : f \in S \} = |f_0| > 0.
\]
Hence given \( 0 \neq f \in Z \), then \( \frac{f}{\|f\|_2} \in S \) so that
\[
m \leq \left| \frac{f}{\|f\|_2} \right| = M \left| \frac{1}{\|f\|_2} \right|
\]
or equivalently
\[
\|f\|_2 \leq \frac{1}{m} |f|.
\]
This shows that \(|\cdot|\) and \( \|\cdot\|_2 \) are equivalent norms. Similarly one shows that \( \|\cdot\|_2 \) and \( \|\cdot\| \) are equivalent and hence so are \(|\cdot|\) and \( \|\cdot\| \).

Corollary 9.40. If \((Z, \|\cdot\|)\) is a finite dimensional normed space, then \( A \subset Z \) is sequentially compact iff \( A \) is closed and bounded relative to the given norm, \( \|\cdot\| \).

Corollary 9.41. Every finite dimensional normed vector space \((Z, \|\cdot\|)\) is complete. In particular any finite dimensional subspace of a normed vector space is automatically closed.

Proof. If \( \{ f_n \}_{n=1}^{\infty} \subset Z \) 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 9.40. It is now routine to show \( \lim_{n \to \infty} f_n = f := \lim_{k \to \infty} g_k \).

9.6 Open Cover Compactness

Definition 9.42. Let \( A \) be a subset of a metric space, \((X, d)\). An open cover of \( A \) is a collection of, \( U \), of open subsets of \( X \) such that \( A \subset \bigcup_{U \in \mathcal{U}} U \).

Definition 9.43. The subset \( A \subset X \) is open cover compact\(^1\) if every open cover (Definition 9.42) of \( A \) has finite sub-cover, i.e. \( U \) is an open cover of \( A \) there exists finite subcollection, \( U_0 \subset U \) such that \( U_0 \) is still a cover of \( A \). (We will write \( A \subset X \) to denote that \( A \subset X \) and \( A \) is compact.) A subset \( A \subset X \) is precompact if \( A \) is compact.

Remark 9.44. If \((X, d)\) is a metric space, we will see below in Theorem 9.53 that \( X \) is sequentially compact if it is open cover compact. We will not use this fact in this section however. In the future though we will just refer to compact sets with out the any extra adjectives.

Example 9.45. Suppose that \( A \) is an unbounded subset of \( X \). Pick \( a_1 \in A \) and then choose \( \{ a_n \}_{n=2}^{\infty} \) inductively so that \( d(a_1, a_n) \geq 1 \) for all \( n \). This sequence then has the property that \( d(a_k, a_l) \geq 1 \) for all \( k \neq l \) and from this it follows that \( F := \{ a_1, a_2, \ldots \} \) is a closed set. We then define an open cover of \( A \) by taking,
\[
U = \{ F^c, B_{a_1}(1/3), B_a(1/3), B_{a_3}(1/3), \ldots \}.
\]
This cover has no finite subcover. Therefore \( A \) can not be open cover compact.

Alternatively. Given \( x \in X \), the collection \( U := \{ B_x(n) : n \in \mathbb{N} \} \) is an open cover of \( X \). So if \( K \) is an open cover compact subset of \( X \), there must exist \( n_1 < n_2 < \cdots < n_l \) so that
\[
K \subset B_x(n_1) \cup B_x(n_2) \cup \cdots \cup B_x(n_l).
\]
This shows \( K \) is bounded.

Lemma 9.46. Suppose that \( K \subset X \) is an open cover compact set, then \( K \) is closed.

Proof. We will shows that \( K^c \) is open. To this end suppose \( x \in K^c \). Then let \( \xi_k := \frac{1}{2} d(x, k) > 0 \) for all \( k \in K \). It then follows that \( B_{x}(\xi_k) \cap B_k(\xi_k) = \emptyset \) for all \( k \in K \). As \( \{ B_k(\xi_k) \}_{k \in K} \) is an open cover \( K \), there exists \( A \subset F \) such that \( K \subset \bigcup_{k \in A} B_k(\xi_k) \). If we now let \( \delta := \min_{k \in A} \xi_k > 0 \), then
\[
B_x(\delta) \cap B_k(\xi_k) \subset B_x(\xi_k) \cap B_k(\xi_k) = \emptyset \text{ for all } k \in A
\]
and therefore
\[
B_x(\delta) \cap K \subset B_x(\delta) \cap \bigcup_{k \in K} B_k(\xi_k) = \emptyset.
\]
\(^1\) We will usually simply say \( A \) is compact in this case.
Theorem 9.47. The open cover compact subsets of \( \mathbb{R}^n \) are the closed and bounded sets.

Proof. Let us first suppose that \( K = [-M,M]^n \) for some positive integer \( M \) and for sake of contradiction let us suppose that \( \mathcal{U} \) is an open cover of \( K \) with no finite subcover. For any \( k \in \mathbb{N} \) let \( A_k := [-M,M] \cap \{ t2^{-k} : t \in \mathbb{Z} \} \) and for \( x \in A_k \), let

\[
C^k_x := [x_1, x_1 + 2^{-k}] \times \cdots \times [x_n, x_n + 2^{-k}].
\]

We now let \( C^k = \{ C^k_x : x \in A_n \} \). The cubes have the following properties;

1. \( K = \cup_{C \in C} C \) for all \( k \) and if \( C \in C_k \), then \( C = \cup_{F \in C_{k+1}} F \).
2. \( \text{diam}(C) = \sqrt{n} \cdot 2^{-k} \) for all \( C \in C_k \).

By item 1 there must be a \( C_1 \in C^1 \) such that \( \mathcal{U} \) has no finite subcover of \( C_1 \). Similarly there exists \( C_2 \in C^2 \) such that \( C_2 \subset C_1 \) such that \( \mathcal{U} \) has no finite subcover of \( C_2 \). Continuing this way inductively we may construct \( C_k \in C^k \) such that \( C_1 \supset C_2 \supset C_3 \supset \ldots \) and \( \mathcal{U} \) has no finite subcover of \( C_k \) for any \( k \). Choose a point \( z_k \in C_k \) for all \( k \in \mathbb{N} \). Because \( \text{diam}(C_k) = \sqrt{n} \cdot 2^{-k} \to 0 \) one learns that \( \{ z_k \}_{k=1}^\infty \) is a Cauchy sequence and hence convergent. Let \( z = \lim_{k \to \infty} z_k \) which is in \( C_k \) for all \( k \) as each of these cubes are closed sets. Since \( \mathcal{U} \) is an open cover of \( K \), there exists \( V \in \mathcal{U} \) such that \( z \in V \). As \( V \) is open, there exists \( \varepsilon > 0 \) such that \( B_z(\varepsilon) \subset V \). On the other hand because diam \( (C_k) \to 0 \) as \( k \to \infty \), it follows that \( C_k \subset B_z(\varepsilon) \subset V \) for all \( k \) sufficiently large which violates the condition that \( \mathcal{U} \) has no finite subcover of \( C_k \) for any \( k \).

Now suppose that \( K \) is a general closed and bounded subset of \( \mathbb{R}^n \) and \( \mathcal{U} \) is an open cover of \( K \). Since \( K \) is bounded, \( K \subset [-M,M]^n \) for some \( M \in \mathbb{N} \). Since \( \mathcal{U} := \mathcal{U} \cup \{ K^c \} \) is an open cover of \( [-M,M]^n \) it has a finite subcover, say \( \{ U_1, \ldots, U_l \} \cup \{ K^c \} \). As \( K^c \cap K = \emptyset \) we must have that \( \{ U_1, \ldots, U_l \} \cap K \) covers \( K \) i.e. \( \{ U_1, \ldots, U_l \} \subset \mathcal{U} \) is a finite subcover of \( K \).

Proposition 9.48. Suppose that \( K \subset X \) is an open cover compact set and \( F \subset K \) is a closed subset. Then \( F \) is open cover compact. If \( \{ K_i \}_{i=1}^\infty \) is a finite collections of open cover compact subsets of \( X \) then \( K = \cup_{i=1}^\infty K_i \) is also an open cover 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 \( \mathcal{U}_0 \subset \mathcal{U} \). 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 \( \mathcal{U}_i \subset \mathcal{U} \) for each \( i \) such that \( K_i \subset \cup \mathcal{U}_i \). Then \( \mathcal{U}_0 := \cup_{i=1}^\infty \mathcal{U}_i \) is a finite cover of \( K \).

Exercise 9.29. Suppose \( f : X \to Y \) is continuous and \( K \subset X \) is open cover compact, then \( f(K) \) is an open cover compact subset of \( Y \). Give an example of continuous map, \( f : X \to Y \), and an open cover compact subset \( K \) of \( Y \) such that \( f^{-1}(K) \) is not open cover compact.

Exercise 9.30 (Extreme value theorem III). Let \( (X,d) \) be an open cover compact metric space and \( f : X \to \mathbb{R} \) be a continuous function. Show \( -\infty < \inf f \leq \sup f < \infty \) and there exists \( a,b \in X \) such that \( f(a) = \inf f \) and \( f(b) = \sup f \). Hint: use Exercise 9.29 and Theorem 9.47.

Exercise 9.31 (Uniform Continuity). Let \( (X,d) \) be an open cover compact metric space, \( (Y,\rho) \) 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 \( y,x \in X \) with \( d(y,x) < \delta \).

Exercise 9.32. Suppose \( f : X \to Y \) is continuous and \( K \subset X \) is open cover compact, then \( f(K) \) is an open cover compact subset of \( Y \). Give an example of continuous map, \( f : X \to Y \), and an open cover compact subset \( K \) of \( Y \) such that \( f^{-1}(K) \) is not open cover compact.

Exercise 9.33 (Dini’s Theorem). Let \( X \) be an open cover compact metric 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 9.49. A collection \( \mathcal{F} \) of closed subsets of a metric space \( (X,d) \) has the finite intersection property if \( \cap \mathcal{F}_0 \neq \emptyset \) for all \( \mathcal{F}_0 \subset \mathcal{F} \).
The notion of open cover compactness may be expressed in terms of closed sets as follows.

**Proposition 9.50.** A metric space $X$ is open cover compact if and only if every family of closed sets $\mathcal{F} \subseteq 2^X$ having the finite intersection property satisfies $\bigcap \mathcal{F} \neq \emptyset$.

**Proof.** The basic point here is that complementation interchanges open and closed sets and open covers go over to collections of closed sets with empty intersection. Here are the details.

($\Rightarrow$) Suppose that $X$ is open cover compact and $\mathcal{F} \subseteq 2^X$ is a collection of closed sets such that $\bigcap \mathcal{F} \neq \emptyset$. Let

$$U = \mathcal{F}^c := \{C^c : C \in \mathcal{F}\} \subseteq \tau,$$

then $U$ is a cover of $X$ and hence has a finite subcover, $U_0$. Let $\mathcal{F}_0 = U_0^c \subseteq \mathcal{F}$, then $\bigcap \mathcal{F}_0 = \emptyset$ so that $\mathcal{F}$ does not have the finite intersection property.

($\Leftarrow$) If $X$ is not open cover 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 $\bigcap \mathcal{F} = \emptyset$. 


9.7 Equivalence of Sequential and Open Cover Compactness in Metric Spaces

**Definition 9.51.** A metric space $(X,d)$ is $\varepsilon$-**bounded** ($\varepsilon > 0$) if there exists a finite cover of $X$ by balls of radius $\varepsilon$. We further say $(X,d)$ is **totally bounded** if it is $\varepsilon$-bounded for all $\varepsilon > 0$.

**Remark 9.52 (Totally bounded means almost finite).** An equivalent way to state that $(X,d)$ is $\varepsilon$-bounded is to say there exists a finite subset $A = A_\varepsilon \subseteq X$ such that $d(A(x)) < \varepsilon$ for all $x \in X$. In other words, $(X,d)$ is $\varepsilon$-bounded iff $X$ is a finite subset within an $\varepsilon$-error.

**Theorem 9.53.** Let $(X,d)$ be a metric space. The following are equivalent.

(a) $X$ is open cover compact.
(b) $X$ is sequentially compact.
(c) $X$ is totally bounded and complete.

**Proof.** The proof will consist of showing that $a \Rightarrow b \Rightarrow c \Rightarrow a$.

(a $\Rightarrow$ b) We will show that not $b \Rightarrow$ not $a$. Suppose there exists \( \{x_n\}_{n=1}^\infty \subseteq X \) which has no convergent subsequence. In this case the set $S := \{x_n : x_n \in X \}$ must be an infinite set as we have already seen finite sets are sequentially compact. For every $x \in X$ we must have

$$\varepsilon(x) := \liminf_{n \to \infty} d(x_n, x) > 0$$

since otherwise there would be a subsequence \( \{x_{n_k}\}_{k=1}^\infty \) such that $\lim_{k \to \infty} d(x_{n_k}, x) = 0$, i.e. $\lim_{k \to \infty} x_{n_k} = x$. Now we let $V_x := B_x (\frac{\varepsilon(x)}{2})$ and observe that $x_n$ can be in $V_x$ for only finitely many $n$ – otherwise we would conclude that $\liminf_{n \to \infty} d(x_n, x) \leq \frac{\varepsilon(x)}{2}$. From these observations, $U := \{V_x : x \in X\}$ is an open cover of $X$ with no finite subcover. Indeed, if $A \subseteq X$, then we must still have $x_n \in \bigcup_{x \in A} V_x$ for only finitely many $n$ and in particular $\bigcup_{x \in A} V_x$ cannot cover $S$.

(b $\Rightarrow$ c) Suppose $\{x_n\}_{n=1}^\infty \subseteq X$ 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. Now for sake of contradiction suppose that $X$ is not totally bounded. There there exists $\varepsilon > 0$ for which $X$ is not $\varepsilon$-bounded. In particular, $\mathcal{U} := \{B_x(\varepsilon) : x \in X\}$ is an open cover of $X$ with no finite subcover. We now use this to construct a sequence $\{x_n\}_{n=1}^\infty \subseteq X$. Choose $x_1 \in X$ at random, then choose $x_2 \in X \setminus B_{x_1}(\varepsilon)$, then $x_3 \in X \setminus (B_{x_1}(\varepsilon) \cup B_{x_2}(\varepsilon)) \cdots$

$$x_n \in X \setminus \left[ B_{x_1}(\varepsilon) \cup B_{x_2}(\varepsilon) \cup \cdots \cup B_{x_{n-1}}(\varepsilon) \right], \ldots.$$ The process may be continued indefinitely as $\mathcal{U}$ has no finite subcover. By construction we have chosen $\{x_n\}_{n=1}^\infty$ such that $d(x_1, \ldots, x_{n-1}, x_n) \geq \varepsilon$ for all $n$ and therefore $d(x_k, x_l) \geq \varepsilon$ for all $k \neq l$. Every subsequence will share this property, i.e. not be Cauchy, and hence can not be convergent.

(c $\Rightarrow$ a) For sake of contradiction, assume there exists an open cover $\mathcal{V} := \{V_a\}_{a \in A}$ of $X$ with no finite subcover. Since $X$ is totally bounded for each $n \in \mathbb{N}$ there exists $A_n \subseteq \mathcal{F} X$ such that

$$X = \bigcup_{x \in A_n} B_x (1/n) \subseteq \bigcup_{x \in A_n} C_x (1/n).$$

Choose $x_1 \in A_1$ such that no finite subset of $\mathcal{V}$ covers $K_1 := C_{x_1}(1)$. Since $K_1 = \bigcup_{a \in A_1} K_1 \cap C_x (1/2)$, there exists $x_2 \in A_2$ 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 9.6. Continuing this way inductively, we construct sets $K_n = K_{n-1} \cap C_{x_n}(1/n)$ with $x_n \in A_n$ 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 $\text{diam}(K_n) \leq 2/n$, it follows that $\{y_n\}$ is a Cauchy and hence convergent with
Suppose that \( y = \lim_{n \to \infty} y_n \in \cap_{m=1}^{\infty} K_m. \)

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 \( \text{diam}(K_n) \to 0 \), 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 \( \mathcal{V} \).

**Corollary 9.54.** The compact subsets of \( \mathbb{R}^n \) are the closed and bounded sets.

**Proof.** If \( K \) is closed and bounded then \( K \) is compact (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

\[
A_\delta = \delta \mathbb{Z}^n \cap [-M,M]^n := \{ \delta x : x \in \mathbb{Z}^n \text{ and } \delta |x_i| \leq M \text{ for } i = 1,2,\ldots,n \}.
\]

We will show, by choosing \( \delta > 0 \) sufficiently small, that

\[
K \subset [-M,M]^n \subset \cup_{x \in A_\delta} B(x,\varepsilon)
\]

which shows that \( K \) is totally bounded. Hence by Theorem 9.53 \( K \) is compact. Suppose that \( y \in [-M,M]^n \), then there exists \( x \in A_\delta \) such that \( |y_i - x_i| \leq \delta \) for \( i = 1,2,\ldots,n \). Hence

\[
d(x,y) = \sum_{i=1}^{n} (y_i - x_i)^2 \leq n\delta^2
\]

which shows that \( d(x,y) \leq \sqrt{n}\delta \). Hence if choose \( \delta < \varepsilon/\sqrt{n} \) we have shows that \( d(x,y) < \varepsilon \), i.e. Eq. (9.7) holds.

**9.8 Connectedness**

**Definition 9.55.** Let \((X,d)\) be a metric space. Two subset \( A \) and \( B \) of \( X \) are separated if \( A \cap B = \emptyset = A \cap B \). A set \( E \subset X \) is disconnected if \( E = A \cup B \) where \( A \) and \( B \) are two non-empty separated sets, otherwise \( E \) is said to be connected.

**Theorem 9.56 (The Connected Subsets of \( \mathbb{R} \)).** The connected subsets of \( \mathbb{R} \) are intervals.

**Proof.** We will break the proof into two parts. First we show if \( E \) is connected then \( E \) is an interval. Then we show if \( E \) is disconnected then \( E \) is not an interval.

1) Suppose that \( E \subset \mathbb{R} \) is a connected subset and that \( a,b \in E \) with \( a < b \). If there exists \( c \in (a,b) \) such that \( c \notin E \), then \( A := (-\infty,c) \cap E \) and \( B := (c,\infty) \cap E \) would be two non-empty separated subsets such that \( E = A \cup B \). Hence \( \{a,b\} \subset E \). Let \( \alpha := \inf(E) \) and \( \beta := \sup(E) \) and choose \( \alpha_n,\beta_n \in E \) 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, \( (\alpha_n,\beta_n) \subset E \) for all \( n \) and hence \( (\alpha,\beta) = \cup_{n=1}^{\infty} (\alpha_n,\beta_n) \subset E \). From this it follows that \( E = (\alpha,\beta), [\alpha,\beta), (\alpha,\beta) \text{ or } [\alpha,\beta), \text{ i.e. } E \) is an interval.

2) Now suppose that \( E \) is a disconnected subset of \( \mathbb{R} \). Then \( E = A \cup B \) where \( A \) and \( B \) are non-empty separated sets and let \( a \in A \) and \( b \in B \). After relabelling \( A \) and \( B \) if necessary we may assume that \( a < b \). Let \( p = \sup ([a,b] \cap A) \). Since \( p \in Â \cap [a,b] \) it follows that \( p \notin B \) and \( a \leq p \leq b \). Since \( b \in B, p \neq b \) and hence \( a \leq p < b \).

i) if \( p \notin A \) then \( p \notin E \) and \( a < p < b \) which shows \( E \) is not an interval.

ii) if \( p \in A \), then \( p \notin B \) and there exist \( p_1 \) such that \( a \leq p < p_1 < b \) and \( p_1 \notin B \). Again \( p_1 \notin A \) (for otherwise \( p \geq p_1 \)) and so \( p_1 \notin E \) and hence again \( E \) is not an interval.

**Lemma 9.57.** Suppose that \( f : X \to Y \) is a continuous function between two metric spaces \( (X,Y) \) and \( A \) and \( B \) are separated subsets of \( Y \). Then \( \alpha := f^{-1}(A) \) and \( \beta := f^{-1}(B) \) are separated subsets of \( X \). In particular, if \( E \subset X \) is connected, then \( f(E) \) is connected in \( Y \).

**Proof.** Since \( \alpha := f^{-1}(A) \subset f^{-1}(\overline{A}) \) and \( \overline{f^{-1}(\overline{A})} \) is the closed being the inverse image under a continuous function of the closed set \( \overline{A} \), it follows that \( \alpha \subset f^{-1}(\overline{A}) \). Thus if \( x \in \alpha \cap \beta \), then \( f(x) \in A \cap B = \emptyset \) and hence \( \alpha \cap \beta = \emptyset \). Similarly one shows \( \alpha \cap \overline{\beta} = \emptyset \) as well.

\(^2\) Notice that separated sets are disjoint. The sets \( A = (0,1) \) and \( B = [1,\infty) \) are disjoint but not separated.

\(^3\) This is because \( B^c \) is an open subset of \( \mathbb{R} \).
If \( f(E) \) disconnected, there exists non-empty separated subsets, \( A \) and \( B \) of \( f(E) \), so that \( f(X) = A \cup B \). The sets \( \alpha := f^{-1}(A) \) and \( \beta := f^{-1}(B) \) are now non-empty separated subsets of \( X \). It now follows that \( \alpha_0 := \alpha \cap E \) and \( \beta_0 := \beta \cap E \) are non-empty sets such that
\[
\alpha_0 \cap \beta_0 \subset \alpha \cap \beta = \emptyset \quad \text{and} \quad \alpha_0 \cap \beta_0 \subset \alpha \cap \beta = \emptyset
\]
which would imply \( E \) is disconnected. Thus if \( E \) is connected we must have that \( f(E) \) is connected.

**Theorem 9.58 (Intermediate Value Theorem).** Suppose that \((X, d)\) is a connected metric 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 exists \( z \in X \) such that \( f(z) = c \).

**Proof.** By Lemma 9.57 \( f(X) \) is a connected subset of \( \mathbb{R} \). So by Theorem 9.56 \( f(X) \) is a subinterval of \( \mathbb{R} \) and this completes the proof.

**Lemma 9.59.** If \( E \subset X \) is a connected set and \( E \subset A \cup B \) where \( A \) and \( B \) are two separated sets, then \( E \subset A \) or \( E \subset B \).

**Proof.** Let \( \alpha := E \cap A \) and \( \beta := E \cap B \), then \( \alpha \cap \beta \subset A \cap B = \emptyset \) and similarly \( \alpha \cap \beta = \emptyset \). Thus \( E = \alpha \cup \beta \) with \( \alpha \) and \( \beta \) being separated sets. Since \( E \) is connected we must have \( \alpha = \emptyset \) or \( \beta = \emptyset \), i.e. \( E \subset A \) or \( E \subset B \).

**Proposition 9.60.** Suppose that \( F \) and \( G \) are connected subsets of \( X \) such that \( F \cap G \neq \emptyset \), then \( F \cap G \cap \overline{X} = \emptyset \).

**Proof.** Suppose that \( E = A \cup B \) where \( A \) and \( B \) are separated sets. Then from Lemma 9.59 we know that \( F \subset A \) or \( F \subset B \) and similarly \( G \subset A \) or \( G \subset B \). If both \( F \) and \( G \) are in the same set (say \( A \)), then \( E \subset A \subset E \) and \( B \) must be empty. On the other hand if \( F \subset A \) and \( G \subset B \), then \( \emptyset \neq F \cap G \subset A \cap B \) which would violate \( A \) and \( B \) being separated. Thus we have shown there does not exist two non-empty separated sets \( A \) and \( B \) such that \( E = A \cup B \), i.e. \( E \) is connected.

**Definition 9.61.** A subset \( E \) of a metric space \( X \) is path connected if to every pair of points \( \{x_0, x_1\} \subset E \) there exists \( \sigma \in C([0, 1], E) \), such that \( \sigma(0) = x_0 \) and \( \sigma(1) = x_1 \). We refer to \( \sigma \) as a path joining \( x_0 \) to \( x_1 \).

**Proposition 9.62.** Every path connected subset, \( E \), of a metric space \( X \) is connected.

**Exercise 9.34.** Prove Proposition 9.62 i.e. if \( E \subset X \) is path connected then \( E \) is connected. **Hint:** for sake of contradiction suppose that \( A \) and \( B \) are two non-empty separated subsets of \( X \) such that \( E = A \cup B \) and choose a path connecting a point in \( A \) to a point in \( B \).

**Definition 9.63.** A subset, \( C \), of a vector space, \( X \), is convex if for all \( a, b \in C \) the path,
\[
\sigma(t) = a + t(b - a) = (1 - t)a + tb \quad \text{for} \quad 0 \leq t \leq 1
\]
is contained in \( C \).

**Example 9.64.** Every convex subset, \( C \), of a normed vector space \((X, \|\cdot\|)\) is path connected and hence connected.

**Exercise 9.35.** Suppose that \((X, \|\cdot\|)\) is a normed space, \( x \in X \), and \( R > 0 \). Show the open and closed balls in \( X, B_x(R) \) and \( C_x(R) \), are both convex sets and hence path connected.

**Definition 9.65.** A metric space \( X \) is locally path connected if for each \( x \in X \), there is an open neighborhood \( V \subset X \) of \( x \) which is path connected.

**Proposition 9.66.** Let \( X \) be a metric space.

1. If \( X \) is connected and locally path connected, then \( X \) is path connected.
2. If \( X \) is any connected open subset of \( \mathbb{R}^n \), then \( X \) is path connected.

**Exercise 9.36.** Prove item 1. of Proposition 9.66 i.e. if \( X \) is connected and locally path connected, then \( X \) is path connected. **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 9.37.** Prove item 2. of Proposition 9.66 i.e. if \( X \) is any connected open subset of \( \mathbb{R}^n \), then \( X \) is path connected.

**9.8.1 Connectedness Problems**

**Exercise 9.38.** In this exercise we will work inside the metric space \( \mathbb{Q} \) with \( d(x, y) := |x - y| \) for all \( x, y \in \mathbb{Q} \). Let \( a, b \in \mathbb{Q} \) with \( a < b \) and let
\[
J := [a, b] \cap \mathbb{Q} = \{x \in \mathbb{Q} : a \leq x \leq b\}
\]
Show \( J \) is disconnected in \( \mathbb{Q} \).

**Exercise 9.39.** 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 9.58), then \( f \) is continuous.

**Exercise 9.40.** Suppose \( -\infty < a < b \leq \infty \) and \( f : [a, b) \to \mathbb{R} \) is a strictly increasing continuous function. Using the intermediate value theorem, one sees that \( f([a, b)) \) is an interval and since \( f \) is strictly increasing it must of the form \([c, d)\) for some \( c \in \mathbb{R} \) and \( d \in \mathbb{R} \) 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 \( n^{\text{th}} \) root of a real number. Compare with Exercise ??.
Exercise 9.41. Let

\[ X := \{(x, y) \in \mathbb{R}^2 : y = \sin(x^{-1}) \text{ with } x \neq 0\} \cup \{(0, 0)\} \]

equipped with the relative topology induced from the standard topology on \( \mathbb{R}^2 \).

Show \( X \) is connected but not path connected.

Remark 9.67 (Structure of open sets in \( \mathbb{R} \)). Let \( V \subset \mathbb{R} \) be an open set. For \( x \in V \), let \( a_x := \inf \{a : (a, x] \subset V\} \) and \( b_x := \sup \{b : [x, b) \subset V\} \). Since \( V \) is open, \( a_x < x < b_x \) and it is easily seen that \( J_x := (a_x, b_x) \subset V \). Moreover if \( y \in V \) and \( J_x \cap J_y \neq \emptyset \), then \( J_x = J_y \). The collection, \( \{J_x : x \in V\} \), is at most countable since we may label each \( J \in \{J_x : x \in V\} \) by choosing a rational number \( r \in J \). Letting \( \{J_n : n < N\} \), with \( N = \infty \) allowed, be an enumeration of \( \{J_x : x \in V\} \), we have \( V = \bigsqcup_{n<N} J_n \) as desired.
Part III

Calculus
Differential Calculus in One Real Variable

10.1 Basics of the Derivative

Definition 10.1. If \( f: J \to X \) is a function we say \( \lim_{t \to t_0} f(t) = x \in X \) iff for all \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that

\[
\|f(t) - x\| \leq \varepsilon \text{ whenever } 0 < |t - t_0| \leq \delta.
\]

A function \( f: J \to X \) is said to be differentiable at \( t_0 \in J \) iff \( \lim_{t \to t_0} \frac{f(t) - f(t_0)}{t - t_0} \) exists in \( X \). We denote the limit by \( \dot{f}(t_0) \) or \( f'(t_0) \) or by \( df(t_0) \), i.e.

\[
\dot{f}(t_0) = \lim_{t \to t_0} \frac{f(t) - f(t_0)}{t - t_0}
\]

Example 10.2. If \( f(t) = at + b \), then \( \dot{f}(t) = a \) for all \( t \in \mathbb{R} \). Indeed,

\[
\frac{f(t) - f(t_0)}{t - t_0} = \frac{a(t - t_0)}{t - t_0} = a \text{ for all } t \neq t_0.
\]

Remark 10.3. One may also define the derivative of \( f \) by

\[
\dot{f}(t_0) = \lim_{h \to 0} \frac{f(t_0 + h) - f(t_0)}{h}
\]

provided the limit exists. To see this is the case let \( \Delta(t) := \left| \frac{f(t) - f(t_0)}{t - t_0} - \dot{f}(t_0) \right| \).

Then \( \lim_{h \to 0} \Delta(t_0 + h) = 0 \) iff for all \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that

\( \Delta(t_0 + h) \leq \varepsilon \) if \( 0 < |h| \leq \delta \). Letting \( t = t_0 + h \), we see this is equivalent to \( \Delta(t) \leq \varepsilon \) if \( 0 < |h| = |t - t_0| \leq \delta \), i.e. iff \( \lim_{h \to 0} \Delta(t) = 0 \).

Proposition 10.4. The function \( f(t) := 1/t = t^{-1} \) defined for \( t \neq 0 \) is differentiable and \( \dot{f}(t) = -1/t^2 \) for all \( t \neq 0 \).

Proof. We have

\[
\frac{f(t+h) - f(t)}{h} = \frac{1}{h} \left[ \frac{1}{t+h} - \frac{1}{t} \right] = \frac{1}{h} \cdot \frac{t - (t+h)}{(t+h)t}
\]

\[
= -\frac{1}{(t+h)t} \to -\frac{1}{t^2} \text{ as } h \to 0.
\]

Here we have used the fact that \( \frac{1}{t} \) is continuous for the last limit.

Notation 10.5 \( (\varepsilon(\cdot), O(\cdot), o(\cdot) \text{ Notation}) \) To simplify proofs below, we will often let \( \varepsilon(h), O(h), \) and \( o(h) \) denote generic \( X \) or \( \mathbb{R} \) - valued functions defined in some deleted neighborhood of \( 0 \in \mathbb{R} \) (i.e. for \( h \in (-\delta, \delta) \setminus \{0\} \)) such that

1. \( \lim_{h \to 0} \varepsilon(h) = 0 \),
2. there exists \( C < \infty \) such that \( \|O(h)\| \leq C|h| \) for \( h \) near 0 or equivalently,
3. \( \lim_{h \to 0} \frac{o(h)}{h} = 0 \), i.e. for all \( c > 0 \) there exists \( \delta = \delta(c) > 0 \) such that \( \|o(h)\| \leq c|h| \) for 0 < \( |h| \leq \delta \).

We will often assume that each of these functions has been extended to \( h = 0 \) by setting there values at 0 to be 0.

Remark 10.6. Using the above notation we have, \( O(h) \varepsilon(h) = o(h) \), \( O(h) + O(h) = O(h) + o(h) \), \( O(h) + \varepsilon(h) = \varepsilon(h) + \varepsilon(h) \), etc. etc.

We may also write, \( \lim_{h \to 0} f(t) = x \) as \( f(t_0 + h) = x + \varepsilon(h) \) and \( f \) is differentiable at \( t_0 \) with derivative \( \dot{f}(t_0) \) iff

\[
f(t_0 + h) = f(t_0) + \dot{f}(t_0) h + o(h).
\]

Lemma 10.7. If \( f \) is differentiable at \( t_0 \) then it is continuous at \( t_0 \).

Proof. By definition,

\[
f(t_0 + h) = f(t_0) + f'(t_0) h + o(h) = f(t_0) + \varepsilon(h)
\]

and so \( \lim_{h \to 0} f(t_0 + h) = f(t_0) \).

Theorem 10.8 (Linearity of \( \frac{d}{dt} \)). Suppose that \( f: (a,b) \to X \) and \( g: (a,b) \to X \) are differentiable at \( t_0 \in (a,b) \) and \( \lambda \in \mathbb{R} \), then

\[
\frac{d}{dt} (f + \lambda g)(t_0) = f'(t_0) + \lambda g'(t_0).
\]

Proof. We are given,

\[
f(t_0 + h) = f(t_0) + \dot{f}(t_0) h + o(h) \quad \text{and} \quad g(t_0 + h) = g(t_0) + \dot{g}(t_0) h + o(h)
\]

and therefore,
Exercise 10.2. Suppose that \( f : (a, b) \to \mathbb{R} \) and \( g : (a, b) \to X \) are differentiable at \( t_0 \in (a, b) \), then

\[
\frac{d}{dt} (f \cdot g) (t_0) = f'(t_0) g(t_0) + f(t_0) g'(t_0).
\]  

(10.1)

**Proof.** We have,

\[
(f \cdot g)(t_0 + h) = f(t_0 + h) \cdot g(t_0 + h)
\]

\[
= \left( f(t_0) + f'(t_0) h + o(h) \right) \left( g(t_0) + g'(t_0) h + o(h) \right)
\]

\[
= f(t_0)g(t_0) + f'(t_0) g(t_0) h + f(t_0) g'(t_0) h + o(h)
\]  

(10.2)

wherein we have used \( h^2 = o(h) \), \( ho(h) = o(h) \), and \( o(h) o(h) = o(h) \). From Eq. (10.2) we deduce Eq. (10.1).  

---

Lemma 10.10. We have the following estimates, \( o(O(h)) = o(h) \) and \( O(o(h)) = o(h) \).

**Proof.** Suppose that \( f(h) \) and \( g(h) \) are functions such that \( f(h) = o(h) \) and \( g(h) = O(h) \). These statements are equivalent to the existence of a constant \( C < \infty \) and a function \( \varepsilon(h) \) such that \( \lim_{h \to 0} \varepsilon(h) = 0 \) and

\[
|f(h)| \leq |\varepsilon(h)| |h| \quad \text{and} \quad |g(h)| \leq C |h|
\]

for \( h \) near 0.

With these estimates in hand we have

\[
|f(g(h))| \leq |\varepsilon(g(h))| |g(h)| \leq C |g(h)| |h|
\]

and

\[
|g(f(h))| \leq C |f(h)| \leq C |\varepsilon(h)| |h|.
\]

Since \( \lim_{h \to 0} C |\varepsilon(h)| = 0 \) we see that \( f(g(h)) = o(h) \) and since \( \lim_{h \to 0} O |\varepsilon(h)| = 0 \) it follows that \( g(f(h)) = o(h) \).

---

Theorem 10.11 (Chain Rule). Suppose that \( g : (c, d) \to X \) and \( f : (a, b) \to (c, d) \) are given functions such that \( f(t_0) \) at \( t_0 \in (a, b) \) and \( g(w_0) \) exists at \( z_0 := f(t_0) \). Then

\[
\frac{d}{dt} g(f(t)) \big|_{t=t_0} = g'(z_0) f'(t_0) = g'(f(t)) f'(t_0).
\]

**Proof.** Let \( h(t) := g(f(t)) \). Then by assumption,

\[
(f(t_0 + \Delta t) - f(t_0)) = g(z_0 + \Delta z) = g(z_0) + g'(z_0) \Delta z + o(\Delta z).
\]

Letting \( \Delta z := f(t_0 + \Delta t) - f(t_0) = f'(t_0) \Delta t + o(\Delta t) \) which is small for \( \Delta t \) small, we find,

\[
h(t_0 + \Delta t) = g(f(t_0 + \Delta t)) = g(z_0 + \Delta z)
\]

\[
= g(z_0) + g'(z_0) \Delta z + o(\Delta z)
\]

\[
= h(t_0) + g'(z_0) [f'(t_0) \Delta t + o(\Delta t)] + o(\Delta t) = h(t_0) + g'(z_0) f'(t_0) \Delta t + o(\Delta t).
\]

(10.4)

wherein we have used Lemma 10.10 to conclude, \( o(f'(t_0) \Delta t + o(\Delta t)) = o(\Delta t) \). The result follows Eq. (10.4) and the definition of the derivative.

---

Corollary 10.12 (Quotient Rule). Suppose that \( f : (a, b) \to \mathbb{R} \) and \( g : (a, b) \to X \) are differentiable at \( t_0 \in (a, b) \) and \( f(t_0) \neq 0 \), then

\[
\frac{d}{dt} \left( \frac{g}{f} \right)(t_0) = \frac{f(t_0)g'(t_0) - f'(t_0)g(t_0)}{f^2(t_0)}.
\]
10.2 First Derivative Test and the Mean Value Theorems

Definition 10.13. A function \( f : (a, b) \to \mathbb{R} \) has as local minimum (maximum) at \( t_0 \in (a, b) \) if there exists \( \delta > 0 \) such that \( f(t) \geq f(t_0) \) (\( f(t) \leq f(t_0) \)) when \( |t - t_0| \leq \delta \).

Theorem 10.14 (First Derivative Test). If \( f : (a, b) \to \mathbb{R} \) has a local extremum (i.e., local minimum or maximum) at \( t_0 \in (a, b) \) and \( f \) is differentiable at \( t_0 \), then \( f'(t_0) = 0 \).

Proof. For sake of definiteness, suppose that \( t_0 \) is a local minimum for \( f \). Then for \( h > 0 \) small,

\[
\frac{f(t_0 + h) - f(t_0)}{h} \geq 0 \quad \text{and} \quad \frac{f(t_0 - h) - f(t_0)}{-h} \leq 0.
\]

Therefore letting \( h \downarrow 0 \) in these two inequalities leads to \( f'(t_0) \geq 0 \) and \( f'(t_0) \leq 0 \) which implies \( f'(t_0) = 0 \).

Exercise 10.4 (Existence of Eigenvectors for Symmetric Matrices). Let \( A \) be an \( n \times n \) symmetric real matrix. Recall this implies (and is equivalent to) \( Ax \cdot y = x \cdot Ay \) for all \( x, y \in \mathbb{R}^n \) where \( \cdot \) is the dot product on \( \mathbb{R}^n \) as in Eq. \( \ref{eq:dot_product} \). In this problem you are going to show that \( A \) has an eigenvector by proving the following statements.

1. Show \( |Ax \cdot y| \leq C \|x\| \|y\| \) where \( C := \sqrt{\sum_{i,j} A_{ij}^2} \) – the Hilbert-Schmidt norm of \( A \).

2. Show \( q(x) := Ax \cdot x \) is a continuous function on \( \mathbb{R}^n \).

Let \( S := \{ x \in \mathbb{R}^n : \|x\| = 1 \} \) be the closed and bounded and hence compact unit sphere in \( \mathbb{R}^n \). By the extreme value theorem, \( \lambda := \max_{x \in S} q(x) \) exists. Let \( x_0 \in S \) be chosen so that

\[
\lambda = q(x_0) = Ax_0 \cdot x_0.
\]

3. Explain why \( R(x) = \frac{Ax_0}{\|x_0\|} = q \left( \frac{x}{\|x\|} \right) \) is a continuous function on \( \mathbb{R}^n \setminus \{0\} \) and show \( x_0 \) is also a maximizer of \( R \).

4. For any \( v \in \mathbb{R}^n \) let \( f_v(t) = R(x_0 + tv) \) which is defined for \( t \) such that \( \|x_0 + tv\| \neq 0 \) and in particular for \( t \) near zero. Compute \( f_v'(t) \) for \( t \) near 0 and in particular show

\[
f_v(0) = 2(Ax_0 - \lambda x_0) \cdot v.
\]

5. Noting that \( f_v(t) \) has a maximum at \( t = 0 \), the first derivative implies \( f_v'(0) = 0 \). By making a judicious choice for \( v \), show \( Ax_0 = \lambda x_0 \), i.e. \( x_0 \) is an eigenvector of \( A \) with eigenvalue \( \lambda \).

Exercise 10.5 (Multiplication Operators). Let \( V := C([0, 1], \mathbb{R}) \) and \( A : V \to V \) be the operation of multiplication by \( t \), i.e. for \( f \in V \) let \( Af \in V \) be defined by

\[
(Af)(t) = tf(t) \quad \text{for all} \quad t \in [0, 1].
\]

Show: if there is a \( \lambda \in \mathbb{R} \) and \( f \in V \) such that \( Af = \lambda f \), then \( f \equiv 0 \), i.e. \( A \) has no eigenvectors (or eigenfunctions if you prefer).

[Moral: The reader should compare this exercise with Exercise 10.4. Notice that if we define an inner product on \( V \) by

\[
(f, g) = \int_0^1 f(t) g(t) \, dt
\]

then \( (Af, g) = (f, Ag) \), i.e. \( A \) is symmetric. Nevertheless, \( A \) has no eigenvectors. This may be considered as another proof the that unit sphere in \( V \) is not compact!]

Theorem 10.15 (Rolle’s Theorem). Suppose that \( f : [a, b] \to \mathbb{R} \) is a continuous function such that \( f(a) = L \) and \( f(b) = f(a) \). Then there exists \( t_0 \in (a, b) \) such that \( f'(t_0) = 0 \).

Proof. If \( f \) is not constant then there exists \( t \in (a, b) \) such that either \( f(t) > L \) or \( f(t) < L \). For definiteness let assume the first case. Then by the extreme value theorem, there exists \( t_0 \in [a, b] \) such that \( f(t_0) \geq f(t) \) for all \( t \in [a, b] \). Moreover this maximizer can not be at \( a \) or \( b \) as \( f(a) = f(b) = f(a) \). Therefore \( f \) has a maximum at some point \( t_0 \in (a, b) \). It then follows by the first derivative test that \( f'(t_0) = 0 \).
The result easily follow from this identity.

**Theorem 10.16 (Mean Value Theorem).** Suppose that \( f : [a, b] \to \mathbb{R} \) is a continuous function such that \( f \) is differentiable on \((a, b)\). Then there exists \( t_0 \in (a, b) \) such that

\[
\dot{f}(t_0) = \frac{f(b) - f(a)}{b - a}.
\]

**Proof.** Let

\[
h(t) := f(t) - \left( f(a) + \frac{t - a}{b - a} (f(b) - f(a)) \right).
\]

Then \( h(a) = 0 = h(b) \) and so by Rolle’s theorem, there exist \( t_0 \in (a, b) \) such that

\[
0 = h(t_0) = \dot{f}(t_0) - \frac{1}{b - a} (f(b) - f(a)) = \dot{f}(t_0) - \frac{f(b) - f(a)}{b - a}.
\]

Alternatively, apply the generalized mean value Theorem \[10.46\] below with \( g(t) = t \).

**Corollary 10.17.** Suppose that \( f : (a, b) \to \mathbb{R} \) is differentiable on \((a, b)\).

1. If \( f' \geq 0 \) on \((a, b)\) then \( f \) is non-decreasing on \((a, b)\).
2. If \( f' \leq 0 \) on \((a, b)\) then \( f \) is non-increasing on \((a, b)\).
3. If \( f' = 0 \) on \((a, b)\), then \( f \) is constant.

**Proof.** For \( a < x_1 < x_2 < b \) there exists \( c = c(x_1, x_2) \in (x_1, x_2) \) such that

\[
f(x_2) - f(x_1) = f'(c(x_1, x_2))(x_2 - x_1).
\]

The result easily follow from this identity.

**Remark 10.18.** The mean value Theorem \[10.16\] does not hold for vector valued functions and in particular not for \( \mathbb{C} \) valued functions. For example let \( f(t) = (u(t), v(t)) \in \mathbb{R}^2 \) be differentiable function such that \( f(0) = f(1) = 0 \in \mathbb{R}^2 \).

We can arrange for \( \dot{u}(t) \) to only be zero at \( \frac{1}{3} \) and \( \dot{v}(t) \) to be zero only at \( 3/4 \). In this case both \( u \) and \( v \) satisfy the mean value theorem with two different times and there is no common time \( t_0 \) such that \( \dot{u}(t_0) = 0 = \dot{v}(t_0) \). However, the following mean value inequality does still hold.

**Theorem 10.19 (Mean Value Inequality).** Suppose that \( X = \mathbb{R}^n \) and \( f : [a, b] \to X = \mathbb{R}^n \) is a continuous function which is differentiable on \((a, b)\). Then there exists \( c \in (a, b) \) such that

\[
\|f(b) - f(a)\| \leq \|\dot{f}(c)\|(b - a).
\]

In particular, if \( f : [a, b] \to \mathbb{C} \cong \mathbb{R}^2 \) then there exists \( c \in (a, b) \) such that

\[
|f(b) - f(a)| \leq \|\dot{f}(c)\|(b - a).
\]

[Although we do not prove it now, this result remains true when \( X \) is any normed vector space however the proof would require the “Hahn-Banach” theorem. We will give a proof of a closely related result later, see Proposition \[12.12\] and Corollary \[12.21\] below.]

**Proof.** For \( v \in \mathbb{R}^n \), let \( f_v(t) := v \cdot f(t) \). Applying the mean value theorem to \( f_v \) shows there exists \( c(v) \in (a, b) \) such that

\[
v \cdot [f(b) - f(a)] = v \cdot f(b) - v \cdot f(a) = \frac{d}{dt} v \cdot f(t) \big|_{t=c(v)} (b - a)
\]

\[
= [v \cdot \dot{f}(c(v))] (b - a)
\]

in Eq. \[10.5\] in order to learn,

\[
\|f(b) - f(a)\| = v \cdot [f(b) - f(a)] = \|v\| \|\dot{f}(c(v))\|(b - a)
\]

\[
\leq \|v\| \|\dot{f}(c(v))\| (b - a) = \|\dot{f}(c(v))\| (b - a).
\]

**Corollary 10.20.** Suppose that \( f : [a, b] \to \mathbb{R}^n \) is a continuous function such that \( f'(t) = 0 \) for all \( t \in (a, b) \), then \( f(t) = f(a) \) for all \( t \in [a, b] \), i.e. \( f \) is constant on \([a, b]\).

**Proof.** By the mean value inequality for all \( \beta \in [a, b] \) we have

\[
\|f(\beta) - f(a)\| \leq \|f'(c)\| (\beta - a) = 0
\]

for some \( c \in (a, \beta) \). This shows that \( f(\beta) = f(a) \) for all \( \beta \in [a, b] \).

**10.3 Limsup and Liminf revisited**

For proofs to follow it will be convenient to introduce the notion of limsup and liminf for function \( g : Y \setminus \{y_0\} \to \mathbb{R} \), where \( Y \) is a metric space and \( y_0 \in Y \). [The reader will find many more details regarding the material in this section in the Appendix \[3\] which has been generously supplied by Ali Behzadan.]
Notation 10.21 The deleted open ball at \( y_0 \) with radius \( \delta > 0 \) is defined by
\[
B'_{y_0}(\delta) := B_{y_0}(\delta) \setminus \{y_0\} = \{y \in Y : 0 < d(y, y_0) < \delta\}.
\]

Definition 10.22. Given \( g : Y \setminus \{y_0\} \to \mathbb{R} \) we let
\[
\limsup_{y \to y_0} g(y) := \lim_{\delta \downarrow 0} \sup_{y \in B'_{y_0}(\delta)} g(y)
\]
and
\[
\liminf_{y \to y_0} g(y) := \lim_{\delta \downarrow 0} \inf_{y \in B'_{y_0}(\delta)} g(y).
\]

These limits always exist since \( \sup_{y \in B'_{y_0}(\delta)} g(y) \) decreases as \( \delta \) decreases while \( \inf_{y \in B'_{y_0}(\delta)} g(y) \) increases as \( \delta \) decreases. Moreover
\[
\inf_{y \in B'_{y_0}(\delta)} g(y) \leq \sup_{y \in B'_{y_0}(\delta)} g(y) \text{ for all } \delta > 0,
\]
and hence by passing to the limit as \( \delta \downarrow 0 \),
\[
\liminf_{y \to y_0} g(y) \leq \limsup_{y \to y_0} g(y).
\]
The following proposition is analogous to Proposition 3.28 and the reader may find a more general version in Proposition 10.9.

Proposition 10.23. Keeping the notation introduced above, \( \lim_{y \to y_0} g(y) \) exists in \( \mathbb{R} \) iff
\[
\liminf_{y \to y_0} g(y) = \limsup_{y \to y_0} g(y)
\]
in which case
\[
\lim_{y \to y_0} g(y) = \liminf_{y \to y_0} g(y) = \limsup_{y \to y_0} g(y). \tag{10.6}
\]

Proof. (\( \Longleftarrow \)) Suppose that \( L := \liminf_{y \to y_0} g(y) = \limsup_{y \to y_0} g(y) \). If \( L \in \mathbb{R} \), then given \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that
\[
\left| \inf_{y \in B'_{y_0}(\delta)} g(y) - L \right| \leq \varepsilon \text{ and } \left| \sup_{y \in B'_{y_0}(\delta)} g(y) - L \right| \leq \varepsilon.
\]
In particular we will have
\[
L - \varepsilon \leq \inf_{y \in B'_{y_0}(\delta)} g(y) \leq \sup_{y \in B'_{y_0}(\delta)} g(y) \leq L + \varepsilon
\]
and so
\[
L - \varepsilon \leq g(y) \leq L + \varepsilon \text{ for all } y \in B'_{y_0}(\delta).
\]

which, by definition, implies \( \lim_{y \to y_0} g(y) = L \).

Next suppose that \( L = \liminf_{y \to y_0} g(y) = \limsup_{y \to y_0} g(y) \). Then for all \( 0 < M < \infty \) there exists \( \delta > 0 \) such that \( \inf_{y \in B'_{y_0}(\delta)} g(y) \geq M \), i.e.
\[
g(y) \geq M \text{ if } 0 < d(y, y_0) \leq \delta.
\]
By definition, this implies \( \lim_{y \to y_0} g(y) = \infty \).
The case \( L = -\infty \) is similar and will be omitted.

(\( \Longrightarrow \)) Suppose that \( L := \lim_{y \to y_0} g(y) \) exists in \( \mathbb{R} \). If \( L \in \mathbb{R} \) then for all \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that
\[
|g(y) - L| \leq \varepsilon \text{ if } 0 < d(y, y_0) < \delta,
\]
i.e.
\[
L - \varepsilon \leq g(y) \leq L + \varepsilon \text{ if } y \in B'_{y_0}(\delta).
\]
From this it follows that
\[
L - \varepsilon \leq \inf_{y \in B'_{y_0}(\delta)} g(y) \leq \sup_{y \in B'_{y_0}(\delta)} g(y) \leq L + \varepsilon
\]
and therefore letting \( \delta \downarrow 0 \) shows
\[
L - \varepsilon \leq \liminf_{y \to y_0} g(y) \leq \limsup_{y \to y_0} g(y) \leq L + \varepsilon.
\]
We may now let \( \varepsilon \downarrow 0 \) to conclude that
\[
L = \lim_{y \to y_0} g(y) \leq \limsup_{y \to y_0} g(y) = L
\]
which suffices to prove omitted [10.6] in this case.

Next suppose that \( L = \infty \). Then for every \( M \in (0, \infty) \) there exists \( \delta = \delta(M) \) such that \( g(y) \geq M \) if \( y \in B'_{y_0}(\delta) \) and therefore,
\[
\lim_{y \to y_0} g(y) = \lim_{\delta \downarrow 0} \inf_{y \in B'_{y_0}(\delta)} g(y) \geq M.
\]
We may now let \( M \uparrow \infty \) to learn
\[
\lim_{y \to y_0} g(y) = \infty \leq \liminf_{y \to y_0} g(y) \leq \limsup_{y \to y_0} g(y)
\]
which again proves Eq. (10.6). The case where \( L = -\infty \) is proved similarly so will be omitted. \( \blacksquare \)
Given a function,

**Lemma 10.25.** If \( n \in \mathbb{N} \) and \( g_k : Y \setminus \{y_0\} \to [0, \infty] \) are given functions for \( 1 \leq k \leq n \), then

\[
\limsup_{y \to y_0} \sum_{k=1}^{n} g_k(y) \leq \sum_{k=1}^{n} \limsup_{y \to y_0} g_k(y).
\]

**Proof.** Since

\[
\sum_{k=1}^{n} g_k(z) \leq \sum_{k=1}^{n} \sup_{y \in B_{y_0}(\delta)} g_k(y)
\]

for all \( z \in B'_{y_0}(\delta) \), it follows that

\[
\sup_{z \in B'_{y_0}(\delta)} \sum_{k=1}^{n} g_k(z) \leq \sum_{k=1}^{n} g_k(y).
\]

Passing the limit as \( \delta \to 0 \) in this inequality gives Eq. (10.7).

**Notation 10.26 (One sided limits)** Given a function, \( f : (a, b) \to \mathbb{R} \) we let

\[
\limsup_{x \uparrow b} f(x) := \lim \sup_{x \uparrow b} f(y) \quad \text{and} \quad \liminf_{x \uparrow b} f(x) := \lim \inf_{x \uparrow b} f(y).
\]

Similarly we let

\[
\limsup_{x \downarrow a} f(x) := \lim \sup_{x \downarrow a} f(y) \quad \text{and} \quad \liminf_{x \downarrow a} f(x) := \lim \inf_{x \downarrow a} f(y).
\]

**Proposition 10.27.** Keeping the notation introduced;

1. \( \liminf_{x \uparrow b} f(x) \leq \limsup_{x \uparrow b} f(x) \) and \( \liminf_{x \downarrow a} f(x) \leq \limsup_{x \downarrow a} f(x) \),
2. \( \liminf_{x \uparrow b} f(x) = \limsup_{x \uparrow b} f(x) = L \in \mathbb{R} \) if \( \lim_{x \uparrow b} f(x) = L \in \mathbb{R} \), and
3. \( \liminf_{x \downarrow a} f(x) = \limsup_{x \downarrow a} f(x) = L \in \mathbb{R} \) if \( \lim_{x \downarrow a} f(x) = L \in \mathbb{R} \).

### 10.4 Differentiating past infinite sums

**Theorem 10.28 (Weierstrass M-test II).** Suppose that \((X, \|\cdot\|)\) is a Banach space, \((Y, d)\) is a metric space, for each \( n \in \mathbb{N} \), \( f_n : Y \setminus \{y_0\} \to X \) is a function, and there exists \( \{M_n\}_{n=1}^{\infty} \subset [0, \infty) \) with \( \sum_{n=1}^{\infty} M_n < \infty \) such that

\[
\sup_{y \in Y} \|f_n(y)\|_X \leq M_n \quad \forall n \in \mathbb{N}.
\]

Under the above assumptions, if \( \lim_{y \to y_0} f_n(y) \) exists in \( X \) for all \( n \in \mathbb{N} \), then

\[
\lim_{y \to y_0} \sum_{n=1}^{\infty} f_n(y) = \sum_{n=1}^{\infty} \lim_{y \to y_0} f_n(y).
\]

**Proof.** Let \( x_n : = \lim_{y \to y_0} f_n(y) \). As the norm on \( X \) is continuous it follows that \( \|x_n\| \leq M_n \) for all \( n \) and therefore \( \sum_{n=1}^{\infty} x_n \) is absolutely convergent. To finish the proof we must show \( \lim_{y \to y_0} \sum_{n=1}^{\infty} f_n(y) = x \). We begin with the estimate;

\[
\left\| x - \sum_{n=1}^{\infty} f_n(y) \right\| = \left\| \sum_{n=1}^{\infty} [x_n - f_n(y)] \right\|
\]

\[
= \left\| \sum_{n=1}^{N} [x_n - f_n(y)] + \sum_{n=N+1}^{\infty} [x_n - f_n(y)] \right\|
\]

\[
\leq \sum_{n=1}^{N} \|x_n - f_n(y)\| + \sum_{n=N+1}^{\infty} \|x_n + f_n(y)\|
\]

\[
\leq \sum_{n=1}^{N} \|x_n - f_n(y)\| + 2 \sum_{n=N+1}^{\infty} M_n.
\]

Taking the \( \limsup_{y \to y_0} \) of this inequality and letting \( N \to \infty \) shows

\[
\limsup_{y \to y_0} \left\| x - \sum_{n=1}^{\infty} f_n(y) \right\| \leq \limsup_{y \to y_0} \sum_{n=1}^{N} \|x_n - f_n(y)\| + 2 \sum_{n=N+1}^{\infty} M_n
\]

\[
\leq \sum_{n=1}^{N} \limsup_{y \to y_0} \|x_n - f_n(y)\| + 2 \sum_{n=N+1}^{\infty} M_n
\]

\[
= 2 \sum_{n=N+1}^{\infty} M_n \to 0 \quad \text{as} \quad N \to \infty.
\]

**Alternative ending #1:** for those that skipped Section 10.3 above, taking supremum over \( y \) such that \( sup_{0<d(y,y_0)\leq \varepsilon} \) in Eq. (10.8) gives
It is now easy to see that \( \lim_{\varepsilon \downarrow 0} \sup_{0 < d(y, y_0) \leq \varepsilon} \left\| x - \sum_{n=1}^{N} f_n (y) \right\| \leq \sum_{n=1}^{N} \sup_{0 < d(y, y_0) \leq \varepsilon} \left\| [x_n - f_n (y)] \right\| + 2 \sum_{n=N+1}^{\infty} M_n. \)

Letting \( \varepsilon \downarrow 0 \) (notice the functions involved are decreasing in \( \varepsilon \) and so the limits exist) we learn,

\[
\lim_{\varepsilon \downarrow 0} \sup_{0 < d(y, y_0) \leq \varepsilon} \left\| x - \sum_{n=1}^{N} f_n (y) \right\| \leq \sum_{n=1}^{N} \lim_{\varepsilon \downarrow 0} \sup_{0 < d(y, y_0) \leq \varepsilon} \left\| [x_n - f_n (y)] \right\| + 2 \sum_{n=N+1}^{\infty} M_n
\]

\[
= 2 \sum_{n=N+1}^{\infty} M_n \to 0 \text{ as } N \to \infty.
\]

It is now easy to see that \( \lim_{\varepsilon \downarrow 0} \sup_{0 < d(y, y_0) \leq \varepsilon} \left\| x - \sum_{n=1}^{\infty} f_n (y) \right\| = 0 \) implies \( \lim_{y \to y_0} \left\| x - \sum_{n=1}^{\infty} f_n (y) \right\| = 0. \)

**Alternative ending #2:** for those that don’t want to use supremums. Let \( \varepsilon > 0 \) be given and then choose \( N = N (\varepsilon) \) large (but now fixed) so that

\[
2 \sum_{n=N+1}^{\infty} M_n \leq \frac{\varepsilon}{2}.
\]

Then choose \( \delta = \delta (\varepsilon, N (\varepsilon)) > 0 \) (so \( \delta \) only depends on \( \varepsilon \)) so small that

\[
\left\| [x_n - f_n (y)] \right\| \leq \frac{\varepsilon}{2N (\varepsilon)} \text{ when } 0 < d(y, y_0) \leq \delta.
\]

Using these estimates in Eq. Eq. (10.8) shows, gives

\[
\left\| x - \sum_{n=1}^{N} f_n (y) \right\| \leq \sum_{n=1}^{N} \frac{\varepsilon}{2N} + 2 \cdot \frac{\varepsilon}{2} = \varepsilon \text{ when } 0 < d(y, y_0) \leq \delta.
\]

Thus by the definition of the limit we have shown \( \lim_{y \to y_0} \sum_{n=1}^{\infty} f_n (y) = x. \)

**Theorem 10.29 (Differentiating past a sum).** Suppose that \( f_n : J := (a, b) \to X \) is a sequence of functions satisfying:

1. there exists \( c \in J \) such that \( F (c) := \sum_{n=1}^{\infty} f_n (c) \) is convergent in \( X \).
2. For all \( n \in \mathbb{N} \), \( f_n \) is differentiable on \( J \) and \( \hat{f}_n \) is continuous on \( J \).
3. If \( M_n := \sup_{t \in J} \left\| \hat{f}_n (t) \right\|_X \), then \( \sum_{n=1}^{\infty} M_n < \infty. \)

Then \( F (t) := \sum_{n=1}^{\infty} f_n (t) \) is convergent for \( t \in J \), \( F \) is differentiable on \( J \), \( \hat{F} \) is continuous on \( J \), and

\[
\hat{F} (t) = \sum_{n=1}^{\infty} \hat{f}_n (t) \text{ for all } t \in J. \quad (10.9)
\]

**Proof.** For this proof we will need the mean value inequality (Theorem 10.19 or Proposition 12.12 below) for \( X \) – valued functions. As of yet we have only proved this when \( X = \mathbb{R}^n \) but we will see later that it holds more generally.

By the mean value inequality (Theorem 10.19 or Proposition 12.12 below) for any \( s, t \in J, \)

\[
\left\| f_n (t) - f_n (s) \right\| \leq M_n \left| t - s \right|
\]

and hence,

\[
\sum_{n=1}^{\infty} \left\| f_n (t) - f_n (s) \right\| \leq \left| t - s \right| \sum_{n=1}^{\infty} M_n < \infty
\]

from which it follows \( \sum_{n=1}^{\infty} \left[ f_n (t) - f_n (s) \right] \) is absolutely convergent. Therefore,

\[
F (t) = \sum_{n=1}^{\infty} f_n (t) = \sum_{n=1}^{\infty} \left( f_n (c) + [f_n (t) - f_n (c)] \right)
\]

is convergent as well. Furthermore the assumptions imply \( G (t) := \sum_{n=1}^{\infty} \hat{f}_n (t) \) is absolutely convergent and that \( G \) is a continuous function by the Weierstrass M - test, Theorem 7.11 Since for all \( n \in \mathbb{N} \) and \( h \neq 0 \) but small,

\[
\left\| f_n (t + h) - f_n (t) \right\| \leq M_n,
\]

it follows from Theorem 10.28 that

\[
\lim_{h \to 0} \frac{F (t + h) - F (t)}{h} = \lim_{h \to 0} \sum_{n=1}^{\infty} \frac{f_n (t + h) - f_n (t)}{h}
\]

\[
= \sum_{n=1}^{\infty} \lim_{h \to 0} \frac{f_n (t + h) - f_n (t)}{h}
\]

\[
= \sum_{n=1}^{\infty} \hat{f}_n (t) = G (t)
\]

which completes the proof of Eq. (10.9).

**Alternative computation of \( \hat{F} \).** Let \( F (t) = \sum_{n=1}^{\infty} f_n (t) \) as above. For \( t \in J \) and \( \varepsilon > 0 \) small so that \( (t - \varepsilon, t + \varepsilon) \subset J \), let

\[
g_n (h) = \begin{cases} 
\frac{f_n (t+h) - f_n (t)}{h} & \text{if } h \neq 0 \\
\hat{f}_n (t) & \text{if } h = 0.
\end{cases}
\]

By the definition of the derivative, \( \lim_{h \to 0} g_n (h) = \hat{f}_n (t) = g_n (0) \) which shows \( g_n \) is continuous at 0 and hence \( g_n \) is continuous for \( h \in (-\varepsilon, \varepsilon) \). By the mean value inequality (Theorem 10.19 or Proposition 12.12 below), there exists \( \varepsilon \) between \( t \) and \( t + h \) such that
order to conclude Eq. (10.10) holds for $|x - x_0| < r$. As $r < R$ was arbitrary, we may conclude that Eq. (10.10) holds for $|x - x_0| < R$. The computation above shows the series for $f''(x)$ converges of $|x - x_0| < R$. Thus these same results apply to $f'(x)$ and then $f''(x)$, etc. etc. So the result now follows by a straightforward induction argument.

Recall for $a \in \mathbb{N}$ that the binomial theorem states,

$$
(1 + x)^a = \sum_{k=0}^{a} \binom{a}{k} x^k
$$

(10.11)

where $\binom{a}{k} := \frac{a!}{k! (a-k)!} = \frac{a (a-1) \ldots (a-k+1)}{k!}$. We are going to generalize this theorem to all $a \in \mathbb{R}$, see Corollary 10.33 and Exercise 10.17 below for the general case.

**Definition 10.32 (Generalized Binomial Coefficients).** For $a \in \mathbb{R} \setminus \{0\}$ let

$$
\binom{a}{k} := \begin{cases} 
\frac{a(a-1) \ldots (a-k+1)}{k!} & \text{if } k \in \mathbb{N} \\
\frac{1}{k!} & \text{if } k = 0
\end{cases}
$$

i.e.

$$
\binom{a}{0} = 1, \quad \binom{a}{1} = a, \quad \binom{a}{2} = \frac{a^2 - a}{2}, \quad \binom{a}{3} = \frac{a^3 - a - 2}{3}, \ldots,
$$

$$
\binom{a}{k} = \prod_{l=1}^{k} \frac{a-l+1}{l} = \frac{a a-1 a-2 \ldots a-k+1}{1 2 3 \ldots k}.
$$

(10.12)

Notice that if $a \in \mathbb{N}$, we will have $\binom{a}{k} = 0$ if $k > a$ and for $0 \leq k \leq a$, so that Eq. (10.11) may be written as

$$
(1 + x)^a = \sum_{k=0}^{a} \binom{a}{k} x^k
$$

for $x \in \mathbb{R}$ and $a \in \mathbb{N}$.

It is also often useful to observe that

$$
\binom{a}{k+1} = \frac{a-k}{k+1} \binom{a}{k},
$$

(10.13)

and
\[-\binom{a+1}{k} = \frac{a+1}{k} \frac{a+1}{k-1} \cdots \frac{a+k-2}{k-1} \]
\[-= \frac{a+1}{k} \frac{a}{k-1} \cdots \frac{a-k+2}{k-1} \]
\[-= \frac{a+1}{k} \binom{a}{k-1}. \quad (10.14) \]

Letting \(a \to a-1\) and \(k \to k+1\) in this last identity gives the identity,
\[-\binom{a}{k+1} = \frac{a}{k+1} \binom{a-1}{k}. \quad (10.15) \]

**Corollary 10.33.** For \(a \in \mathbb{N}\),

\[(1+x)^{a} = \sum_{k=0}^{\infty} \binom{a}{k} x^{k} \text{ for } |x| < 1. \quad (10.16) \]

**Proof.** Let \(n = -a \in \mathbb{N}\). We are going to arrive at this result by differentiating the geometric series,
\[-\frac{1}{1-x} = \sum_{k=0}^{\infty} x^{k} \text{ for } |x| < 1 \]
repeatedly. For example,
\[-1 \cdot \left(\frac{1}{1-x}\right)^{2} = \frac{d}{dx} \frac{1}{1-x} \]
\[-= \sum_{k=1}^{\infty} kx^{k-1} = \sum_{k=0}^{\infty} (k+1) x^{k} \]
and then similarly
\[-2! \cdot \left(\frac{1}{1-x}\right)^{3} = \sum_{k=1}^{\infty} (k+1) kx^{k-1} = \sum_{k=0}^{\infty} (k+2) (k+1) x^{k}, \]
\[-\vdots \]
\[-(n-1)! \cdot \left(\frac{1}{1-x}\right)^{n} = \sum_{k=0}^{\infty} (k+n-1) \ldots (k+2) (k+1) x^{k}. \]
So replacing \(x\) by \(-x\) in the above expression implies
\[-(1+x)^{a} = \sum_{k=0}^{\infty} \frac{(k+n-1) \ldots (k+2) (k+1)}{(n-1)!} (-1)^{k} x^{k} \text{ for } |x| < 1. \]

To complete the proof we observe that the coefficient of \(x^{k}\) may be expressed as,
\[-(-1)^{k} \frac{(n+k-1)!}{k! \cdot (n-1)!} = \frac{(n+k-1) \cdot (n+k-2) \ldots n}{k!} \]
\[-= \frac{(a-k+1) \cdot (a-k+2) \ldots a}{k!} \]
\[-= \prod_{l=1}^{k} \frac{a-l+1}{l} = \binom{a}{k}. \]

**Alternatively,** we can prove Eq. (10.16) by induction;
\[-a(1+x)^{a-1} = \frac{d}{dx} (1+x)^{a} = \frac{d}{dx} \sum_{k=0}^{\infty} \binom{a}{k} x^{k} \]
\[-= \sum_{k=1}^{\infty} \binom{a}{k} kx^{k-1} = \sum_{k=0}^{\infty} \binom{a}{k+1} (k+1) x^{k} \]
\[-= \sum_{k=0}^{\infty} \frac{a}{k+1} \binom{a-1}{k} (k+1) x^{k} \]
\[-= a \sum_{k=0}^{\infty} \binom{a-1}{k} x^{k}, \]
where we have made us of Eq. (10.15). This completes the induction step. \[\blacksquare\]

**Exercise 10.6.** For \(z \in \mathbb{C}\) recall we have defined
\[-e^{z} := \sum_{n=0}^{\infty} \frac{z^{n}}{n!}. \]

The goal here is to give another proof of Proposition 7.23 using the following outline. Show;
1. \(\frac{d}{dz} e^{iz} = e^{iz} \) with \(e^{0z} = 1\).
2. Use this results and the obvious product rule to show
\[-\frac{d}{dt} \left[ e^{iz} e^{tw} e^{-t(z+w)} \right] = 0 \text{ for all } z, w \in \mathbb{C} \text{ and } t \in \mathbb{R}. \quad (10.17) \]
3. Use Eq. (10.17) to show \(e^{z} e^{w} e^{-(z+w)} = 1 \) for all \(z, w \in \mathbb{C}\).
4. Now, using item 3., show \(e^{-z} = (e^{z})^{-1}\) and then that \(e^{z} e^{w} = e^{z+w}\) for all \(z, w \in \mathbb{C}\).
Corollary 10.34. For \( x \in \mathbb{R} \), the functions
\[
e^x := \sum_{n=0}^{\infty} \frac{x^n}{n!}, \quad \sin x := \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} \quad \text{and} \quad \cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}
\]
are infinitely differentiable and they satisfy
\[
\frac{d}{dx} e^x = e^x \quad \text{with} \quad e^0 = 1
\]
\[
\frac{d}{dx} \sin x = \cos x \quad \text{with} \quad \sin(0) = 0
\]
\[
\frac{d}{dx} \cos x = -\sin x \quad \text{with} \quad \cos(0) = 1.
\]

Proof. Uses Theorem 10.31 to justify differentiating the power series expansions term by term to arrive at the desired results. Alternatively, take \( z = 1 \) in Exercise 10.6 in order to show \( \frac{d}{dx} e^x = e^x \). Then take \( z = i \) in Exercise 10.6 to learn that
\[
\frac{d}{dx} e^{ix} = ie^{ix}.
\]

Exercise 10.7. Show;
1. \( \cos^2 y + \sin^2 y = 1 \) for all \( y \in \mathbb{R} \). Hint: compute \( \frac{d}{dy} (\cos^2 y + \sin^2 y) = 0 \).
2. Conclude that \( |e^y| = 1 \) for all \( y \in \mathbb{R} \).
3. Show \( |e^z| = e^{\Re z} \) for all \( z \in \mathbb{C} \).

Remark 10.35 (cos and sin are the functions you think they are). Please see page 182-184 of Rudin where he shows;
1. There is a number called \( \pi/2 \) which is the first positive zero of \( \cos(\theta) \).
2. The functions \( e^{i\theta}, \cos(\theta), \) and \( \sin(\theta) \) are \( 2\pi \) - periodic. Here we say a function, \( f : \mathbb{R} \rightarrow \mathbb{C} \) is \( 2\pi \) - periodic iff \( f(\theta + 2\pi) = f(\theta) \) for all \( \theta \in \mathbb{R} \).
3. \( e^{i\theta} \neq 1 \) for \( 0 < \theta < 2\pi \) and hence \( [0, 2\pi) \ni \theta \rightarrow e^{i\theta} \in \mathbb{C} \) is one to one.
4. For all \( z \in \mathbb{C} \) with \( |z| = 1 \), there exists a unique \( \theta \in [0, 2\pi) \) such that \( e^{i\theta} = z \).
5. The arc length of the curve \( [0, \alpha) \ni \theta \rightarrow e^{i\theta} \) is \( \alpha \) for all \( \alpha \geq 0 \). [We have to wait until we do integration theory in order to define arc-length.] In particular, \( 2\pi \) is the arc-length of the unit circle in \( \mathbb{C} \).

Exercise 10.8 (Theory behind separation of variables). 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. (10.18) by \( ix \) and comparing the result to Eqs. (10.19) and (10.20).
1. \( F(t,x) \) is continuous for \( (t,x) \in [0, \infty) \times \mathbb{R} \). Hint: Weierstrass M - test, Theorem 7.11
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: Use Theorem 10.29 to compute these derivatives. In computing the \( t \) derivative, you should let \( \varepsilon > 0 \) and apply Theorem 10.29 with \( t > \varepsilon \) and then afterwards let \( \varepsilon \downarrow 0 \).
3. \( F \) satisfies the heat equation, namely
\[
\partial F(t,x)/\partial t = \partial^2 F(t,x)/\partial x^2 \quad \text{for} \quad t > 0 \quad \text{and} \quad x \in \mathbb{R}.
\]

10.5 Taylor’s Theorem

Remark 10.36. If \( f(x) := \sum_{n=0}^{\infty} a_n x^n \) converges of \( x \in (-R,R) \), then by induction along with Theorem 10.31, one shows for \( |x| < R \) that,
\[
f^{(k)}(x) = \sum_{n=k}^{\infty} a_n \cdot n(n-1)(n-2)\ldots(n-k+1)x^{n-k}
\]
and in particular that
\[
f^{(k)}(0) = k! \cdot a_k.
\]
Therefore we may express \( f \) as,
\[
f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k.
\]
Exercise 10.9. Suppose \( f : \mathbb{R} \to \mathbb{R} \) is a continuous function such that \( f \) is differentiable to all orders on \( \mathbb{R} \setminus \{0\} \). Further suppose that \( \lim_{t \to 0} f^{(n)}(t) = L_n \) exists in \( \mathbb{R} \) for each \( n \in \mathbb{N} \). Show \( f \) is differentiable at \( t = 0 \) to all orders and that \( f^{(n)}(0) = L_n \).

Exercise 10.10. [In this exercise we will take for granted that \( \frac{d}{dx}e^x = e^x \). This will be proved a little later in this chapter.] Let \( f : \mathbb{R} \to \mathbb{R} \) be the function defined by

\[
 f(t) := \begin{cases} 
 0 & \text{if } t \leq 0 \\
 e^{-1/t} & \text{if } t > 0 
\end{cases}.
\]

Show \( f \) is differentiable to all orders on \( \mathbb{R} \) and that \( f^{(n)}(0) = 0 \) for all \( n \in \mathbb{N} \).

Hints: first show there exists a polynomial functions, \( p_n(x) \), such that \( f^{(n)}(t) = p_n(t^{-1}) f(t) \) for \( t > 0 \). You may also use \( e^x \geq \sum_{n=0}^N \frac{x^n}{n!} \) when \( x \geq 0 \) and therefore \( \lim_{x \to \infty} \frac{x^n}{n!} = 0 \) for all \( n \in \mathbb{N} \).

Remark 10.37. Exercise 10.10 shows that not all infinitely differentiable functions may be represented as a power series! Nevertheless we can try to approximate a function \( f : (a, b) \to X \) near \( x_0 \in (a, b) \) by its “Taylor Polynomial of degree \( n \)” defined by

\[
p_n(x) := \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k. \tag{10.22}
\]

It is easy to check that \( p_n \) is the unique degree \( n \) polynomial such that \( \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!}(t-x_0)^k \) for \( 0 \leq k \leq n \) with the convention that \( f^{(0)} = f \). Indeed if

\[
p(x) = \sum_{k=0}^n p_k(x-x_0)^k \quad \text{where } p_k \in \mathbb{R},
\]

then a simple computation shows \( p^{(k)}(x_0) = p_k \cdot k! \). So if we want \( p^{(k)}(x_0) = f^{(k)}(x_0) \) for \( 0 \leq k \leq n \) we must require \( p_k := \frac{f^{(k)}(x_0)}{k!} \). The question now becomes what is the error between \( f \) and \( p_n \).

Theorem 10.38 (Taylor’s Theorem). Suppose that \( f : \mathbb{R} \to \mathbb{R} \) is \( n+1 \) times differentiable on \((a, b)\) and \( x_0 \in (a, b) \). Let \( p_n(x) \) be as in Eq. \(10.22\). Then for all \( x \in (a, b) \) there exists a real number \( c \) between \( x_0 \) and \( x \) such that

\[
f(x) = p_n(x) + \frac{f^{(n+1)}(c)}{(n+1)!}(x-x_0)^{n+1}. \tag{10.23}
\]

Proof. Let me first give the proof in the special case that \( n = 2 \). Given \( x \in (a, b) \) with \( x \neq x_0 \) and let \(1 \) fix \( M = M_x \) so that

\[
 f(x) = p_2(x) + M (x-x_0)^3.
\]

Our goal is to show, for some \( c \) between \( x_0 \) and \( x \), that

\[
 M = \frac{f^{(3)}(c)}{3!}.
\]

To prove this we let

\[
h(t) := f(t) - p_2(t) - M (t-t_0)^3
\]

and observe that \( h(x) = 0 \) by construction of \( M \) and that \( h^{(k)}(x_0) = 0 \) for \( 0 \leq k \leq 2 \) and

\[
h^3(t) = f^3(t) - M \cdot 3!.
\]

So to finish the proof it suffices to find a \( c \) between \( x_0 \) and \( x \) such that \( h^3(c) = 0 \). This means we take \( c = c_3 \) which is between \( x_0 \) and \( x \).

For the general case, let \( x \in (a, b) \) with \( x \neq x_0 \) and define \( M = M_x \) so that

\[
f(x) = p_n(x) + \frac{f^{(n+1)}(c)}{(n+1)!}(x-x_0)^{n+1}
\]

and set

\[
h(t) := f(t) - p_n(t) - M (t-t_0)^{n+1}.
\]

By construction; \( h(x) = 0 \), \( h^{(k)}(x_0) = 0 \) for \( 0 \leq k \leq n \), and

\[
h^{(n+1)}(t) = f^{(n+1)}(t) - M \cdot (n+1)!.
\]

For sake of definiteness, suppose that \( x > x_0 \). Since \( h(x) - h(x_0) = 0 - 0 = 0 \), then by the Mean value theorem (or use Rolle’s theorem directly) there exists \( c_1 \) between \( x_0 \) and \( x \) so that \( h’(c_1) = 0 \). Similarly since \( h’(c_1) - h’(x_0) = 0 - 0 = 0 \), there exists \( c_2 \) between \( x_0 \) and \( c_1 \) such that \( h''(c_2) = 0 \). By repeating this procedure we find

\[
x_0 < c_{n+1} < c_n < \cdots < c_1 < x
\]

so that \( h^{(k)}(c_k) = 0 \) for \( 1 \leq k \leq n + 1 \). Setting \( c = c_{n+1} \), we then have

\[
0 = h^{(n+1)}(c) = f^{(n+1)}(c) - M \cdot (n+1)!
\]
which implies,

\[ M = \frac{f^{(n+1)}(c)}{(n+1)!} \]

Using this back in Eq. (10.24) gives Eq. (10.23).

**Example 10.39 (Exponential Function revisited).** Suppose \( f : \mathbb{R} \to \mathbb{R} \), such that \( f'(x) = f(x) \) for all \( x \in \mathbb{R} \) and \( f(0) = 1 \). Then by Taylor’s theorem with \( \alpha = 0 \), for all \( x \in \mathbb{R} \) and \( n \in \mathbb{N}_0 \) there exists a point \( c = c_n(x) \) between 0 and \( x \) such that

\[ f(x) = \sum_{k=0}^{n} \frac{1}{k!} x^k + \frac{x^{n+1}}{(n+1)!} f(c_n(x)) \]

wherein we have used \( f^{(n)} = f \) for all \( n \). Letting \( M_x := \max \{|f(s)| : |s| \leq |x|\} \) we may conclude that

\[ \left| f(x) - \sum_{k=0}^{n} \frac{1}{k!} x^k \right| = \frac{x^{n+1}}{(n+1)!} |f(c_n(x))| \leq \frac{x^{n+1}}{(n+1)!} M_x \to 0 \text{ as } n \to \infty. \]

Thus there is at most one such function and if it exists it must be given by

\[ f(x) = \sum_{k=0}^{\infty} \frac{1}{k!} x^k, \]

i.e. the function we have defined previously to be \( \exp(x) = e^x \). We have seen in Corollary 10.34 that \( \exp(x) \) is differentiable and the \( \frac{d}{dx} \exp(x) = \exp(x) \) with \( \exp(0) = 1 \) as required.

**Exercise 10.11 (Taylor’s Theorem II).** Suppose that \( f : \mathbb{R} \to \mathbb{R}^d \) is \( n+1 \) times differentiable on \((a, b)\) and \( x_0 \in (a, b) \) Let

\[ p_n(x) := \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!} (x-x_0)^k \]

as in Eq. (10.22). Then for all \( x \in (a, b) \) there exists a \( c \) between \( x_0 \) and \( x \) such that

\[ \|f(x) - p_n(x)\| \leq \|f^{(n+1)}(c)\| \frac{|x-x_0|^{n+1}}{(n+1)!}. \]

**Hint:** use the idea in the proof of Theorem 10.19 to reduce this to the case where \( d = 1 \).

### 10.6 Inverse Function Theorem I

**Theorem 10.40 (Converse to the Chain Rule).** Let \( U \) and \( V \) be open subset of \( \mathbb{R} \) and suppose \( f : U \to V \) and \( g : V \to \mathbb{R} \) are continuous functions such that \( g \) is differentiable on \( V \) and \( h := g \circ f \) is differentiable on \( U \). Then \( f \) is differentiable at all point \( x \in U \) where \( g'(f(x)) \neq 0 \) and at such a point,

\[ f'(x) = g'(f(x))^{-1} h'(x). \quad (10.25) \]

**Proof.** Suppose that \( x \in U \) and \( g'(f(x)) \neq 0 \). Let \( \Delta f = f(x + \Delta x) - f(x) \) and notice that \( \Delta f = \varepsilon (\Delta x) \) because \( f \) is continuous at \( x \). Then on one hand,

\[ \Delta h := f(x + \Delta f) - f(x) = g'(f(x)) \Delta f + o(\Delta f) \]

while on the other hand,

\[ \Delta h = h(x + \Delta x) - h(x) = h'(x) \Delta x + o(\Delta x). \]

Comparing these equations implies,

\[ h'(x) \Delta x + o(\Delta x) = g'(f(x)) \Delta f + o(\Delta f) \]

or equivalently,

\[ \Delta f = g'(f(x))^{-1} [h'(x) \Delta x + o(\Delta x) - o(\Delta f)] = g'(f(x))^{-1} h'(x) \Delta x + o(\Delta x) + o(\Delta f). \quad (10.26) \]

Since \( f \) is continuous, \( \Delta f \to 0 \) as \( \Delta x \to 0 \). Therefore there exists \( \delta > 0 \) such that \( |o(\Delta f)| \leq \frac{1}{2} |\Delta f| \) if \( |\Delta x| \leq \delta \). Using this in Eq. (10.26) implies, for some \( C < \infty \) and \( |\Delta x| \leq \delta \) that

\[ |\Delta f| = \left| g'(f(x))^{-1} h'(x) \Delta x + o(\Delta x) + o(\Delta f) \right| \leq \left| g'(f(x))^{-1} h'(x) \Delta x \right| + o(\Delta x) + \frac{1}{2} |\Delta f| \leq C |\Delta x| + \frac{1}{2} |\Delta f|, \]

i.e. \( |\Delta f| \leq 2C |\Delta x| \) for \( |\Delta x| \leq \delta \). From this we may now conclude that \( o(\Delta f) = o(\Delta x) \) and therefore Eq. (10.26) asserts,

\[ f(x + \Delta x) - f(x) = g'(f(x))^{-1} h'(x) \Delta x + o(\Delta x). \]

i.e. \( f'(x) \) exists and is give by Eq. (10.25).
Theorem 10.41 (Inverse Function Theorem). Suppose that \( g : (a, b) \to \mathbb{R} \) is a differentiable function such that \( g'(x) > 0 \) (or \( g'(x) < 0 \)) for all \( a < x < b \). Let \( \alpha = \inf \{ g(x) : a < x < b \} \) and \( \beta := \sup \{ g(x) : a < x < b \} \); then \( g : (a, b) \to (\alpha, \beta) \) is a bijection and there exists a unique inverse function, \( f : (\alpha, \beta) \to (a, b) \) such that \( g(f(y)) = y \) for all \( y \in (\alpha, \beta) \) and \( f(g(x)) = x \) for all \( x \in (a, b) \). Moreover the function \( f \) is differentiable on \( (\alpha, \beta) \) and

\[
    f'(y) = \frac{1}{g'(f(y))} \quad \text{for all } y \in (\alpha, \beta).
\]

Proof. I will assume that \( g' > 0 \), if not apply the results to \( -g \) instead. It follows by the mean value theorem that \( g(x) < g(x') \) if \( x < x' \), i.e. \( g \) is strictly increasing. Since \( g \) is continuous and \( (a, b) \) is connected we know that \( J := g(a, b) \) is connected and hence equal to an interval. If \( \beta = \sup J \in J \), then there would exists \( x \in (a, b) \) such that \( g(x) = \beta \) and then for \( x' \in (a, b) \) we would have \( g(x') \in J \) while \( g(x') > \beta \) which violates the fact that \( \beta \) is an upper bound for \( J \). Similarly we see that \( \alpha \notin J \) and therefore \( J = (\alpha, \beta) \). Hence we have shown \( g : (a, b) \to (\alpha, \beta) \) is a bijection and hence there exists a unique inverse map, \( f : (\alpha, \beta) \to (a, b) \) and this map is also strictly increasing.

The continuity of \( f \) has already been addressed in Exercise 6.20 and Exercise 9.40 Here is yet another variant of the continuity proof. Suppose \( y \in (\alpha, \beta) \), \( \alpha < a < y < b < \beta \) and \( y_n \in [a, b] \) is a sequence such that \( \lim_{n \to \infty} y_n = y \). Then \( x_n := f(y_n) \in [f(a), f(b)] \) and we wish to show \( \lim_{n \to \infty} x_n = f(y) \). For sake of contradiction suppose not, then there exists \( \varepsilon > 0 \) and as subsequence, \( \tilde{x}_k := x_{n_k} \) such that \( |\tilde{x}_k - f(y)| \geq \varepsilon \). By passing to a subsequence if necessary, using the sequential compactness of \([f(a), f(b)]\), we may assume that \( \lim_{k \to \infty} \tilde{x}_k = x_0 \) exists. Since \( g \) is continuous it follows that

\[
    g(x_0) = \lim_{k \to \infty} g(\tilde{x}_k) = \lim_{k \to \infty} g(f(y_{n_k})) = \lim_{k \to \infty} y_{n_k} = y
\]

which shows \( x_0 = f(y) \). But this then leads to the contradiction;

\[
    0 < \varepsilon \leq \lim_{k \to \infty} |\tilde{x}_k - f(y)| = |x_0 - f(y)| = 0.
\]

The differentiability assertions now follows from Theorem 10.41.

Since the function \( g(x) = e^x \) satisfies \( g'(x) > 0 \) for all \( x \in \mathbb{R} \), \( \lim_{x \to -\infty} e^x = 0 \), and \( \lim_{x \to \infty} e^x = \infty \) there is a unique function, \( \ln : (0, \infty) \to \mathbb{R} \) such that \( e^\ln y = y \) for all \( y > 0 \) and \( \ln (e^x) = x \) and \( x \in \mathbb{R} \). We refer to this function as the natural logarithm.

Theorem 10.42 (Natural logarithm). Let \( \ln : (0, \infty) \to \mathbb{R} \) be the inverse function to \( \exp(x) = e^x \). Then;

1. \( \ln \) is an increasing differentiable function such that \( \ln 1 = 0 \) and

\[
    \frac{d}{dy} \ln y = \frac{1}{y} \quad \text{for all } y > 0.
\]

2. For \( a, b \in (0, \infty) \), \( \ln(ab) = \ln a + \ln b \).

3. For \( a > 0 \), \( \ln a^{-1} = -\ln a \).

4. For any \( k \in \mathbb{Z} \) and \( a > 0 \), \( \ln a^k = k \ln a \) or equivalently \( a^k = e^{k \ln a} \).

5. For any \( q \in \mathbb{Q} \) and \( a > 0 \), \( \ln a^q = q \ln a \) or equivalently \( a^q = e^{q \ln a} \).

Proof. 1. The first assertion follows directly from Theorem 10.41 noting that

\[
    1 = \frac{d}{dy} y = \frac{d}{dy} e^{\ln y} = e^{\ln y} \frac{d}{dy} \ln y = y \ln y
\]

and \( e^0 = 1 \) so that \( \ln 1 = 0 \).

2. If \( x = \ln a \) and \( y = \ln b \), then \( x = e^x \) and \( b = e^y \) so that \( ab = e^x e^y = e^{x+y} \) and hence \( \ln(ab) = x + y = \ln a + \ln b \).

3. Take \( b = a^{-1} \) in item 2. to conclude that \( 0 = \ln 1 = \ln (a \cdot a^{-1}) = \ln a + \ln a^{-1} \), i.e. \( \ln a^{-1} = -\ln a \).

4. If \( k = 0 \) and \( a > 0 \) then \( a^0 = 1 \) so that \( \ln a^0 = \ln 1 = 0 = 0 \cdot \ln a \). For \( k \in \mathbb{N} \) we show using item 1. and induction that \( \ln a^k = k \ln a \). If \( k \in \mathbb{N} \) then

\[
    \ln a^k = \ln (a^k)^{1} = -\ln |a|^k = -|k| \ln a = \ln k \ln a.
\]

5. If \( b = \frac{1}{n} \), then \( e^{\frac{1}{n} \ln a} > 0 \) and

\[
    \left( e^{\frac{1}{n} \ln a} \right)^n = e^{\frac{1}{n} \ln a} = e^{\ln a} = a.
\]

As we know there is a unique positive \( n^{th} \) root of \( a \) (we have already seen this but it also follows from Theorem 10.41) and therefore, \( e^{\frac{1}{n} \ln a} = a^{1/n} \). So if \( q = m/n \) with \( m \in \mathbb{Z} \) and \( n \in \mathbb{N} \) we find,

\[
    a^q = (a^{1/n})^m = \left( e^{\frac{1}{n} \ln a} \right)^m = e^{\frac{m}{n} \ln a} = e^{\ln a}.
\]

Definition 10.43. For \( a > 0 \) and \( b \in \mathbb{R} \) we define \( a^b := e^{b \ln a} \). [This agrees with our previous definition of \( a^q \) when \( b \in \mathbb{Q} \) by item 5. of Theorem 10.42]

Corollary 10.44. For \( a > 0 \) and \( b \in \mathbb{R} \),

\[
    \frac{d}{dx} a^x = \ln a \cdot a^x \quad \text{for all } x \in \mathbb{R}
\]

and

\[
    \frac{d}{dx} x^b = bx^{b-1} \quad \text{for all } x > 0.
\]
Exercise 10.12. Prove Corollary \[\text{[10.44]}\] i.e. for \(a > 0\) and \(b \in \mathbb{R}\),
\[
\frac{d}{dx} a^x = \ln a \cdot a^x \quad \text{for all } x \in \mathbb{R}
\]
and
\[
\frac{d}{dx} x^b = bx^{b-1} \quad \text{for all } x > 0.
\]

Exercise 10.13. Suppose that \(f : \mathbb{R} \to \mathbb{R}\) is a continuous function such that \(f(0) \neq 0\) and
\[
f(x + y) = f(x)f(y) \quad \text{for all } x, y \in \mathbb{R}.
\]
(10.27)
Show:
1. \(f(0) = 1\) and \(f(x) > 0\) for all \(x \in \mathbb{R}\).
2. Under the further assumption that \(f\) is differentiable, show \(f(x) = e^{cx}\) for some \(c \in \mathbb{R}\).
3. Show the result in 2. holds even if we only assume that \(f\) is continuous.

Exercise 10.14. Let \(\mathcal{P} = \{2, 3, 5, 7, 11, \ldots \} \subset \mathbb{N}\) be the prime numbers and for \(N \in \mathbb{N}\) let \(\mathcal{P}_N = \mathcal{P} \cap \{1, 2, \ldots, N\}\). Show:
1. Fix \(N \in \mathbb{N}\) and let \(\mathcal{P}_N = \{p_1, \ldots, p_k\}\). Then for each \(1 \leq n \leq N\) (by the prime factorization of \(n\)) there exists a unique vector \(a(n) = (a_1(n), \ldots, a_k(n)) \in \mathbb{N}_0^k\) such that \(n = p_1^a_1(n) \cdots p_k^a_k(n)\) and hence
\[
\sum_{n=1}^{N} \frac{1}{n} = \sum_{n=1}^{N} \frac{1}{a_1(n)} \cdots \frac{1}{a_k(n)}.
\]
(10.28)
Use this observation to show,
\[
\sum_{n=1}^{N} \frac{1}{n} \leq \prod_{j=1}^{k} \sum_{n=0}^{N} \frac{1}{p_j} \leq \prod_{j=1}^{k} \frac{1}{1 - \frac{1}{p_j}} = \prod_{p \in \mathcal{P}_N} \left(1 - \frac{1}{p}\right)^{-1}.
\]
(10.29)
2. Show \((1 - x)^{-1} \leq e^{2x}\) for all \(0 \leq x \leq \frac{1}{2}\).
3. Combining items 1. and 2. together to show,
\[
\sum_{n=1}^{N} \frac{1}{n} \leq \prod_{p \in \mathcal{P}_N} e^{2 \frac{1}{p}} = \exp \left(2 \cdot \sum_{p \in \mathcal{P}_N} \frac{1}{p}\right).
\]
Letting \(N \to \infty\) then shows \(\sum_{p \in \mathcal{P}} \frac{1}{p} = \infty\) which implies the primes are not sparse in \(\mathbb{N}\) and in particular implies \(#(\mathcal{P}) = \infty\).

Example 10.45. In a similar fashion one may use Theorem \[\text{[10.41]}\] to prove the usual differentiation formulas for the inverse trigonometric functions:
\[
\frac{d}{dx} \arctan (x) = \frac{1}{x^2 + 1},
\]
\[
\frac{d}{dx} \arcsin (x) = \frac{1}{\sqrt{1 - x^2}},
\]
\[
\frac{d}{dx} \arccos (x) = -\frac{1}{\sqrt{1 - x^2}}, \quad \text{and}
\]
\[
\frac{d}{dx} \arcsec (x) = \frac{1}{|x| \sqrt{x^2 - 1}}
\]
(where they are defined).
1. For example,
\[
\frac{d}{dy} \tan (y) = \frac{d}{dy} \sin y \cos y = \cos^2 y + \sin^2 y = \frac{1}{\cos^2 y}
\]
which is positive on \((-\pi/2, \pi/2)\). Moreover we have \(\lim_{y \to \pi/2} \tan y = \infty\) and \(\lim_{y \to -\pi/2} \tan y = -\infty\). Thus by the inverse function theorem asserts there exist a increasing differentiable function, \(f : \mathbb{R} \to (-\pi/2, \pi/2)\) such that \(\tan (f(x)) = x\) for all \(x \in \mathbb{R}\). Differentiating this equation then implies
\[
1 = \frac{1}{\cos^2 (f(x))} f'(x), \quad \text{i.e. } f'(x) = \cos^2 (f(x)).
\]
Letting \(y = f(x)\), we have
\[
x^2 = \tan^2 y = \frac{\sin^2 y}{\cos^2 y} \implies 1 - \cos^2 y = x^2 \cos^2 y \implies \cos^2 y = \frac{1}{1 + x^2}.
\]
Therefore it follows that \(f'(x) = \frac{1}{1 + x^2}\). The functions \(f(x)\) is more commonly known as \(\arctan (x)\).
2. As another example, let us observe that \(\frac{d}{dx} \sin (x) = \cos (x) > 0\) on \((-\pi/2, \pi/2)\) and in this case \(\lim_{y \to \pi/2} \sin y = 1\) and \(\lim_{y \to -\pi/2} \sin y = -1\). Thus the inverse function theorem asserts there exists a differentiable increasing function, \(f : (-1,1) \to (-\pi/2, \pi/2)\) such that \(\sin (f(x)) = x\). Of course \(f(x) = \arcsin (x)\). Differentiating this equation then implies
\[
1 = \cos (f(x)) f'(x), \quad \text{i.e. } f'(x) = \frac{1}{\cos (f(x))}.
\]
Letting \(y = f(x)\), we have \(x = \sin y\) and so \(\cos y = \pm \sqrt{1 - x^2}\). We know we must take the + sign here since \(\cos y > 0\) for \(y \in (-\pi/2, \pi/2)\). Thus we have shown,
Exercise 10.17 (Binomial Expansions). Let \( a \in \mathbb{R} \) and \( f(x) = (1 + x)^a \). By direct computation using induction one shows, \( f(0) = 1, f'(x) = a(1 + x)^{a-1} \),

\[
\begin{align*}
\frac{f''(x)}{2!} &= a(a-1) (1 + x)^{a-2}, \\
\frac{f^{(3)}(x)}{3!} &= a(a-1)\cdots(a-2) (1 + x)^{a-2} \\
\vdots \quad \vdots \quad \vdots \\
\frac{f^{(n)}(x)}{n!} &= a(a-1)\cdots(a-n+1) (1 + x)^{a-n} \\
&= \binom{a}{n} (1 + x)^{a-n}.
\end{align*}
\]

So we expect, based on Taylor’s Theorem \[\text{Exercise 10.38} \]

\[ (1 + x)^a = \sum_{n=0}^{\infty} \binom{a}{n} x^n. \quad (10.30) \]

for \(|x| < R \) where \( R \) is the radius of convergence of the series above. Show;

1. \( R = 1. \) Hint: you might use the ratio test here.
2. Show \((1 + x)^a \) solves the differential equation,

\[ (1 + x) \frac{d}{dx} (1 + x)^a = a(1 + x)^a \quad \text{for } x > -1. \quad (10.31) \]

Exercise 10.16. Show the series

\[ f(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n \]

converges for \(|x| < 1\) and that \( f(x) = \ln(1 + x) \) for \(|x| < 1\). **Hint**: this problem can be done independently from Exercise 10.15. Your goal is to show \( f(x) = \ln(1 + x) \) for \(|x| < 1\).

Exercise 10.18. Exercise 10.16.

Use Taylor’s Theorem 10.38 to show Exercise 10.15. So we expect, based on Taylor’s Theorem 10.38, that

\[ \ln(1 + x) = 0 \quad \text{for } x = 0. \]

**Hint**: this problem can be done independently from Exercise 10.15. Your goal is to show \( g(x) := \ln(1 + x) - \ln(1 + x) = 0 \) for \(|x| < 1\).

Exercise 10.17 (Binomial Expansions). Let \( a \in \mathbb{R} \) and \( f(x) = (1 + x)^a \). By direct computation using induction one shows, \( f(0) = 1, f'(x) = a(1 + x)^{a-1} \),

\[
\begin{align*}
\frac{f''(x)}{2!} &= a(a-1) (1 + x)^{a-2}, \\
\frac{f^{(3)}(x)}{3!} &= a(a-1)\cdots(a-2) (1 + x)^{a-2} \\
\vdots \quad \vdots \quad \vdots \\
\frac{f^{(n)}(x)}{n!} &= a(a-1)\cdots(a-n+1) (1 + x)^{a-n} \\
&= \binom{a}{n} (1 + x)^{a-n}.
\end{align*}
\]

So we expect, based on Taylor’s Theorem 10.38, that

\[ (1 + x)^a = \sum_{n=0}^{\infty} \binom{a}{n} x^n. \quad (10.30) \]

for \(|x| < R \) where \( R \) is the radius of convergence of the series above. Show;

1. \( R = 1. \) Hint: you might use the ratio test here.
2. Show \((1 + x)^a \) solves the differential equation,

\[ (1 + x) \frac{d}{dx} (1 + x)^a = a(1 + x)^a \quad \text{for } x > -1. \quad (10.31) \]

3. Show Eq. \[10.30\] is valid for \(|x| < 1\) as follows.
   a) Show

\[ g(x) := \sum_{n=0}^{\infty} \binom{a}{n} x^n \quad (10.32) \]

satisfies \((1 + x) g'(x) = ag(x) \) for \(|x| < 1\), the same differential equation as \((1 + x)^a \).

b) Next compute the derivative of \( h(x) := (1 + x)^{-a} g(x) \) and use to show \( h(x) = 1 \) for \(|x| < 1\).


10.7 L’hopitals Rule

**Theorem 10.46 (Cauchy’s Generalized Mean Value Theorem).** Suppose that \( f, g : [a, b] \to \mathbb{R} \) are continuous functions which are differentiable on \((a, b)\). Then there exists \( t_0 \in (a, b) \) such that

\[ [f(b) - f(a)] g'(t_0) = [f(b) - f(a)] f'(t_0). \]

**Proof.** Let

\[ h(t) := [f(b) - f(a)] g(t) - [f(b) - f(a)] f(t). \]

Then

\[ h(b) = [f(b) - f(a)] g(b) - [f(b) - f(a)] f(b) = g(a) f(b) - f(a) g(b) \]

and

\[ h(a) = [f(b) - f(a)] g(a) - [f(b) - f(a)] f(a) = g(a) f(b) - f(a) g(b) \]

so that \( h(a) = h(b) \). Applying Rolle’s theorem shows there exists \( t_0 \in (a, b) \) such that

\[ 0 = h'(t_0) = [f(b) - f(a)] g'(t_0) - [g(b) - g(a)] f'(t_0). \]

**Theorem 10.47 (L’hopitals Rule).** Let \( c \) and \( L \) be in \( \mathbb{R} \). The real valued functions \( f \) and \( g \) are assumed to be differentiable on an open interval \( J \) with endpoint \( c \), and additionally \( g'(x) \neq 0 \) and \( g(x) \neq 0 \) for \( x \in J \). It is also assumed that

\[ \lim_{x \to c} \frac{f'(x)}{g'(x)} = L, \]

\[ \lim_{x \to c} f(x) = \lim_{x \to c} g(x) = 0. \]
where \( \lim_{x \to c} \) means \( \lim_{x \to c^+} \) if \( c \) is the right end point of \( J \) and \( \lim_{x \to c} \) if \( c \) is the left endpoint of \( J \). If either:

1) \( \lim_{x \to c^-} f(x) = \lim_{x \to c^-} g(x) = 0 \) or

2) \( \lim_{x \to c^-} |f(x)| = \lim_{x \to c^-} |g(x)| = \infty \),

then

\[
\lim_{x \to c} \frac{f(x)}{g(x)} = L.
\]

The following proof is taken fairly directly from Wikipedia. For definiteness we assume that \( J = (a, c) \) for some \( -\infty < a < c \), we allow for \( c = \infty \) here. The case where \( c \) is the left endpoint of \( J \) is proved similarly. For \( x \in J \) let

\[
m(x) = \inf_{\xi \in (x, c)} \frac{f'(\xi)}{g'(\xi)} \in \mathbb{R} \quad \text{and} \quad M(x) = \sup_{\xi \in (x, c)} \frac{f'\xi}{g'\xi} \in \mathbb{R}.
\]

By Cauchy’s mean value theorem if \( a < x < y < c \) there exist \( \xi \in (x, y) \) such that

\[
f(x) - f(y) = \frac{f'(\xi)}{g'(\xi)} (x - y).
\]

[Notice \( g(x) - g(y) \neq 0 \) by the mean value theorem and the assumption that \( g'(\xi) \neq 0 \) for \( \xi \in J \).] In particular we have

\[
m(x) \leq \frac{f(x) - f(y)}{g(x) - g(y)} \leq M(x) \quad \text{for all} \quad a < x < y < c.
\] \hfill (10.33)

**Case 1.** Suppose that \( \lim_{x \to c^-} f(x) = \lim_{x \to c^-} g(x) = 0 \). In this case we may let \( y \uparrow c \) in Eq. \( (10.33) \) in order to learn

\[
m(x) \leq \frac{f(x)}{g(x)} \leq M(x).
\] \hfill (10.34)

**Case 2.** If \( \lim_{x \to c^-} |g(x)| = \infty \), then

\[
m(x) \leq \frac{f(y) - f(x)}{g(y) - g(x)} = \frac{\frac{f(y)}{g(y)} - \frac{f(x)}{g(x)}}{1 - \frac{g(x)}{g(y)}} \leq M(x).
\] \hfill (10.35)

Since

\[
\lim_{y \uparrow c^-} \frac{f(x)}{g(x)} = 0 = \lim_{y \uparrow c^-} \frac{f(y)}{g(y)}
\]

we may pass to the limit in Eq. \( (10.35) \) in order to find,

\[
m(x) \leq \liminf_{y \uparrow c^-} \frac{f(y)}{g(y)} \leq \limsup_{y \uparrow c^-} \frac{f(y)}{g(y)} \leq M(x).
\] \hfill (10.36)

Since

\[
\lim_{x \uparrow c} \frac{f(x)}{g(x)} = \liminf_{y \uparrow c^-} \frac{f'(x)}{g'(x)} = L = \limsup_{x \uparrow c} \frac{f'(x)}{g'(x)} = \lim M(x)
\]

we may pass to the limit in Eqs. \( (10.34) \) and \( (10.36) \) using the squeeze theorem in order to learn \( L = \lim_{x \uparrow c} \frac{f(x)}{g(x)} \) in Case 1) and

\[
L \leq \liminf_{y \uparrow c^-} \frac{f(y)}{g(y)} \leq \limsup_{y \uparrow c^-} \frac{f(y)}{g(y)} \leq L
\]

in Case 2. The last equation implies \( \lim_{y \uparrow c^-} \frac{f(y)}{g(y)} \) exists and is equal to \( L \).

**Example 10.48.** By L’Hospital’s Rule,

\[
\lim_{x \to 0} \frac{\sin(x)}{x} = \lim_{x \to 0} \frac{\frac{d}{dx} \sin(x)}{\frac{d}{dx} x} = \lim_{x \to 0} \frac{\cos(x)}{1} = 1
\]

and by two applications of L’Hospital’s Rule,

\[
\lim_{x \to 0} \frac{1 - \cos(x)}{x^2} = \lim_{x \to 0} \frac{\sin(x)}{2x} = \lim_{x \to 0} \frac{\cos(x)}{2} = \frac{1}{2}
\]

Technically one needs to work backwards in these string of equalities to justify the use of L’Hospital’s Rule. Let me point out you would be wrong to use L’Hospital’s rule for the following limit; \( \lim_{x \to 0} \frac{\cos(x)}{x} = \infty \). If we were to mistakenly use L’Hospital’s rule in this \( \infty/0 \) situation we would conclude mistakenly that

\[
\infty = \lim_{x \to 0} \frac{\cos(x)}{x} = \lim_{x \to 0} \frac{\frac{d}{dx} \cos(x)}{\frac{d}{dx} x} = \lim_{x \to 0} \frac{-\sin(x)}{1} = 0.
\]

**Exercise 10.19.** Rudin problem 5.11 on page 115.

**Exercise 10.20.** Rudin problem 5.13 on page 115. In this problem assume that \( x \in [0, 1] \) rather than \( [-1, 1] \) and assume \( a \neq 0 \) and \( c > 0 \).
11

Simple Integration Theory

In this chapter we will let $\mathbb{F}$ be either $\mathbb{R}$ or $\mathbb{C}$ and $X$ and $Y$ will be Banach spaces over $\mathbb{F}$.

11.1 Banach Spaces of Bounded Functions

**Notation 11.1** Let $\Omega$ be a set and $Y$ be a Banach space and for any function $f : \Omega \to Y$ let

$$\|f\|_u := \sup_{\omega \in \Omega} \|f(\omega)\|_Y.$$  

Also let $B(\Omega, Y)$ denote those functions, $f : \Omega \to Y$ such that $\|f\|_u < \infty$. We refer to $B(\Omega, Y)$ as the space of $Y$-valued bounded functions on $\Omega$.

**Fact 11.2** The collection of all functions $f : \Omega \to Y$, denoted by $Y^\Omega$, is a $\mathbb{F}$-vector space where for $g : \Omega \to Y$ and $\lambda \in \mathbb{F}$ we define

$$(f + \lambda g)(\omega) := f(\omega) + \lambda g(\omega)$$

for all $\omega \in \Omega$.

We are going to take this fact for granted. To prove one needs to open a book on linear algebra and verify that $Y^\Omega$ satisfies all the axioms of a vector space.

**Proposition 11.3.** The space $B(\Omega, Y)$ is a subspace of $Y^\Omega$, $\|\|_u$ is a norm on $B(\Omega, Y)$, and $(B(\Omega, Y), \|\|_u)$ is complete, i.e. $(B(\Omega, Y), \|\|_u)$ is a Banach space.

**Exercise 1.1.** Prove the first part of Proposition 11.3. That is show $B(\Omega, Y)$ is a subspace of $Y^\Omega$, $\|\|_u$ is a norm on $B(\Omega, Y)$. Hint: model your proof on Exercise 6.1.

**Exercise 1.2.** Prove the second part of Proposition 11.3. That is show $(B(\Omega, Y), \|\|_u)$ is complete, i.e. $(B(\Omega, Y), \|\|_u)$ is a Banach space. Hint: model your proof on Lemma 6.27.

**Notation 11.4 (Indicator Functions)** The indicator function, $1_A$, of a subset $A \subset \Omega$ is the function defined by,

$$1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}.$$  

**Notation 11.5** Given a function, $f : \Omega \to Y$ and $y \in Y$, let

$$\{f = y\} := f^{-1}\{\{y\}\} := \{\omega \in \Omega : f(\omega) = y\}.$$  

**Example 11.6 (Simple functions).** Given subsets $A_k \subset \Omega$ and $y_k \in Y$ for $1 \leq k \leq n$, the function,

$$f = \sum_{k=1}^{n} y_k 1_{A_k} \in B(\Omega, Y).$$

Notice that if $S = \{z = \sum_{k \in A} y_k : A \subset \{1, 2, \ldots, n\}\}$ where by convention $z = 0$ if $A = \emptyset$, then $f(\Omega) \subset S$ and in particular $f(\Omega)$ is a finite set with no more that $2^n$ – elements. Conversely if $f : \Omega \to Y$ is a function such that $f(\Omega)$ is a finite subset of $Y$, then

$$f = \sum_{y \in f(\Omega)} y \cdot 1_{\{f=y\}}$$

where

$$\{f = y\} := \{\omega \in \Omega : f(\omega) = y\} = f^{-1}\{\{y\}\}.$$  

A function $f : \Omega \to Y$ such that $f(\Omega)$ is a finite set is called a **simple** function.

We are now going to take $\Omega = [a, b]$ where $-\infty < a < b < \infty$.

**Definition 11.7 (Partitions).** A **partition** of $[a, b]$ is a finite subset, $\Pi$, of $[a, b]$ such that $a, b \in \Pi$. We will typically write $\Pi$ as,

$$\Pi = \{a = t_0 < t_1 < \cdots < t_n = b\}. \quad (11.1)$$

The **mesh size** of $\Pi$ is defined to be,

$$|\Pi| := \max_{1 \leq i \leq n} |t_i - t_{i-1}|.$$  

**Notation 11.8** Given a partition, $\Pi$, of $[a, b]$ as in Eq. (11.1) and $c = (c_1, \ldots, c_n) \in [a, b]$ such that $t_{i-1} \leq c_i \leq t_i$ for $1 \leq i \leq n$ [we will say that $c$ **interlaces** $\Pi$] let

$$f_{\Pi,c} := f(c_1) 1_{[a]} + \sum_{i=1}^{n} f(c_i) 1_{(t_{i-1}, t_i)}.$$
Exercise 11.3 (Uniform Continuity). Show that every continuous function, \( f \): \([a, b] \to Y \) is uniformly continuous, i.e. if \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that \( \| f(t) - f(s) \|_Y \leq \varepsilon \) if \( s, t \in [a, b] \) with \( |t - s| \leq \delta \). \textbf{Hint:} see Exercise 9.24

Proposition 11.9. If \( f \): \([a, b] \to Y \) is a continuous function with \( f([a, b]) \) uniformly continuous, then \( f \in B([a, b], Y) \). Let us further suppose that \( \{ \Pi_m \}_{m=1}^\infty \) is any collection of partitions of \([a, b]\) such that \( \lim_{m \to \infty} \| \Pi_m \| = 0 \) and that \( f_m := f_{\Pi_m}, c_m \) where \( c_m \) interlaces \( \Pi_m \), then \( \lim_{m \to \infty} \| f - f_m \|_Y = 0 \). In particular, every continuous function is the uniform limit of simple functions.

Exercise 11.4. Prove Proposition 11.9. Suppose that \( f \): \([a, b] \to Y \) is a continuous function.

1. Show \( f \in B([a, b], Y) \).
2. Show for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that for every partition \( \Pi \) of \([a, b]\) with \( \| \Pi \| \leq \delta \) we have \( \| f - f_{\Pi}, c \|_Y \leq \varepsilon \) for any collections of times \( c \) which interlace \( \Pi \). In particular if \( \{ \Pi_m \}_{m=1}^\infty \) is a collection of partitions of \([a, b]\) such that \( \lim_{m \to \infty} \| \Pi_m \| = 0 \) and that \( f_m := f_{\Pi_m}, c_m \) where \( c_m \) interlaces \( \Pi_m \), then \( \lim_{m \to \infty} \| f - f_m \|_Y = 0 \). \textbf{Hint:} make use of Exercise 11.3

11.2 Algebras of Sets

Notation 11.10 (Pairwise Disjoint Unions) If \( \{ A_\alpha \}_{\alpha \in I} \) and \( A \) are subsets of \( \Omega \), we write \( A = \sum_{\alpha \in I} A_\alpha \) to mean that \( A = \cup_{\alpha \in I} A_\alpha \) with the added restriction that \( A_\alpha \cap A_\beta = \emptyset \) if \( \alpha \neq \beta \). We will also write \( A \bigcup B \) for \( A \cup B \) with the added condition that \( A \cap B = \emptyset \). Similarly \( A \bigcup B \bigcup C \in A \cup B \cup C \) with the added requirement that \( A \cap B = A \cap C = B \cap C = \emptyset \).

Definition 11.11. A set \( S \subset 2^\Omega \) is said to be an semialgebra or elementary class provided that

- \( \emptyset \in S \)
- \( S \) is closed under finite intersections
- if \( E \in S \), then \( E^c \) is a finite disjoint union of sets from \( S \). (In particular \( \Omega = \emptyset \) is a finite disjoint union of elements from \( S \).)

Notation 11.12 A subset \( J \subset \mathbb{R} \) is said to be a half open interval \( J = (a, b] \subset \mathbb{R} \) for some \( -\infty \leq a \leq b \leq \infty \). In other words a half open interval is a set of the form \( (a, b] \) with \( -\infty \leq a \leq b < \infty \) or of the form \( (a, \infty) \) with \( -\infty \leq a < \infty \). If \( J \) and \( K \) are two half open intervals we will write \( J < K \) iff \( s < t \) for all \( s \in J \) and \( t \in K \). More concretely, \( (a, b] \leq (c, d] \cap \mathbb{R} \) iff \( b \leq c \).

Example 11.13. Let \( \Omega = \mathbb{R} \), then

\[ S := \{(s, t) \cap \mathbb{R} : s, t \in \mathbb{R}\} = \{(s, t) : s \in [-\infty, \infty) \text{ and } s < t < \infty\} \cup \emptyset \mathbb{R} \]

is an elementary class.

Example 11.14. Let \( -\infty < a \leq b < \infty \), \( \Omega = (a, b] \), and \( S := \{(s, t) : a \leq s \leq t \leq b\} \)

is an elementary class.

Example 11.15. Suppose that \( \Omega \) is a finite set, then

\[ S := \{\omega : \omega \in \Omega\} \cup \emptyset \]

is an elementary class.

Example 11.16. Suppose that \( \Omega = (a_1, b_1] \times (a_2, b_2) \subset \mathbb{R}^2 \). Then

\[ S := \{(s_1, t_1) \times (s_2, t_2) : a_i \leq s_i \leq t_i \leq b_i \text{ for } i = 1, 2\} \]

is an elementary class as follows from from Example 11.14 and Exercise 11.5 below or may be verified directly. This example has an obvious generalization to any dimension.
Exercise 11.5. Let $\Omega_1$ and $\Omega_2$ be sets and $\mathcal{A}_1 \subset 2^{\Omega_1}$ and $\mathcal{A}_2 \subset 2^{\Omega_2}$ be elementary classes. Show the collection

$$S := \{ A \times B : A \in \mathcal{A}_1 \text{ and } B \in \mathcal{A}_2 \} \subset 2^{\Omega_1 \times \Omega_2}$$

is also a elementary class.

Proposition 11.17 (An algebra of sets). Let $S$ be a semi-algebra on $\Omega$ and let $\mathcal{A} = \mathcal{A}(S)$ denote the subsets of $\Omega$ which may be written as a finite disjoint union of elements from $S$. So $A \in \mathcal{A}$ iff there $E_i \in S$ such that $A = \sum_{i=1}^n E_i$. Then $\mathcal{A}$ is an algebra of subsets of $\Omega$, i.e.;

1. $\emptyset$ and $\Omega$ belong to $\mathcal{A}$,
2. if $A, B \in \mathcal{A}$, then $A \cap B \in \mathcal{A}$,
3. if $A \in \mathcal{A}$ then $A^c = \Omega \setminus A \in \mathcal{A}$, and
4. if $A, B \in \mathcal{A}$ then $A \cup B \in \mathcal{A}$.

Proof. We check each assertion in turn.

1. By the properties of $S$, we know that $\emptyset, \Omega \in \mathcal{A}$.
2. ($\mathcal{A}$ is closed under intersections.) Suppose that $A = \sum_{i=1}^n E_i \in \mathcal{A}$ and $B = \sum_{j=1}^m F_j \in \mathcal{A}$ where $E_i, F_j \in S$. Then

$$A \cap B = \left( \sum_{i=1}^n E_i \right) \cap \left( \sum_{j=1}^m F_j \right) = \sum_{i,j} E_i \cap F_j$$

where we are justified in writing $\sum_{i,j}$ since,

$$[E_i \cap F_j] \cap [E_{i'} \cap F_{j'}] = [E_i \cap E_{i'}] \cap [F_j \cap F_{j'}] = \emptyset$$

unless $i = i'$ and $j = j'$. Therefore $\mathcal{A}$ is closed under intersections.
3. ($\mathcal{A}$ is closed under complementation.) If $A = \sum_{i=1}^n E_i \in \mathcal{A}$ with $E_i \in S$, then $A^c = \cap_{i=1}^n E_i^c \in \mathcal{A}$ from item 2. and the property of $S$ that implies $E_i^c \in \mathcal{A}$ whenever $E_i \in S$.
4. ($\mathcal{A}$ is closed under unions.) If $A, B \in \mathcal{A}$, then

$$A \cup B = \left[ \left( A \cup B \right) ^c \right]^c = \left[ A^c \cap B^c \right]^c \in \mathcal{A}$$

as $A^c, B^c \in \mathcal{A}$ and so $A^c \cap B^c \in \mathcal{A}$ and therefore $[A^c \cap B^c]^c \in \mathcal{A}$.

Example 11.18. Let $\Omega = \mathbb{R}$, then $\mathcal{A}$ consisting of finite disjoint unions of elements from

$$S := \{(a, b) \cap \mathbb{R} : a, b \in \mathbb{R} \} = \{(a, b) : a \in [-\infty, \infty) \text{ and } a < b < \infty \} \cup \{(\emptyset, \mathbb{R}) \}$$

is an algebra. For example,

$$A = (0, \pi] \cup (2\pi, 7] \cup (11, \infty) \in \mathcal{A}(S).$$

Example 11.19 (Algebra of half open intervals). Let $\Omega = (a, b)$. The algebra of half open intervals, $\mathcal{A}$, on $\Omega$ is the collection of subsets of $(a, b)$ which may be written as a union of pairwise disjoint half open subintervals of $(a, b)$. [The general element of $\mathcal{A}$ may be written as $A = \sum_{i=1}^n J_i$ with $J_1 < J_2 < \cdots < J_n$ and $J_i \subset (a, b)$ for all $i$.]

Lemma 11.20. Suppose that $F$ is a finite collection of half open intervals which are pairwise disjoint, i.e. if $J, K \in F$ with $J \neq K$ then $J \cap K = \emptyset$. If $n = \#(F)$, there exists a unique listing of $F$ as $\{J_1, \ldots, J_n\}$ with $J_1 < J_2 < \cdots < J_n$.

Proof. The proof is by induction on $n$. If $n = 1$, $F = \{J_1\}$ and we are done. Now suppose the statement is true when $\#(F) = n - 1$ for some $n \geq 2$ and now suppose that $\#(F) = n$. Choose $J_1 = (a_1, b_1) \in F$ so that it has the smallest possible left end point. By an induction hypothesis we can list $F \setminus \{J_1\}$ as $\{J_2, \ldots, J_n\}$ with $J_2 < J_3 < \cdots < J_n$. It now only remains to show $J_1 = (a_1, b_1) < J_2 = (a_2, b_2)$. By construction we know that $a_2 \geq a_1$ and by assumption that $J_1 \cap J_2 = \emptyset$. These two conditions can only hold if $a_2 \geq b_1$, i.e. $J_1 < J_2$.

Definition 11.21. Let $\Omega$ be a set. A collection, $\Pi$, of non empty subsets of $\Omega$ is a partition of $\Omega$ if $\Omega = \sum_{\Pi} A$. In other words, every point of $x \in \Omega$, belongs precisely to one and only one element, $A$, in $\Pi$. We say that $\Pi$ is a finite partition if we further assume that $\#(\Pi) < \infty$.

Corollary 11.22. If $\Pi$ is a finite partition of $(a, b)$ consisting of half open sub-intervals of $(a, b)$. Then there exists $\{t_k\}_{k=0}^n \subset (a, b)$ such that $\Pi = \{ \{t_{k-1}, t_k\}_{k=1}^n \}$ where $n = \#(\Pi)$ and $a = t_0 < t_2 < t_3 < \cdots < t_n = b$.

Proof. By Lemma 11.20 we know that $\Pi = \{J_k = (a_k, b_k)\}_{k=1}^n$ with $J_1 < J_2 < \cdots < J_n$. With this ordering we must have $a = a_1$ and $b = b_n$. If $b_k < a_{k+1}$ for some $k < n$ then $\emptyset \neq \{b_k, a_{k+1}\} \subset (a_k, b_k) \cap \Omega$ while on the other hand,

$$(b_k, a_{k+1}) \cap (a, b) \cap \Omega = (b_k, a_{k+1}) \cap \bigcup_{J \in \Pi} J = \emptyset.$$ 

Thus we conclude that $b_k = a_{k+1}$ for $k = 1, 2, \ldots, n - 1$ and therefore the theorem holds with $t_0 = a$ and $t_k = b_k$ for $1 \leq k \leq n$. \[\]
Exercise 11.6 (Structure of finite algebras). Suppose that $\Omega$ is a set and $\mathcal{A}$ is a sub-algebra of $2^\Omega$ with only finitely many elements, i.e. $\#(\mathcal{A}) < \infty$. For each $x \in \Omega$ let

$$A_x = \cap \{ A \in \mathcal{A} : x \in A \} \in \mathcal{A},$$

wherein we have used $\mathcal{A}$ is finite to ensure $A_x \in \mathcal{A}$. Hence $A_x$ is the smallest set in $\mathcal{A}$ which contains $x$, i.e. if $A \in \mathcal{A}$ and $x \in A$ then $A_x \subset A$. Show;

1. If $x, y \in \Omega$ with $A_x \cap A_y \neq \emptyset$ then $A_x = A_y$ so either $A_x = A_y$ or $A_x \cap A_y = \emptyset$.

Hint: if $x \notin A_y$, then $A_x \setminus A_y$ would be a set in $\mathcal{A}$ which is strictly smaller than $A_x$ – this would violate the definition of $A_x$.

2. Let $\Pi := \{A_1, \ldots, A_n\} \subset \mathcal{A}$ be the list of distinct elements in $\{A_x : x \in \Omega\}$. Show that every $A \in \mathcal{A}$ may be written as $A = \sum_{j \in A} A_j$ for some finite subset $A \subset \{1, 2, \ldots, n\}$. By convention, if $A = \emptyset$ we define $\sum_{j \in A} A_j = \emptyset \in 2^\Omega$. In particular this shows $\Omega = \sum_{i=1}^n A_i$ and hence $\Pi$ is a partition of $\Omega$. [Conclusion, finite subalgebras of $2^\Omega$ are in one to one correspondence with finite partitions of $\Omega$.]

3. Explain why there is no algebra $\mathcal{A} \subset 2^\Omega$ such that $\#(\mathcal{A}) = 6$.

3.3 Finitely additive measure on $\mathcal{A}$

Let $Z$ be a vector space (typically we are going to take $Z = \mathbb{R}$).

Definition 11.23. An $Z$-valued finitely additive measure on $\mathcal{A}$ is a function, $\mu : \mathcal{A} \to Z$ such that;

1. $\mu(\emptyset) = 0 \in Z$, and
2. $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A, B \in \mathcal{A}$ with $A \cap B = \emptyset$. [By induction it then follows that $\mu\left(\sum_{j=1}^n A_j\right) = \sum_{j=1}^n \mu(A_j)$ where $A_j \in \mathcal{A}$]

Exercise 11.7. Let $\Omega$ be a set, $\mathcal{A}$ be a sub-algebra of $2^\Omega$, and $\mu : \mathcal{A} \to [0, \infty)$ be a finitely additive measure. Further suppose that $A, B \in \mathcal{A}$ and $\{A_j\}_{j=1}^n \subset \mathcal{A}$.

Show;

1. $\mu$ is monotone If $A \subset B$, then $\mu(A) \leq \mu(B)$ and more precisely, $\mu(B) - \mu(A) = \mu(B \setminus A)$.
2. Show $\mu\left(\sum_{j=1}^n A_j\right) = \sum_{j=1}^n \mu(A_j)$ if $A_j \cap A_i = \emptyset$ when $i \neq j$.
3. For any $A, B \in \mathcal{A}$

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B).$$

(11.2)

4. $\mu$ is finitely subadditive $\mu(\bigcup_{j=1}^n A_j) \leq \sum_{j=1}^n \mu(A_j)$.

Definition 11.24. If $S \subset 2^\Omega$ is a semi-algebra, we say a function $\mu : S \to Z$ is additive on $S$ if whenever $n \in \mathbb{N}$ and $E = \sum_{i=1}^n E_i$ with $E_i, E \in S$, then

$$\mu(E) = \sum_{i=1}^n \mu(E_i).$$

Lemma 11.25. Let $\Omega = (a, b)$, $\alpha : [a, b] \to Z$ be a function, and $S$ be the collection of half open intervals in $\Omega$. Then $\mu : S \to Z$ defined by

$$\mu((s, t]) = \alpha(t) - \alpha(s) \text{ for all } a \leq s \leq t \leq b$$

is additive on $S$.

Proof. Suppose that $J = \sum_{i=1}^n J_i$ with $J = (u, v)$ and $J_i := (u_i, v_i]$ in $S$. We must show

$$\mu(J) = \sum_{i=1}^n \mu(J_i).$$

(11.3)

As the sum in Eq. (11.3) is independent of the ordering we may reorder the terms so that $J_1 < J_2 < \cdots < J_n$ and then use Corollary 11.22 to deduce that $J_i = (t_{i-1}, t_i]$, where $u = t_0 < t_1 < \cdots < t_n = v$. With this notation, we find by a telescoping series arguments (see Theorem 7.3) that

$$\sum_{i=1}^n \mu(J_i) = \sum_{i=1}^n [\alpha(t_i) - \alpha(t_{i-1})] = \alpha(t_n) - \alpha(t_0) = \alpha(v) - \alpha(u) = \mu(J).$$

Theorem 11.26 (Construction of Finitely Additive Measures). Suppose $S \subset 2^\Omega$ is a semi-algebra (see Definition 11.11) and $A = A(S)$ is the algebra generated by $S$, see Proposition 11.17. Then every additive function $\mu : S \to Z$ such that $\mu(\emptyset) = 0$ extends uniquely to an additive measure on $A$. This unique extension will still be denoted by $\mu$.

Proof. Since every element $A \in A$ is of the form $A = \sum_i E_i$ for a finite collection of $E_i \in S$, it is clear that if $\mu$ extends to a measure on $A$ then this extension is unique and we must define $\mu(A)$ by

$$\mu(A) = \sum_i \mu(E_i).$$

(11.4)

To prove existence, the main point is to show that $\mu(A)$ in Eq. (11.4) is well defined; i.e. if we also have $A = \sum_j F_j$ with $F_j \in S$, then we must show
\[ \sum_i \mu(E_i) = \sum_j \mu(F_j). \]  
(11.5)

But \( E_i = \sum_j (E_i \cap F_j) \) and the additivity of \( \mu \) on \( \mathcal{S} \) implies \( \mu(E_i) = \sum_j \mu(E_i \cap F_j) \) and hence

\[ \sum_i \mu(E_i) = \sum_j \sum_i \mu(E_i \cap F_j) = \sum_{i,j} \mu(E_i \cap F_j). \]

Similarly,

\[ \sum_j \mu(F_j) = \sum_{i,j} \mu(E_i \cap F_j) \]

which combined with the previous equation shows that Eq. (11.5) holds.

Finally if \( A = \sum_i E_i \) and \( B = \sum_{j=1}^m F_j \) with \( A \cap B = \emptyset \), then \( A \cup B = (\sum_i E_i) \prod_{j=1}^m F_j \) so that

\[ \mu(A \cup B) = \sum_i \mu(E_i) + \sum_j \mu(F_j) = \mu(A) + \mu(B). \]

This shows that \( \mu \) is a finitely additive measure on \( \mathcal{A} \).

\[ \square \]

**Corollary 11.27.** Let \( \alpha : [a, b] \to Z \) be any function, \( \Omega = (a, b) \), and \( \mathcal{A} \) be the algebra of half open intervals in \( \Omega \). Then there exists a unique finitely additive measure \( \mu = \mu_\alpha : \mathcal{A} \to Z \) such that

\[ \mu ((s, t]) = \alpha(t) - \alpha(s) \] for all \( a \leq s \leq t \leq b \).

**Proof.** This corollary follows directly from Lemma 11.26 and Theorem 11.26.

**Remark 11.28.** Let \( \Omega = (a, b) \) and \( \mathcal{A} \) be the algebra of half open intervals. Given a finitely additive measure \( \mu : \mathcal{A} \to Z \), let \( \alpha : [a, b] \to Z \) be defined by \( \alpha(t) := \mu((a, t]) \in Z \). Then if \( J = (s, t] \subset (a, b) \), then \( (a, t] = (a, s) \cup (s, t] \) and so

\[ \alpha(t) = \mu((a, t]) = \mu((a, s]) + \mu((s, t]) = \alpha(s) + \mu((s, t]) \]

and hence

\[ \mu(J) = \mu((s, t]) = \alpha(t) - \alpha(s). \]

Hence if \( A = \sum_{i=1}^n (a_i, b_i] \) we find

\[ \mu(A) = \sum_{i=1}^n \mu((a_i, b_i]) = \sum_{i=1}^n [\alpha(b_i) - \alpha(a_i)] \]  
(11.6)

which shows \( \mu \) is uniquely determined by its cumulative distribution function, \( \alpha \).

**Example 11.29.** Suppose that \( \Omega \) is a finite set and let \( \mathcal{S} := \{\{\omega\} : \omega \in \Omega\} \cup \emptyset \).

Then \( \mathcal{S} \) is a semi-algebra and the associated algebra is \( \mathcal{A} = 2^\Omega \). Given any function, \( \alpha : \Omega \to Z \) we may define \( \mu : \mathcal{S} \to Z \) by \( \mu(\emptyset) = 0 \) and \( \mu(\{\omega\}) = \alpha(\omega) \).

Then an application of Theorem 11.26 states that \( \mu \) extends to a finitely additive measure on \( \mathcal{A} \) given by

\[ \mu(A) = \mu \left( \sum_{\omega \in A} \{\omega\} \right) = \sum_{\omega \in A} \mu(\{\omega\}) = \sum_{\omega \in A} \alpha(\omega). \]

**11.4 Simple Functions**

In this section, let \( F \) be either \( \mathbb{R} \) or \( \mathbb{C} \) and let \( Y \) be a vector space over \( F \). Further suppose that \( \Omega \) is a set and \( \mathcal{A} \subset 2^\Omega \) is an algebra of sets.

**Definition 11.30 (Simple functions).** A function, \( f : \Omega \to Y \) is said to be simple if \( f(\Omega) \subset Y \) is a finite set. [See Example 11.6 for the structure of simple functions.] We further say that \( f \) is an \( \mathcal{A} \)-simple function if \( \{f = y\} \in \mathcal{A} \) for all \( y \in f(\Omega) \subset Y \). Let \( \mathcal{S}(\mathcal{A}, Y) \) denote the collection of \( Y \)-valued simple functions on \( \Omega \).

**Example 11.31.** If \( y \in Y \) and \( A \in \mathcal{A} \), then \( f := 1_A \cdot y \) (so \( f(\omega) = y \) if \( \omega \in A \) and \( 0 \) otherwise) is an \( \mathcal{A} \)-simple function. Indeed, assuming \( y \neq 0 \), we have \( f(\Omega) = \{0, y\} \) and \( f^{-1}(\{y\}) = A \in \mathcal{A} \) and \( f^{-1}(\{0\}) = A^c \in \mathcal{A} \). If \( y = 0 \), then \( f(\Omega) = \{0\} \) and \( f^{-1}(\{0\}) = \Omega \in \mathcal{A} \).

**Lemma 11.32.** The set of simple functions, \( \mathcal{S}(\mathcal{A}, Y) \), is a vector subspace of all functions from \( \Omega \) to \( Y \). Moreover, \( f : \Omega \to Y \) is an \( \mathcal{A} \)-simple function if it may be written as

\[ f = \sum_{i=1}^N y_i 1_{A_i} \text{ for some } y_i \in Y \text{ and } A_i \in \mathcal{A}. \]  
(11.7)

One may always let \( \{y_1, \ldots, y_N\} \) be a listing of \( f(\Omega) \) and then take \( A_i := f^{-1}(\{y_i\}) \in \mathcal{A} \) for each \( i \) in which case \( A_i \cap A_j = \emptyset \) for \( i \neq j \).

**Proof.** Let us observe that \( 1_\Omega = 1 \) and \( 1_\emptyset = 0 \) are in \( \mathcal{S}(\mathcal{A}, Y) \). If \( f, g \in \mathcal{S}(\mathcal{A}, Y) \) and \( c \in F \setminus \{0\} \), then for any \( y \in Y \),

\[ \{f + cg = y\} = \bigcup_{a,b \in \mathbb{C} : a+cb=y} (\{f = a\} \cap \{g = b\}) \in \mathcal{A}. \]  
(11.8)

This shows that \( \mathcal{S}(\mathcal{A}, Y) \) is a vector subspace.

From Example 11.31 and the fact that \( \mathcal{S}(\mathcal{A}, Y) \) is a linear subspace, any function given in the form of Eq. (11.7) is an \( \mathcal{A} \)-simple function. Conversely if \( f \) is an \( \mathcal{A} \)-simple function, let \( \{y_1, \ldots, y_N\} \) be a listing of \( f(\Omega) \) and set \( A_i := f^{-1}(\{y_i\}) \). Then Eq. (11.7) holds with \( A_i \cap A_j = \emptyset \) if \( i \neq j \).
Corollary 11.33. If \( f \in S(A, Y) \) and \( F : Y \to Z \) is any function from \( Y \) to another vector space \( Z \), then \( F \circ f \in S(A, Z) \). In particular, if \( \| \cdot \|_Y \) is a norm on \( Y \), then \( \| f(\cdot) \|_Y \in S(A, \mathbb{R}) \).

**Proof.** We may write \( f \) as in Eq. (11.7) with \( A_i \cap A_j = \emptyset \) for all \( i \neq j \). We then have,

\[
F \circ f = \sum_{i=1}^{N} F(y_i) 1_{A_i}
\]

which is in \( S(A, Z) \) by Lemma 11.32. \( \blacksquare \)

Lemma 11.34. The set of simple functions, \( S(A, Y) \), is a vector subspace of all functions from \( \Omega \) to \( Y \). Moreover, \( f, g \in S(A, F) \) and \( g \in S(A, Y) \) then \( f \cdot g \in S(A, Y) \). [In algebraic jargon, \( S(A, F) \) is an algebra of functions and \( S(A, Y) \) is a module over \( S(A, F) \).]

**Proof. First proof.** We may write \( f \) and \( g \) as,

\[
f = \sum_{i=1}^{N} \lambda_i 1_{A_i}, \text{ and } g = \sum_{j=1}^{M} y_j 1_{B_j}
\]

with \( \lambda_i \in F, y_j \in Y, \) and \( A_i, B_j \in A \). It then follows that

\[
f \cdot g = \sum_{i=1}^{N} \sum_{j=1}^{M} \lambda_i y_j 1_{A_i \cap B_j}
\]

which is a linear combination of \( A \)–indicator functions and hence is in \( S(A, Y) \).

**Second proof.** We first observe that the \( f \cdot g \) can only take values in \( \{ a y : a \in f(\Omega) \text{ and } y \in g(\Omega) \} \) which is a finite set. Therefore \( f \cdot g \) is a simple function. If \( y \in Y \setminus \{0\} \), then

\[
\{f \cdot g = y\} = \cup_{a \in f(\Omega)} \{a y \} \cap \{g = z\} \subset A \text{ (11.9)}
\]

where the union is over those \( a \in f(\Omega) \setminus \{0\} \) and \( z \in g(\Omega) \setminus \{0\} \) such that \( az = y \). We also have

\[
\{f \cdot g = 0\} = \{f = 0\} \cup \{g = 0\} \in A
\]

and thus we have shown \( f \cdot g \in S(A, F) \). \( \blacksquare \)

Definition 11.35. When \( \Omega = (a, b] \), we say that \( f : \Omega \to Y \) is a step function if there is a partition, \( \Pi = \{a = t_0 < t_1 < \cdots < t_n = b\} \) of \( \Omega \) such that \( f \) is constant on \( (t_{i-1}, t_i] \) for each \( i \), i.e. there exists \( y_i \in Y \) such that

\[
f = \sum_{i=1}^{n} y_i 1_{(t_{i-1}, t_i]}. \quad (11.10)
\]

Lemma 11.36. Let \( \Omega = (a, b] \) and \( f : \Omega \to Y \) be a function, then;

1. \( f \) is an \( A \)–simple function iff \( f \) is a step function.
2. Suppose \( X \) is another vector space and \( F : Y \to X \) is any function then \( F \circ f \in S(A, X) \) for all \( f \in S(A, Y) \). In particular, \( \| f \|_Y \in S(A, \mathbb{R}) \) when \( f \in S(A, Y) \).

**Proof.** 1. All step functions are \( A \)–simple functions by Lemma 11.32. Conversely if \( f \) is an \( A \)–simple function then

\[
f = \sum_{y \in f(\Omega)} y \cdot 1_{\{f = y\}}
\]

where \( \{f = y\} = \cup_{k=1}^{n} J_k^y \) with \( \{J_k^y\} \) being half open intervals. It then follows that \( \Omega = \cup_{k=1}^{n} \cup_{t \in J_k^y} \{t\} \) and so that collection of half open intervals \( \cup_{y \in f(\Omega)} \{J_k^y : 1 \leq k \leq n_y \} \) forms a partition of \( \Omega \). Therefore there exists \( \Pi = \{a = t_0 < t_1 < \cdots < t_N = b\} \) such that \( (t_{i-1}, t_i] \) is in \( \cup_{y \in f(\Omega)} \{J_k^y : 1 \leq k \leq n_y \} \) for each \( i \) and therefore \( f \) is constant on \( (t_{i-1}, t_i] \), i.e. \( f \) is a step function.

2. If \( f \in S(A, Y) \) and \( F : Y \to X \) is any function then

\[
F \circ f = F \circ \left( \sum_{y \in f(\Omega)} y \cdot 1_{\{f = y\}} \right) = \sum_{y \in f(\Omega)} F(y) \cdot 1_{\{f = y\}} \in S(A, X).
\]

\( \blacksquare \)

11.5 Simple Integration

Definition 11.37. A finitely additive measure space is a triple \( (\Omega, A, \mu) \) where \( \Omega \) is a set, \( A \) is a sub-algebra of \( 2^\Omega \), and \( \mu : A \to \mathbb{R}_+ \) is a finitely additive measure.

In this section we will suppose that \( (\Omega, A, \mu) \) is a finitely additive measure space and \( Y \) is a normed space.

Definition 11.38 (Simple Integral). For \( f \in S(A, Y) \) the integral (written as \( \int_\Omega f d\mu \) or \( \int_\Omega f(\omega) d\mu(\omega) \)) is defined by

\[
\int_\Omega f d\mu = \sum_{y \in f(\Omega)} \mu(\{f = y\}) \cdot y \in Y.
\]

(11.11)

where \( \mu(\{f = y\}) \) is shorthand for \( \mu(\{f = y\}) \). It is often convenient to write the last sum as
The integration operator,

Proposition 11.40.

\[
\sum_{y \in Y} \mu(f = y) \cdot y
\]

with the understanding that this is really a finite sum since \( \mu(f = y) \neq 0 \) happens only when \( y \in f(\Omega) \) which is a finite set.

To simplify notation I will often write \( \int f \) for \( \int f \, d\mu \) in this section.

Example 11.39. Suppose that \( A \in \mathcal{A} \) and \( y \in Y \), then

\[
\int_{\Omega} y 1_A \, d\mu = \int_{\Omega} y 1_A \, d\mu = \mu(A) y = y \int_{\Omega} 1_A \, d\mu. \quad (11.12)
\]

Proposition 11.40. The integration operator, \( \int : \mathcal{S}(\mathcal{A}, Y) \to Y \), satisfies:

1. If \( f \in \mathcal{S}(\mathcal{A}, Y) \) and \( \lambda \in \mathbb{F} \), then

\[
\int \lambda f = \lambda \int f. \quad (11.13)
\]

2. If \( f, g \in \mathcal{S}(\mathcal{A}, Y) \), then

\[
\int (f + g) = \int f + \int g. \quad (11.14)
\]

Items 1. and 2. say that integration is a linear functional on \( \mathcal{S}(\mathcal{A}, Y) \).

3. If \( f = \sum_{j=1}^N y_j 1_{A_j} \) for some \( y_j \in Y \) and some \( A_j \in \mathcal{A} \), then

\[
\int f = \sum_{j=1}^N y_j \mu(A_j). \quad (11.15)
\]

4. For all \( f \in \mathcal{S}(\mathcal{A}, Y) \),

\[
\left\| \int f \right\|_Y \leq \int_{\Omega} \| f(t) \| \, d\mu(t) \leq \| f \|_\infty \cdot \mu(\Omega). \quad (11.16)
\]

where

\[
\| f \|_\infty := \sup_{\omega \in \Omega} \| f(\omega) \|_Y.
\]

5. If \( X \) is another vector space and \( T : Y \to X \) is a linear operator, then

\[
T \int_{\Omega} f \, d\mu = \int_{\Omega} T(f) \, d\mu.
\]

Proof.

1. If \( \lambda \in \mathbb{F} \) and \( \lambda \neq 0 \), then

\[
\int \lambda f = \sum_{y \in Y} y \mu(\lambda f = y) = \sum_{y \in Y} y \mu(f = y/\lambda)
\]

\[
= \sum_{z \in Y} \lambda z \mu(f = z) = \lambda \int f.
\]

The case \( \lambda = 0 \) is trivial.

2. Writing \( \{ f = v, g = w \} \) for \( f^{-1}(\{ v \}) \cap g^{-1}(\{ w \}) \), then

\[
\int (f + g) = \sum_{z \in Y} z \mu(f + g = z)
\]

\[
= \sum_{z \in Y} z \left( \sum_{v+w=z} \mu(\{ f = v, g = w \}) \right)
\]

\[
= \sum_{z \in Y} \sum_{v+w=z} (v+w) \mu(\{ f = v, g = w \})
\]

\[
= \sum_{v,w} (v+w) \mu(\{ f = v, g = w \}).
\]

But

\[
\sum_{v,w} v \mu(\{ f = v, g = w \}) = \sum_{v} v \sum_{w} \mu(\{ f = v, g = w \})
\]

\[
= \sum_{v} v \mu(\cup_{w} \{ f = v, g = w \})
\]

\[
= \sum_{v} v \mu(\{ f = v \}) = \int f
\]

and similarly,

\[
\sum_{v,w} w \mu(\{ f = v, g = w \}) = \int g.
\]

Equation \( (11.14) \) is now a consequence of the last three displayed equations.

3. If \( f = \sum_{j=1}^N y_j 1_{A_j} \), then using the linearity of the integral and Example \( \mathbf{11.39} \) we find

\[
\int f = \int \left[ \sum_{j=1}^N y_j 1_{A_j} \right] = \sum_{j=1}^N y_j \int_{A_j} = \sum_{j=1}^N y_j \mu(A_j).
\]
4. By the triangle inequality,
\[ \|f\| = \left\| \sum_{y \in f(\Omega)} y\mu(f = y) \right\| \leq \sum_{y \in f(\Omega)} \|y\| \mu(f = y) = \int_{\Omega} \|f\| \, d\mu, \]
wherein the last equality we have used Eq. (11.15) and the fact that \(|f| = \sum_{y \in Y} |y| 1_{f=y}\). Since for \(y \in f(\Omega)\) we have \(\|y\| \leq \|f\|_{\infty}\), we may conclude that
\[ \sum_{y \in f(\Omega)} \|y\| \mu(f = y) \leq \sum_{y \in f(\Omega)} \|f\|_{\infty} \mu(f = y) = \|f\|_{\infty} \sum_{y \in f(\Omega)} \mu(f = y) = \|f\|_{\infty} \cdot \mu(\Omega) \]
from which we see the second inequality in Eq. (11.16) holds.

5. Since \(T\) is linear,
\[ T \circ f = \sum_{y \in f(\Omega)} T(y) 1_{f=y} \]
and therefore,
\[ \int_{\Omega} T(f) \, d\mu = \sum_{y \in f(\Omega)} \int_{\Omega} T(y) 1_{f=y} \, d\mu = \sum_{y \in f(\Omega)} T(y) \mu(f = y) = T \left( \sum_{y \in f(\Omega)} y\mu(f = y) \right) = T \left( \int_{\Omega} f \, d\mu \right) = T \int_{\Omega} f \, d\mu. \]

Proposition 11.41. If \(Y = \mathbb{F} = \mathbb{R}\), then \(f\) is positive, i.e. \(f \geq 0\) for all \(0 \leq f \in S(\mathcal{A}, \mathbb{R})\). More generally, if \(f, g \in S(\mathcal{A}, \mathbb{R})\) and \(f \leq g\), then \(f \leq g\).

**Proof.** If \(f \geq 0\) then
\[ \int f = \sum_{a \geq 0} a\mu(f = a) \geq 0 \]
and if \(f \leq g\), then \(g - f \geq 0\) so that
\[ \int g - \int f = \int (g - f) \geq 0. \]

Notation 11.42. If \(A \in \mathcal{A}\) and \(f \in S(\mathcal{A}, \mathbb{Y})\), we let
\[ \int_{\mathbb{A}} f \, d\mu := \int_{\Omega} 1_A f \, d\mu. \]

Notation 11.43. When \(\Omega := (a, b)\), \(\alpha : [a, b] \to \mathbb{R}\) is an increasing function, \(A\) is the algebra of half open intervals on \(\Omega\), and \(\mu : A \to \mathbb{R}_{+}\) is the unique measure such that \(\mu((s, t]) = \alpha(t) - \alpha(s)\) for all \(a \leq s \leq t \leq b\), we will denote \(\int_{\Omega} f \, d\mu\) by \(\int_a^b f \, d\mu\) or \(\int_a^b f \, d\alpha\). Moreover, if \(f \in S(\mathcal{A}, \mathbb{Y})\) and \(\alpha, \beta \in [a, b]\) we let
\[ \int_{\alpha}^{\beta} f \, d\mu = \begin{cases} \int_{\alpha}^{\beta} 1_{(\alpha, \beta]} f \, d\mu & \text{if } \alpha \leq \beta, \\ -\int_{\beta}^{\alpha} 1_{(\beta, \alpha]} f \, d\mu & \text{if } \alpha \geq \beta. \end{cases} \]

Proposition 11.44. Let us continue using the setup in Notation 11.43. If \(f \in S(\mathcal{A}, \mathbb{Y})\) and \(\alpha, \beta, \gamma \in [a, b]\), then
\[ \int_{\alpha}^{\gamma} f \, d\mu = \int_{\alpha}^{\beta} f \, d\mu + \int_{\beta}^{\gamma} f \, d\mu. \]
Moreover,
\[ \left\| \int_{\alpha}^{\beta} f \, d\mu \right\| \leq \left| \int_{\alpha}^{\beta} \|f\| \, d\mu \right|. \]

**Proof.** There are a number of cases to consider. I will only check two of them.

First suppose that \(\alpha \leq \beta \leq \gamma\). In this case we have
\[ 1_{(\alpha, \gamma]} f = 1_{(\alpha, \beta]} f + 1_{(\beta, \gamma]} f \]
and therefore,
\[ \int_{\alpha}^{\gamma} f \, d\mu = \int_{\alpha}^{\beta} 1_{(\alpha, \gamma]} f = \int_{\alpha}^{\beta} 1_{(\alpha, \beta]} f + \int_{\beta}^{\gamma} 1_{(\beta, \gamma]} f = \int_{\alpha}^{\beta} f \, d\mu + \int_{\beta}^{\gamma} f \, d\mu. \]
Secondly suppose that \(\alpha \leq \gamma \leq \beta\), then by the first case,
\[ \int_{\alpha}^{\beta} f \, d\mu = \int_{\alpha}^{\gamma} f \, d\mu + \int_{\gamma}^{\beta} f \, d\mu \]
and so
\[ \int_{\alpha}^{\gamma} f \, d\mu = \int_{\alpha}^{\beta} f \, d\mu - \int_{\gamma}^{\beta} f \, d\mu = \int_{\alpha}^{\beta} f \, d\mu + \int_{\gamma}^{\beta} f \, d\mu. \]
The remaining possibilities are verified by similar means.

For the last assertion, we may suppose that $\alpha \leq \beta$ and the result follows from Eq. \[11.16\] since,
\[
\left\| \int_{a}^{b} f \, d\mu \right\| = \left\| \int_{a}^{b} 1_{(\alpha, \beta]} f \, d\mu \right\| \\
\leq \int_{a}^{b} \left\| 1_{(\alpha, \beta]} f \right\| \, d\mu = \int_{a}^{b} 1_{(\alpha, \beta]} \| f \| \, d\mu = \int_{a}^{b} \| f \| \, d\mu.
\]

\[\Box\]

### 11.6 Product Measure

Suppose that $(\Omega_i, A_i, \mu_i)$ for $i = 1, 2$ be two finitely additive measure spaces. From Exercise \[11.5\] we know that $S := \{ A \times B : A \in A_1 \text{ and } B \in A_2 \} \subset 2^{\Omega_1 \times \Omega_2}$ is an elementary class. Then, by Proposition \[11.17\] the subalgebra of $A$ of $2^{\Omega_1 \times \Omega_2}$ generated by $S$ consists of those $C \subset \Omega_1 \times \Omega_2$ which may be written as
\[
C := \sum_{i} A_i \times B_i \text{ for some } A_i \in A_1 \text{ and } B_i \in A_2.
\]

**Definition 11.45 (Product Algebra).** We let $A_1 \otimes A_2$ be the subalgebra of $2^{\Omega_1 \times \Omega_2}$ consisting of sets $C \subset \Omega_1 \times \Omega_2$ of the form in Eq. \[11.17\].

**Theorem 11.46 (Product Measure).** There is a unique finitely additive measure, $\mu_1 \otimes \mu_2$, on $A := A_1 \otimes A_2$, such that
\[
\mu_1 \otimes \mu_2 (A \times B) = \mu_1 (A) \mu_2 (B) \quad \text{for all } A \in A_1 \text{ and } B \in A_2.
\]

Moreover $\mu_1 \otimes \mu_2$ satisfies,
\[
\int_{\Omega_1} \left[ \int_{\Omega_2} 1_C (x, y) \, d\mu_2 (y) \right] d\mu_1 (x) = \mu_1 \otimes \mu_2 (C) = \int_{\Omega_2} \left[ \int_{\Omega_1} 1_C (x, y) \, d\mu_1 (x) \right] d\mu_2 (y)
\]
for all $C \in A$. Part of what is being asserted here is that $y \rightarrow 1_C (x, y)$ is an $A_2$ -- simple function for each $x \in \Omega_1$ and $x \rightarrow \int_{\Omega_2} 1_C (x, y) \, d\mu_2 (y)$ is an $A_1$ -- simple function and similarly, $x \rightarrow 1_C (x, y)$ is an $A_1$ -- simple function for each $y \in \Omega_2$ and $x \rightarrow \int_{\Omega_1} 1_C (x, y) \, d\mu_1 (x)$ is an $A_2$ -- simple function so that all integrals make sense.

**Proof.** Let $C \in A$ be given as in Eq. \[11.17\]. Then we have
\[
1_C (x, y) = \sum_{i} 1_{A_i} (x) 1_{B_i} (y) \quad \text{for all } (x, y) \in \Omega_1 \times \Omega_2
\]
and hence for each fixed $x \in \Omega_1$ the functions $y \rightarrow 1_{C} (x, y)$ is an $A_2$ -- simple function and
\[
\int_{\Omega_2} 1_C (x, y) \, d\mu_2 (y) = \sum_{i} 1_{A_i} (x) \int_{\Omega_2} 1_{B_i} (y) \, d\mu_2 (y) = \sum_{i} 1_{A_i} (x) \mu_2 (B_i).
\]

From this equation we see that $x \rightarrow \int_{\Omega_2} 1_C (x, y) \, d\mu_2 (y)$ is an $A_1$ -- simple function and that
\[
\int_{\Omega_1} \left[ \int_{\Omega_2} 1_C (x, y) \, d\mu_2 (y) \right] d\mu_1 (x) = \sum_{i} \mu_2 (B_i) \int_{\Omega_1} 1_{A_i} (x) \, d\mu_1 (x)
\]
\[
= \sum_{i} \mu_1 (A_i) \cdot \mu_2 (B_i).
\]

A similar computation shows
\[
\int_{\Omega_2} \left[ \int_{\Omega_1} 1_C (x, y) \, d\mu_1 (x) \right] d\mu_2 (y) = \sum_{i} \mu_1 (A_i) \cdot \mu_2 (B_i).
\]

At any rate we may define $\mu_1 \otimes \mu_2 := \lambda : A \rightarrow \mathbb{R}_+$ by
\[
\lambda (C) := \int_{\Omega_1} \left[ \int_{\Omega_2} 1_C (x, y) \, d\mu_2 (y) \right] d\mu_1 (x).
\]

If $C_1, C_2 \in A$ are disjoint sets then
\[
1_{C_1 \cup C_2} = 1_{C_1} + 1_{C_2}
\]
and it follows that
\[
\lambda (C_1 \cup C_2) = \int_{\Omega_1} \left[ \int_{\Omega_2} \left[ 1_{C_1} (x, y) + 1_{C_2} (x, y) \right] \, d\mu_2 (y) \right] d\mu_1 (x)
\]
\[
= \int_{\Omega_1} \left[ \int_{\Omega_2} 1_{C_1} (x, y) \, d\mu_2 (y) + \int_{\Omega_2} 1_{C_2} (x, y) \, d\mu_2 (y) \right] d\mu_1 (x)
\]
\[
= \int_{\Omega_1} \left[ \int_{\Omega_2} 1_{C_1} (x, y) \, d\mu_2 (y) \right] d\mu_1 (x) + \int_{\Omega_1} \left[ \int_{\Omega_2} 1_{C_2} (x, y) \, d\mu_2 (y) \right] d\mu_1 (x)
\]
\[
= \lambda (C_1) + \lambda (C_2).
\]

Thus we have shown $\mu_1 \otimes \mu_2$ is a finitely additive measure on $A$ which satisfies Eq. \[11.18\] and
\[
\mu_1 \otimes \mu_2 (C) = \sum_{i} \mu_1 (A_i) \cdot \mu_2 (B_i)
\]
when $C \in A$ is given as in Eq. \[11.17\]. This last expression uniquely determines $\mu_1 \otimes \mu_2$. \[\Box\]
Corollary 11.47 (Area measure). Let $\Omega := (a, b)^2 \subset \mathbb{R}^2$. Then the collection $\mathcal{A}$ of subsets $C \subset \Omega$ of the form, $C = \sum_{i=1}^{n} [x_i, y_i] \times [u_i, v_i]$, is a subalgebra of $2^\Omega$, where $a \leq x_i \leq y_i \leq b$ and $a \leq u_i \leq v_i \leq b$. Moreover there is a unique finitely additive measure $m : \mathcal{A} \to \mathbb{R}^+$ such that

$$m(C) = \sum_{i=1}^{n} (y_i - x_i) (v_i - u_i)$$

where $C$ is given as above, i.e. $m$ is the total area of $C$.

Theorem 11.48 (Baby Fubini Theorem). Continuing the above notation. If $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ is an $\mathcal{A}$–simple function, then

$$\int_{\Omega_1} \left[ \int_{\Omega_2} f(x, y) \, d\mu_2(y) \right] \, d\mu_1(x) = \int_{\Omega_1 \times \Omega_2} f \, d(\mu_1 \otimes \mu_2) = \int_{\Omega_1} \left[ \int_{\Omega_2} f(x, y) \, d\mu_1(x) \right] \, d\mu_2(y).$$

Exercise 11.8. Prove Theorem 11.48. That is show

$$\int_{\Omega_1} \left[ \int_{\Omega_2} f(x, y) \, d\mu_2(y) \right] \, d\mu_1(x) = \int_{\Omega_1 \times \Omega_2} f \, d(\mu_1 \otimes \mu_2) = \int_{\Omega_1} \left[ \int_{\Omega_2} f(x, y) \, d\mu_1(x) \right] \, d\mu_2(y)$$

for all $\mathcal{A}$–simple functions, $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$. Here $(\Omega_i, \mathcal{A}_i, \mu_i)$ are two finitely additive measure spaces and $\mu_1 \otimes \mu_2$ is the product measure on $(\Omega_1 \times \Omega_2, \mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2)$. [Part of the problem here is to show all integrals are well defined.]
Extending the Integral by Uniform Limits

We now suppose that $Y$ is a Banach space, $\Omega$ is a set, $A$ is a subalgebra of $2^\Omega$, and $\mu : A \to \mathbb{R}_+$ is a finitely additive measure on $A$.

**Theorem 12.1.** Suppose that $f : \Omega \to Y$ is a function such that there exists $f_n \in S(A,Y)$ with $\|f - f_n\|_\infty \to 0$ as $n \to \infty$, then $\lim_{n \to \infty} \int f_n d\mu$ exists in $Y$. Moreover if $g_n \in S(A,Y)$ with $\|f - g_n\|_\infty \to 0$ as $n \to \infty$, then

$$\lim_{n \to \infty} \int f_n d\mu = \lim_{n \to \infty} \int f_n d\mu.$$

We denote this common limit by $\int f d\mu$ and call it the integral of $f$ relative to $\mu$.

**Proof.** To show the limit exists we need only need to use the completeness of $Y$ along with the observation that

$$\left\| \int f_n d\mu - \int f_m d\mu \right\| \leq \|f_n - f_m\|_\infty \mu(\Omega) \to 0 \quad \text{as} \quad m, n \to \infty.$$  

Similarly,

$$\left\| \int f_n d\mu - \int g_m d\mu \right\| \leq \|f_n - g_m\|_\infty \mu(\Omega) \to 0 \quad \text{as} \quad m, n \to \infty $$

since

$$\|f_n - g_n\|_\infty \leq \|f_n - f\|_\infty + \|f - g_n\|_\infty \to 0 + 0 = 0 \quad \text{as} \quad n \to \infty.$$

In light of Theorem 12.1, we have extended the integral from $\bar{S}(A,Y)$ – the closure of $S(A,Y)$ in the collection of bounded functions from $\Omega \to Y$ relative to $\|\cdot\|_\infty$.

**Exercise 12.2 (Closure of subspaces are subspaces).** Suppose that $(B,\|\cdot\|)$ is a Banach space and $S \subset B$ is a linear subspace of $B$. Show that closure, $\bar{S}$, of $S$ is a subspace of $B$ as well. [It now follows that $(\bar{S},\|\cdot\|)$ is a complete normed space, i.e. is a Banach space.]

**Theorem 12.2 (Linearity of the Integral).** The extended integral, $\int \bar{S}(A,Y) \to Y$ is still linear and satisfies the estimates,

$$\left\| \int f d\mu \right\|_Y \leq \int |f(\omega)| \|\cdot\|_Y d\mu(\omega) \leq \|f\|_\infty \mu(\Omega).$$  

(12.1)

In particular if $f_n \in \bar{S}(A,Y)$ and $f = \lim_{n \to \infty} f_n$ exists in $\bar{S}(A,Y)$, i.e. $\|f - f_n\|_\infty \to 0$ as $n \to \infty$, then

$$\lim_{n \to \infty} \int f_n d\mu = \int f d\mu = \int \left( \lim_{n \to \infty} f_n \right) d\mu.$$  

(12.2)

**Proof.** Let $\lambda \in \mathbb{F}$ and $h, g \in \bar{S}(A,Y)$ and choose $g_n, h_n \in S(A,Y)$ are chosen so that $h_n \to h$, and $g_n \to g$ uniformly on $\Omega$. Then

$$\int (h + \lambda g) d\mu = \lim_{n \to \infty} \int (h_n + \lambda g_n) d\mu$$

$$= \lim_{n \to \infty} \left[ \int h_n d\mu + \lambda \int g_n d\mu \right]$$

$$= \int h d\mu + \lambda \int g d\mu.$$

Similarly using $\|h_n\| \to \|h\|$ uniformly on $\Omega$ we have,

$$\left\| \int h d\mu \right\| = \lim_{n \to \infty} \left\| \int h_n d\mu \right\| \leq \lim_{n \to \infty} \int \|h_n\| d\mu = \int \|h\| d\mu$$

and

$$\int \|h_n\| d\mu \leq \|h_n\|_\infty \mu(\Omega)$$

so that

$$\lim_{n \to \infty} \int \|h_n\| d\mu \leq \lim_{n \to \infty} \|h_n\|_\infty \mu(\Omega) = \|h\|_\infty \mu(\Omega).$$
For the last assertion we need only observe that
\[
\left\| \int_{\Omega} f d\mu - \int_{\Omega} f_n d\mu \right\| = \left\| \int_{\Omega} (f - f_n) d\mu \right\| \\
\leq \|f - f_n\|_{\infty} \mu(\Omega) \to 0 \text{ as } n \to \infty.
\]

\begin{proof}

Notation 12.3 If \( A \in \mathcal{A} \) and \( f \in \mathfrak{S}(A,Y) \) we let
\[
\int_{A} f d\mu := \int \mathbf{1}_{A} f d\mu.
\]

Theorem 12.4 (Summary of Integration). Given a measure space, \((\Omega, \mathcal{A}, \mu)\), there is a unique linear map,
\[
\mathfrak{S}(A,Y) \in f \to \int_{\Omega} f d\mu \in Y
\]

such that \( f = y \mathbf{1}_{A} \) for some \( y \in Y \) and \( A \in \mathcal{A} \), then \( \int_{\Omega} f d\mu = y \mu(A) \). Moreover this integral satisfies:

1. \( \left\| \int_{\Omega} f d\mu \right\|_{Y} \leq \int_{\Omega} \|f(\omega)\|_{Y} d\mu(\omega) \leq \|f\|_{\infty} \mu(\Omega) \) for all \( f \in \mathfrak{S}(A,Y) \),
2. If \( A, B \in \mathcal{A} \) and \( A \cap B = \emptyset \), then
\[
\int_{A \cup B} f d\mu = \int_{A} f d\mu + \int_{B} f d\mu.
\]
3. If \( X \) is another Banach space and \( T : X \to Y \) is a continuous linear map, then
\[
T \int_{\Omega} f d\mu = \int_{\Omega} T f d\mu.
\]
4. If \( Y = \mathbb{R} \), \( f, g \in \mathfrak{S}(A,\mathbb{R}) \), and \( f(\omega) \geq g(\omega) \) for all \( \omega \in \Omega \) then \( \int_{\Omega} f d\mu \geq \int_{\Omega} g d\mu \).

Proof. Everything but items 2. – 4. have already been proved. Item 2. is easy since
\[
\mathbf{1}_{A \cup B} f = \mathbf{1}_{A} f + \mathbf{1}_{B} f
\]
and so the result follows by linearity of the integral.
3. If \( f = \sum_{i=1}^{n} y_{i} \mathbf{1}_{A_{i}} \), with \( y_{i} \in Y \) and \( A_{i} \in \mathcal{A} \), then \( Tf = \sum_{i=1}^{n} T y_{i} \mathbf{1}_{A_{i}} \) and
\[
\int_{\Omega} T f d\mu = \sum_{i=1}^{n} T y_{i} \mu(A_{i}) = T \left[ \sum_{i=1}^{n} y_{i} \mu(A_{i}) \right] = T \int_{\Omega} f d\mu.
\]
If \( f \in \mathfrak{S}(A,Y) \) and \( f_n \in \mathfrak{S}(A,Y) \) with \( \|f - f_n\|_{\infty} \to 0 \), then
\[
T \int_{\Omega} f d\mu = T \lim_{n \to \infty} \int_{\Omega} f_n d\mu = \lim_{n \to \infty} \int_{\Omega} f_n d\mu \quad \text{[} T \text{ is continuous]}.
\]

Now from Exercise 7.12 there exists \( C < \infty \) such that \( \|T y\|_{X} \leq C \|y\|_{Y} \) for all \( y \in Y \) and therefore,
\[
\|T f - T f_n\|_{\infty} = \sup_{\omega} \|T f(\omega) - T f_n(\omega)\|_{Y} \leq C \sup_{\omega} \|f(\omega) - f_n(\omega)\|_{Y} = C \|f - f_n\|_{\infty}
\]
and hence \( T f_n \to T f \) uniformly and hence \( T f \in \mathfrak{S}(A,X) \) and
\[
T \int_{\Omega} f d\mu = \lim_{n \to \infty} \int_{\Omega} T f_n d\mu = \int_{\Omega} T f d\mu.
\]

Exercise 12.3. Let \( f \in \mathfrak{S}(A,\mathbb{R}) \) and \( y \in Y \) where \( Y \) is a Banach space. Show
\[
\int_{\Omega} y f d\mu = y \int_{\Omega} f d\mu.
\]

Hint: this can be proved by from first principles but it would be easier to use Theorem 12.4 for a well chosen continuous linear transformation, \( T : F \to Y \).

Exercise 12.4 (Interchanging sums with integrals). Suppose that \( \{f_n\}_{n=1}^{\infty} \subset \mathfrak{S}(A,Y) \) and \( \sum_{n=1}^{\infty} \|f_n\|_{\infty} < \infty \). Show \( \sum_{n=1}^{\infty} f_n \) is in \( \mathfrak{S}(A,Y) \) and
\[
\int_{\Omega} \left( \sum_{n=1}^{\infty} f_n \right) d\mu = \sum_{n=1}^{\infty} \int_{\Omega} f_n d\mu
\]
and moreover
\[ \left\| \int_{\Omega} \left( \sum_{n=1}^{\infty} f_{n} \right) d\mu \right\| \leq \sum_{n=1}^{\infty} \int_{\Omega} \| f_{n} \| d\mu \leq \mu(\Omega) \sum_{n=1}^{\infty} \| f_{n} \|_{\infty}. \]

**Exercise 12.5.** Suppose that \( \rho \in \bar{S}(A, \mathbb{R}) \) with \( \rho \geq 0 \).

1. Show \( \nu(A) := \int_{A} \rho d\mu \) defines a finitely additive measure on \( A \).
2. If \( f \in \bar{S}(A, Y) \), show,
\[ \int_{\Omega} f d\nu = \int_{\Omega} f \rho d\mu. \quad (12.3) \]

**Hint:** first verify Eq. (12.3) when \( f = y \cdot 1_{A} \) for some \( y \in Y \) and \( A \in \mathcal{A} \).

At this stage you should make use of Exercise 12.3. Then verify Eq. (12.3) for \( f \in \bar{S}(A, Y) \) and then go on to the general case.

Let us now specialize to the case where \( \Omega := (a, b), \alpha : [a, b] \to \mathbb{R} \) is an increasing function, \( A \) is the algebra of half open intervals on \( \Omega \), and \( \mu : A \to \mathbb{R}_{+} \) is the unique measure such that \( \mu((s, t]) = \alpha(t) - \alpha(s) \) for all \( a \leq s \leq t \leq b \).

**Theorem 12.5.** Suppose that \( \Omega = (a, b) \) and let \( C([a, b], Y) \) be the space of continuous functions from \([a, b] \to Y \). Then \( C([a, b], Y) \subset \bar{S}(A, Y) \). (More precisely, if \( f \in C([a, b], Y) \) then \( f_{(a, b]} \in \bar{S}(A, Y) \). Further suppose that \( f \in C([a, b], Y) \) and to each partition \( \Pi := \{a = t_{0} < t_{1} < \cdots < t_{n} = b\} \) of \([a, b]\) we have chosen \( c_{i}^{\Pi} \in [t_{i-1}, t_{i}] \) for \( 1 \leq i \leq n \). Then
\[ \int_{a}^{b} f d\mu = \lim_{|\Pi| \to 0} \sum_{i=1}^{n} f(c_{i}^{\Pi})(\alpha(t_{i}) - \alpha(t_{i-1})). \]

In particular if \( \alpha(t) = t \), then we write \( dt \) for \( d\mu \) or \( d\alpha \) and we have,
\[ \int_{a}^{b} f(t) dt = \lim_{|\Pi| \to 0} \sum_{i=1}^{n} f(c_{i}^{\Pi})(t_{i} - t_{i-1}). \]

**Proof.** From Exercise 11.4, we know that \( f \) may be written as a uniform limit of step functions and hence is in \( \bar{S}(A, Y) \). Moreover that same exercise shows; if \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that
\[ \| f - f_{H, \varepsilon} \|_{u} \leq \varepsilon \text{ if } |\Pi| \leq \delta. \]

Therefore if \( |\Pi| \leq \delta \), then
\[ \left\| \int_{a}^{b} f d\mu - \int_{a}^{b} f_{H, \varepsilon} d\mu \right\| \leq \| f - f_{H, \varepsilon} \|_{u} \cdot \mu(\Omega) \leq \varepsilon \cdot \mu(\Omega) \]
and this suffices to show
\[ \lim_{|\Pi| \to 0} \int_{a}^{b} f_{H, \varepsilon} d\mu = \int_{a}^{b} f d\mu. \]

This completes the proof since,
\[ f_{H, \varepsilon} = \sum_{i=1}^{n} f(c_{i}) 1_{(t_{i-1}, t_{i}]} \text{ on } \Omega = (a, b) \]

and hence
\[ \int_{a}^{b} f_{H, \varepsilon} d\mu = \sum_{i=1}^{n} f(c_{i}) \mu((t_{i-1}, t_{i}]) = \sum_{i=1}^{n} f(c_{i}^{\Pi})(\alpha(t_{i}) - \alpha(t_{i-1})). \]

If \( f_{n} \in S \) and \( f \in \bar{S} \) such that \( \lim_{n \to \infty} \| f - f_{n} \|_{\infty} = 0 \), then for \( a \leq u < v \leq b \), then \( 1_{(u, v)} f_{n} \in S \) and \( \lim_{n \to \infty} \| 1_{(u, v)} f - 1_{(u, v)} f_{n} \|_{\infty} = 0 \). This shows \( 1_{(u, v)} f \in \bar{S} \) whenever \( f \in \bar{S} \).

**Notation 12.6.** For \( u, v \in [a, b] \) and \( f \in \bar{S}(A, Y) \) we let
\[ \int_{u}^{v} f d\mu := \begin{cases} \int_{\Omega} 1_{(u, v]} f d\mu & \text{if } u \leq v, \\ -\int_{\Omega} 1_{[v, w]} f d\mu & \text{if } u \geq v. \end{cases} \]

The next lemma explains there reason for the convention in Notation 12.6.

**Lemma 12.7.** For all \( u, v, w \in [a, b] \) and \( f \in \bar{S}(A, Y) \) we have
\[ \int_{u}^{w} f d\mu = \int_{u}^{v} f d\mu + \int_{v}^{w} f d\mu. \quad (12.4) \]

**Proof.** If \( u \leq v \leq w \), then the formula in Eq. (12.4) is a consequence of the linearity of the integral and the fact that
\[ 1_{(u, w]} = 1_{(u, v]} + 1_{(v, w]} \]
All other cases may be reduced to this case. For example if \( w \leq u \leq v \), then we know
\[ \int_{w}^{u} f d\mu = \int_{w}^{v} f d\mu + \int_{v}^{u} f d\mu = -\int_{u}^{w} f d\mu + \int_{u}^{v} f d\mu. \]
Solving this equation for \( \int_{w}^{u} f d\mu \) then gives
\[ \int_{u}^{w} f d\mu = \int_{u}^{v} f d\mu - \int_{v}^{u} f d\mu = \int_{u}^{v} f d\mu + \int_{w}^{v} f d\mu. \]
Similar arguments work for the 4 other orderings of \( u, v, w \) and are left to the reader. (That is you should verify Eq. (12.4) in the following cases; 1) \( u \leq w \leq v \), 2) \( v \leq u \leq w \), 3) \( v \leq w \leq u \), and 4) \( w \leq v \leq u \).
Lemma 12.8. Suppose that \( u, v \in [a, b] \) and \( f \in \tilde{S}(A, Y) \), then

\[
\left\| \int_u^v f \, d\mu \right\| \leq \int_u^v \| f \| \, d\mu \leq \| f \|_\infty \mu(J)
\]

where

\[ J = \begin{cases} (u, v) & \text{if } u < v \\ (v, u) & \text{if } v \leq u. \end{cases} \]

Proof. By definition we have

\[
\left\| \int_u^v f \, d\mu \right\| = \left| \int_J f \, d\mu \right| = \left\| \int_{(a,b]} 1_J f \, d\mu \right\|.
\]

Therefore,

\[
\left\| \int_u^v f \, d\mu \right\| \leq \int_{(a,b]} \| 1_J f \| \, d\mu = \| 1_J f \|_\infty \cdot \mu(J)
\]

which combined with the identity,

\[
\int_u^v \| f \| \, d\mu = \left| \int_J f \, d\mu \right| = \int_{(a,b]} \| 1_J f \| \, d\mu = \int_J \| f \| \, d\mu,
\]

proves the result. \( \Box \)

The next lemma lists a number of the many familiar properties of the integral and really is just a special case of Theorem 12.4.

Lemma 12.9. For \( f \in \tilde{S}((a,b], X) \) and \( u, v, w \in [a, b] \), the integral satisfies:

1. \( \| \int_u^w f(t) \, dt \|_X \leq (v - u) \sup \{ \| f(t) \| : u \leq t \leq v \} \).
2. \( \int_u^w f(t) \, dt = \int_u^v f(t) \, dt + \int_v^w f(t) \, dt \).
3. If \( Y \) is another Banach space and \( T : X \to Y \) be a continuous linear operator\(^1\) then \( T f \in \tilde{S}((a,b], Y) \) and

\[
T \left( \int_u^v f(t) \, dt \right) = \int_u^v T f(t) \, dt.
\]

4. The function \( t \to \| f(t) \|_X \) is in \( \tilde{S}((a,b], \mathbb{R}) \) and

\[
\left\| \int_u^v f(t) \, dt \right\|_X \leq \int_u^v \| f(t) \|_X \, dt.
\]

5. If \( f, g \in \tilde{S}([a,b], \mathbb{R}) \) and \( f \leq g \), then

\[
\int_a^b f(t) \, dt \leq \int_a^b g(t) \, dt.
\]

Analogous statements also hold if \( dt \) is replaced by \( d\mu \) where \( \mu \) is any finitely additive measure from \( A_{[a,b]} \to \mathbb{R}^+ \). The only modification is in item 1. where we would now have,

\[
\left\| \int_u^v f(t) \, dt \right\|_X \leq \mu((u, v]) \cdot \sup \{ \| f(t) \| : u \leq t \leq v \}.
\]

Proof.

1. \( \| \int_u^w f(t) \, dt \|_\infty \leq (v - u) \sup \{ \| f(t) \| : u \leq t \leq v \} \) is easily seen to hold for \( f \in \mathcal{S} \). The general statement then holds by passing to the limit.

2. The function \( G(t) := \int_t^w f(\tau) \, d\tau \) is continuous on \( [a,b] \), indeed,

\[
\| G(t) - G(\tau) \| = \left\| \int_\tau^t f(\tau) \, d\tau \right\| \leq \| f \|_\infty |t - \tau|.
\]

3. If \( T \in L(X,Y) \) and \( f \in \mathcal{S}((a,b], X) \) then \( T f \in \mathcal{S}((a,b], Y) \) and because \( T \) commutes with finite sums,

\[
T \left( \int_u^v f(t) \, dt \right) = \int_u^v T f(t) \, dt.
\]

For general \( f \in \tilde{S}((a,b], X) \), there exists \( f_n \in \mathcal{S}((a,b], X) \) such that \( \| f - f_n \|_\infty \to 0 \). Because \( T \) is bounded it follows that

\[
\| T f - T f_n \|_\infty \leq \| T \| \cdot \| f - f_n \|_\infty \to 0.
\]

which shows \( T f \in \tilde{S}((a,b], Y) \). Moreover we have

\[
T \left( \int_u^v f(t) \, dt \right) = T \lim_{n \to \infty} \int_u^v f_n(t) \, dt = \lim_{n \to \infty} T \int_u^v f_n(t) \, dt
\]

\[
= \lim_{n \to \infty} \int_u^v T f_n(t) \, dt = \int_u^v T f(t) \, dt.
\]

4. The function \( t \to \| f(t) \|_X \) is in \( \mathcal{S}((a,b], \mathbb{R}) \) for \( f \in \mathcal{S}((a,b], X) \) and by the triangle inequality for finite sums,

\[
\left\| \int_a^v f(t) \, dt \right\| \leq \int_a^v \| f(t) \| \, dt.
\] (12.5)
For general \( f \in \tilde{S}((a, b], X) \), choose \( f_n \in \mathcal{S}((a, b], X) \) such that \( \|f - f_n\|_\infty \to 0 \) as above. Since
\[
\sup_i ||f(t) - f_n(t)|| \leq \sup_i \|f(t) - f_n(t)\| = \|f - f_n\|_\infty \to 0 \quad \text{as } n \to \infty
\]
it follows that \( t \to \|f(t)\|_X \) is in \( \tilde{S}((a, b], \mathbb{R}) \). A similar limiting argument to that in item 4. shows that Eq. (12.5) holds for \( f \in \tilde{S}((a, b], X) \).

5. It suffices to show if \( f \in \tilde{S}((a, b], \mathbb{R}) \) with \( f \geq 0 \) then
\[
0 \leq \int_a^b f(t)dt.
\]
This is certainly true if \( f \in \tilde{S}((a, b], \mathbb{R}) \). Again for general \( f \in \tilde{S}((a, b], \mathbb{R}) \) with \( f \geq 0 \), choose \( f_n \in \tilde{S}((a, b], \mathbb{R}) \) such that \( \|f - f_n\|_\infty \to 0 \). By replacing \( f_n \) with \( \max(f_n, 0) \in \tilde{S}((a, b], \mathbb{R}) \) we may assume \( f_n \geq 0 \) for all \( n \). So passing to the limit in the inequality,
\[
0 \leq \int_a^b f_n(t)dt
\]
leads to the desired conclusion.

**Exercise 12.6.** Show \( f = (f_1, \ldots, f_n) \in \tilde{S}((a, b], \mathbb{R}^n) \) iff \( f_i \in \tilde{S}((a, b], \mathbb{R}) \) for \( i = 1, 2, \ldots, n \)
\[
\int_a^b f(t)dt = \left( \int_a^b f_1(t)dt, \ldots, \int_a^b f_n(t)dt \right). \tag{12.6}
\]
Here \( \mathbb{R}^n \) is to be equipped with the usual Euclidean norm. **Hint:** Use item 4. of Lemma 12.9 with judiciously chosen \( T : \mathbb{R}^n \to \mathbb{R} \) for point one direction. For the opposite direction, use Exercise 12.1 and the observation that \( f(t) = \sum_{i=1}^n f_i(t) e_i \) where \( \{e_i\}_{i=1}^n \) is the standard basis for \( \mathbb{R}^n \).

**Exercise 12.7.** Let \( 0 < a < T < \infty \) and \( \lambda > 0 \) be given and let \( \alpha : [0, T] \to \mathbb{R}_+ \) be defined by
\[
\alpha(x) = \lambda \cdot 1_{x \geq a} = \begin{cases} 
\lambda & \text{if } x \geq a \\
0 & \text{if } x < a.
\end{cases}
\]
Further let \( \mu_\alpha \) be the unique finitely additive measure on \( \mathcal{A}([0, T]) \) such that \( \mu_\alpha((s, t]) = \alpha(t) - \alpha(s) \) for all \( 0 \leq s \leq t \leq T \). Show
\[
\int_{[0, T]} fd\alpha = \int_{[0, T]} f d\mu_\alpha = \lambda f(a) \quad \text{for all } f \in C([0, T], \mathbb{C}).
\]

**Exercise 12.8.** Let \( 0 < a < T < \infty \) and \( \lambda > 0 \) be given and let \( \beta : [0, T] \to \mathbb{R}_+ \) be defined by
\[
\beta(x) = \lambda \cdot 1_{x > a} = \begin{cases} 
\lambda & \text{if } x > a \\
0 & \text{if } x \leq a.
\end{cases}
\]
Further let \( \mu_\beta \) be the unique finitely additive measure on \( \mathcal{A}([0, T]) \) such that \( \mu_\beta((s, t]) = \beta(t) - \beta(s) \) for all \( 0 \leq s \leq t \leq T \). Show
\[
\int_{[0, T]} f d\beta = \int_{[0, T]} f d\mu_\beta = \lambda f(a) \quad \text{for all } f \in C([0, T], \mathbb{C}).
\]
[Thus it is possible that]
\[
\int_{[0, T]} f d\alpha = \int_{[0, T]} f d\beta \quad \text{for all } f \in C([0, T], \mathbb{C})
\]
even though \( \alpha \neq \beta \). This point may be addressed more in the third quarter.]

**Remark 12.10.** If \( \alpha, \beta : [0, T] \to \mathbb{R} \) are two increasing functions, it is easy to verify that
\[
\mu_{\alpha + \beta} = \mu_\alpha + \mu_\beta
\]
and
\[
\int_0^T f d(\alpha + \beta) = \int_0^T f d\alpha + \int_0^T f d\beta
\]
for all \( f \in \tilde{S}(\mathcal{A}([0, T]), Y) \).

**Corollary 12.11.** If \( \alpha : [0, T] \to \mathbb{R} \) is define by
\[
\alpha(x) := \sum_{i=1}^n \lambda_i 1_{x \geq a_i},
\]
where \( 0 < a_1 < a_2 < \cdots < a_n < T \) and \( \lambda_i > 0 \) for \( 1 \leq i \leq n \), then
\[
\int_0^T f d\alpha = \sum_{i=1}^n \lambda_i f(a_i) \quad \text{for all } f \in C([0, T], \mathbb{C}).
\]
Moreover if \( \beta(x) = \alpha(x) + x \), then for all \( f \in C([0, T], \mathbb{C}) \),
\[
\int_0^T f d\beta = \sum_{i=1}^n \lambda_i f(a_i) + \int_0^T f(x) dx.
\]
This shows that our integral may be a combination of the usual Riemann integral along with a finite sum.
12.1 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. The next proposition follows directly from Theorem 10.19. Since we have not proven this theorem in full generality, I will supply a proof of the proposition which is all we really need for our purposes.

Proposition 12.12 (A Crude Mean Value Inequality). Suppose that \(-\infty < a < b < \infty\) and \(f : [a, b] \to X\) is a continuous function such that \(f(t)\) exists for all \(t \in (a, b)\). If \(M := \sup_{a < t < b} \|f'(t)\| < \infty\), then

\[
\|f(\beta) - f(\alpha)\| \leq M(\beta - \alpha) \quad \text{for all } a \leq \alpha \leq \beta \leq b.
\]

In particular if \(f'(t) = 0\) for all \(t \in (a, b)\), then \(f\) is constant.

Proof. For every \(\varepsilon > 0\) and \(\tau \in (a, b)\) there exists (by definition of the derivative) a \(\delta_\tau > 0\) such that

\[
\|f(t) - f(\tau) - f'(\tau)(t - \tau)\| \leq \varepsilon |t - \tau| \quad \text{if } |t - \tau| \leq \delta_\tau.
\]

This in equality along with the triangle inequality implies for \(|t - \tau| \leq \delta_\tau\) with \(t \in (a, b)\) that

\[
\|f(t) - f(\tau)\| \leq \|f'(\tau)(t - \tau)\| + \varepsilon |t - \tau| \leq (M + \varepsilon)|t - \tau|.
\]

Now suppose that \(\alpha \in (a, b)\) and \(\varepsilon > 0\) is given. (We will later let \(\varepsilon \downarrow 0\).) Let

\[
S := \{t \in [a, b] : \|f(t) - f(\alpha)\| > (M + \varepsilon)(t - \alpha)\}.
\]

From Eq. (12.8) with \(\tau = \alpha\), we see that \([a, \alpha + \delta_\alpha] \cap S = \emptyset\). For sake of contradiction suppose that \(S\) is not empty. In this case let \(c := \inf S \in [a + \delta_\alpha, \beta]\). Then for all \(t \in [a, c]\) we have

\[
\|f(t) - f(\alpha)\| \leq (M + \varepsilon)(t - \alpha).
\]

By continuity of \(f\) and the norm, the previous estimate remains true for \(t = c\) as well. However, using Eq. (12.8) with \(\tau = c\) and Eq. (12.9) at \(t = c\), we find for \(0 \leq h \leq \delta_\varepsilon\) that

\[
\|f(c + h) - f(\alpha)\| \leq \|f(c + h) - f(c)\| + \|f(c) - f(\alpha)\|
\leq (M + \varepsilon)h + (M + \varepsilon)(c - \alpha) = (M + \varepsilon)(c + h - \alpha).
\]

This however then implies that \(c = \inf S > c\) which is a contradiction and we conclude that \(S = \emptyset\).

The previous argument has shown

\[
\|f(t) - f(\alpha)\| \leq (M + \varepsilon)(t - \alpha) \quad \text{for all } \alpha \leq t \leq b.
\]

We may now let \(\varepsilon \downarrow 0\) in order to conclude that

\[
\|f(t) - f(\alpha)\| \leq M(t - \alpha) \quad \text{for all } a < \alpha \leq t \leq b.
\]

By continuity this inequality also holds for \(\alpha = a\). Finally if \(M = 0\) we have shown \(\|f(t) - f(\alpha)\| = 0\) for all \(a \leq t \leq b\); i.e. \(f(t) = f(a)\) for all \(t \in [a, b]\).

Alternate Proof. To simplify notation let us now write \(f(\partial [u, v]) := f(v) - f(u)\) and \([u, v] := v - u\) for all \(a \leq u < v \leq b\). Suppose for sake of contradiction that there exits \(a \leq \alpha < \beta \leq b\) such that

\[
\|f(\partial [\alpha, \beta])\| = \|f(\beta) - f(\alpha)\| > M(\beta - \alpha) = M[\alpha, \beta].
\]

By continuity of \(f\) we may move \(\alpha\) and \(\beta\) if necessary in order to further assume \(a < \alpha < \beta < b\). We now choose \(M' > M\) so that

\[
\|f(\partial [\alpha, \beta])\| > M'[\alpha, \beta] > M[\alpha, \beta].
\]

Let \(J_0 = [\alpha, \beta]\) and \(t = \frac{1}{2}(\alpha + \beta)\) be the mid-point of \(J_0\). Since

\[
M'(\beta - t) + M'(t - \alpha) = M'(\beta - \alpha) \quad < \|f(\beta) - f(\alpha)\| \leq \|f(\beta) - f(t)\| + \|f(t) - f(\alpha)\|
\]

we must have either

\[
M'[t, \beta] < \|f(\partial [t, \beta])\| \quad \text{or} \quad M'[\alpha, t] < \|f(\partial [\alpha, t])\|.
\]

If the first case holds we let \(J_1 = [\alpha, t]\) and if only the second case holds we take \(J_1 = [t, \alpha]\). We may now work similarly and split \(J_1\) in half an choose one one of the two sub-intervals in \(J_2\) so that \(\|f(\partial J_2)\| > M'[J_2]\). Continuing this way inductively we construct, \(J_1 \subseteq J_0\) so that \(J_0 \supset J_1 \supset J_2 \supset \ldots\) and \(\|f(\partial J_i)\| > M'[J_i]\) for all \(i\) and diam \((J_i) = 2^{-i}(\beta - \alpha)\). As the sets \(J_i\) are closed it follows that \(\cap J_i = \{\tau\}\) for some point \(\tau \in J_0\).

Let \(\varepsilon := M' - M\) and use the definition of derivative of \(f\) at \(\tau\) to find a \(\delta > 0\) such that

\[
\|f(t) - f(\tau) - f'(\tau)(t - \tau)\| \leq \varepsilon |t - \tau| \quad \text{if } |t - \tau| \leq \delta.
\]

From this estimate we learn that

\[
\|f(t) - f(\tau)\| \leq \left(\|f'(\tau)\| + \varepsilon\right)|t - \tau| = M'|t - \tau|.
\]

\footnote{This proof is reminiscent of Morera’s theorem for triangles from complex variables.}
Choosing \( n \) large enough so that \( \tau \in J_n = [\alpha_n, \beta_n] \subset [\tau - \delta, \tau + \delta] \), we find,

\[
M' \, |J_n| < \|f(\partial J_n)\| \leq \|f(b_n) - f(t)\| + \|f(t) - f(\alpha_n)\| \\
\leq M'(\beta_n - \tau + \tau - \alpha_n) = M'|J_n|
\]

which leads to the contradiction that \( M'|J_n| < M'|J_n| \).

**Definition 12.13.** Given a function \( f \in C([a,b], X) \) we say that \( F \in C([a,b], X) \) is an anti-derivative of \( f \) if \( F(t) = f(t) \) for all \( t \in (a,b) \).

**Lemma 12.14.** For \( f \in C([a,b], X) \), the function\( G(t) := \int_a^t f(\tau) \, d\tau \) is Lipschitz continuous in \( t \in [a,b] \) and therefore,

\[
\|G(t) - G(s)\|_X \leq \int_s^t \|f(\tau)\| \, d\tau \leq \|f\|_\infty |t-s| \quad \text{for all } s, t \in [a,b].
\]

**Proof.** Suppose that \( a \leq s \leq t \leq b \). Then

\[
G(t) - G(s) = \int_a^t f(\tau) \, d\tau - \int_a^s f(\tau) \, d\tau = \int_s^t f(\tau) \, d\tau
\]

and therefore,

\[
\|G(t) - G(s)\|_X \leq \int_s^t \|f(\tau)\| \, d\tau \leq \|f\|_\infty |t-s| .
\]

The following elementary but important lemma and its relatives will frequently be used throughout the rest of these notes.

**Lemma 12.15.** For all \( x \in X \) and \( a, b \in \mathbb{R} \) we have,

\[
\int_a^b x \, d\tau = x \cdot (b-a) . \tag{12.10}
\]

**Proof.** Proof if \( a \leq b \) then Eq. (12.10) holds by definition of the integral on simple functions. If \( b < a \), then

\[
\int_a^b x \, d\tau = -\int_b^a x \, d\tau = -x \cdot (a-b) = x \cdot (b-a) .
\]

**Theorem 12.16 (Fundamental Theorem of Calculus).** Suppose that \( f \in C([a,b], X) \), Then

\[\text{Note that } G(a) = \int_{[a,a]} f \, dm = 0.\]

1. \( \frac{d}{dt} \int_a^t f(\tau) \, d\tau = f(t) \) for all \( t \in (a,b) \), i.e. \( G(t) := \int_a^t f(\tau) \, d\tau \) is an anti-derivative of \( f \).

2. If \( F \in C([a,b], X) \) is an anti-derivative of \( f \), then

\[
\int_a^b f(t) \, dt = F(b) - F(a).
\]

**Proof.** 1. By Lemma 12.14 the function \( G(t) := \int_a^t f(\tau) \, d\tau \) is continuous for \( t \in [a,b] \). Now suppose that \( t \in (a,b) \) and \( h \in \mathbb{R} \setminus \{0\} \) is small so that \( t + h \in (a,b) \). Let

\[
J_h := \begin{cases} 
(t, t + h) & \text{if } h > 0 \\
(t + h, t) & \text{if } h < 0
\end{cases}
\]

Applying Lemma 12.15 with \( x = f(t) \), \( a = t \) and \( b = t + h \) implies,

\[
G(t + h) - G(t) = \int_t^{t+h} f(\tau) \, d\tau - \int_t^t f(\tau) \, d\tau = \int_t^{t+h} f(\tau) \, d\tau - f(t)h
\]

\[
= \int_t^{t+h} f(\tau) \, d\tau - f(t)h
\]

\[
= \int_t^{t+h} [f(\tau) - f(t)] \, d\tau.
\]

Taking norms of this equation along with Lemma 12.18 gives,

\[
\|G(t + h) - G(t) - f(t)h\|_X = \left\| \int_t^{t+h} [f(\tau) - f(t)] \, d\tau \right\|_X
\]

\[
\leq \int_{J_h} \|f(\tau) - f(t)\|_X \, d\tau \leq \varepsilon(h) |h|
\]

where

\[
\varepsilon(h) := \max \|f(\tau) - f(t)\|_X \leq \max_{\tau \in [t-|h|, t+|h|]} \|f(\tau) - f(t)\|_X.
\]

The continuity of \( f \) at \( t \) implies \( \lim_{h \to 0} \varepsilon(h) = 0 \) and so by the definition of the derivative we have shown, \( G(t) = f(t) \) exists for all \( t \in (a,b) \).

2. For the second item, let \( h(t) := G(t) - F(t) \). Then \( h \in C([a,b], X) \) with

\[
h(t) = \dot{G}(t) - \dot{F}(t) = f(t) - f(t) = 0 \quad \text{for all } t \in (a,b).
\]

Hence by the mean value inequality in any one its guises (see Theorem 10.19 or Proposition 12.12) we know that \( h(t) \) is constant. In particular we may conclude that

\[
G(b) - F(b) = h(b) = h(a) = G(a) - F(a) = -F(a).
\]
\[ \int_a^b f(\tau) \, d\tau = G(b) - G(a). \]

**Example 12.17.** For all \( a, b \in \mathbb{R} \) and \( z \in \mathbb{C} \setminus \{0\} \),
\[ \int_a^b e^{zt} \, dt = \frac{1}{z} e^{zt} \bigg|_{\tau=b}^{\tau=a} = \frac{1}{z} \left[ e^{zb} - e^{za} \right]. \]  

(12.11)

and in particular,
\[ \int_0^b e^{zt} \, dt = \frac{1}{z} \left[ e^{zb} - 1 \right]. \]

Let us now write \( z = x + iy \) so that
\[ e^{zt} = e^{x \tau + iy \tau} = e^{x \tau} \cos(y \tau) + ie^{x \tau} \sin(y \tau). \]

Then according to Exercise 12.9 and Eq. (12.11) we may conclude
\[ \int_0^b e^{zt} \, dt = \int_a^b e^{z \tau} \cos(y \tau) \, d\tau + i \int_a^b e^{z \tau} \sin(y \tau) \, d\tau. \]

On the other hand,
\[ \frac{1}{z} \left[ e^{zb} - 1 \right] = \frac{1}{x^2 + y^2} \left( x - iy \right) \left( e^{xb} \cos(yb) + ie^{xb} \sin(yb) - 1 \right). \]

So comparing real and imaginary parts of these two equations implies,
\[ \int_0^b e^{z \tau} \cos(y \tau) \, d\tau = \frac{1}{x^2 + y^2} \left[ e^{xb} (x \cos(yb) + y \sin(yb)) - x \right] \]

and
\[ \int_0^b e^{z \tau} \sin(y \tau) \, d\tau = \frac{1}{x^2 + y^2} \left[ e^{zb} (-y \cos(yb) + x \sin(yb)) + y \right]. \]

**Exercise 12.9 (Integrating Power Series).** Suppose that \( f(t) := \sum_{n=0}^{\infty} c_n t^n \) has radius of convergence \( R > 0 \). Show for any \(-R < a < b < R\) that
\[ \int_a^b \sum_{n=0}^{\infty} c_n t^n \, dt = \sum_{n=0}^{\infty} c_n \int_a^b t^n \, dt = \sum_{n=0}^{\infty} c_n \left[ \frac{b^{n+1} - a^{n+1}}{n+1} \right] \]

**Hint:** use Exercise 12.4 above.

**Example 12.18.** By the fundamental theorem of calculus along with Exercise 12.9 we find for \(|x| < 1\) that
\[ \ln(1 + x) = \int_0^x \frac{1}{1 + t} \, dt = \int_0^x \sum_{n=0}^{\infty} (-1)^n t^n \, dt \]
\[ = \sum_{n=0}^{\infty} (-1)^n \int_0^x t^n \, dt = \sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{n+1} \]

and
\[ \tan^{-1}(x) = \int_0^x \frac{1}{1 + t^2} \, dt = \int_0^x \sum_{n=0}^{\infty} (-1)^n t^{2n} \, dt \]
\[ = \sum_{n=0}^{\infty} (-1)^n \int_0^x t^{2n} \, dt = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1}. \]

**Exercise 12.10.** Suppose \( \alpha : (a - \varepsilon, b + \varepsilon) \to \mathbb{R} \) is a non-decreasing \( C^1 \) function. Show
\[ \int_a^b f(\alpha(t)) \, dt = \int_a^b f(t) \, d\alpha(t) \] for all \( f \in \mathbb{R} \).

**Hint:** Let \( \mu_\alpha \) be the unique measure on \( A \) such that \( \mu_\alpha((s, t]) = \alpha(t) - \alpha(s) \) for all \( s \leq a \leq t \leq b \) and first show \( \mu_\alpha(A) = \int_A \alpha(\tau) \, d\tau \) for all \( A \in \mathcal{A} \). Now apply Exercise 12.5.

**Remark 12.19.** Suppose \( f : [a, b] \to Y \) is continuous in Exercise 12.10 and let \( \Pi = \{a = t_0 < t_1 < \cdots < t_n = b\} \) be a partition of \([a, b]\). By the mean value theorem, there exists \( c_i \in (t_{i-1}, t_i) \) such that \( \alpha(t_i) - \alpha(t_{i-1}) = \alpha(c_i)(t_i - t_{i-1}) \). Therefore,
\[ \sum_{i=1}^{n} f(c_i) [\alpha(t_i) - \alpha(t_{i-1})] = \sum_{i=1}^{n} f(c_i) \alpha(c_i) (t_i - t_{i-1}) = \sum_{i=1}^{n} f(c_i) \alpha(c_i) (t_i - t_{i-1}). \]

(12.12)

We may now use Theorem 12.5 twice to conclude that
\[ \lim_{|\Pi| \to 0} \sum_{i=1}^{n} f(c_i) [\alpha(t_i) - \alpha(t_{i-1})] = \int_a^b f(\alpha) \, d\alpha \] and
\[ \lim_{|\Pi| \to 0} \sum_{i=1}^{n} f(c_i) \alpha(c_i) (t_i - t_{i-1}) = \int_a^b f(t) \, d\alpha(t) \]

which combined with Eq. (12.12) shows
\[ \int_a^b f(\alpha) \, d\alpha = \int_a^b f(t) \, d\alpha(t). \]
Remark 12.20. Suppose that \( \alpha : [a,b] \to \mathbb{R} \) is an non-decreasing function such that there a finite subset, \( \Pi \subseteq (a,b) \), such that \( \alpha = C^1 \) on \( (a,b) \setminus \Pi \) with 
\[ \hat{\alpha}'|_{(a,b) \setminus \Pi} \]
 extending to a continuous function on \([a,b]\). Then combining the ideas in Remark 12.19 along with those in Corollary 12.11, for every \( f : [a,b] \to \mathbb{R} \) which is continuous we have
\[
\int_a^b f \, d\alpha = \int_a^b f(t) \, \dot{\alpha}(t) \, dt + \sum_{t \in \Pi} f(t) [\alpha(t+) - \alpha(t-)],
\]  
(12.13)
where \( \alpha(t+) = \lim_{\tau \to t^+} \alpha(\tau) \) and \( \alpha(t-) = \lim_{\tau \to t^-} \alpha(\tau) \). However, it should be noted that there exists continuous strictly increasing functions \( \hat{\alpha} : [a,b] \to \mathbb{R} \) such that \( \dot{\alpha}(t) = 0 \) “almost everywhere!” For such a function, \( \alpha \), Eq. (12.13) is not longer valid!

Corollary 12.21 (Another Mean Value Inequality). Suppose that \( f : [a,b] \to X \) is a continuous function such that \( f(t) \) exists for \( t \in (a,b) \) and \( f \) extends to a continuous function on \([a,b]\). Then
\[
\|f(b) - f(a)\| \leq \int_a^b \|\dot{f}(t)\| \, dt \leq (b-a) \cdot \|f\|_\infty.
\]  
(12.14)

Proof. By the fundamental theorem of calculus, \( f(b) - f(a) = \int_a^b f'(t) \, dt \) and then by Lemma 12.9
\[
\|f(b) - f(a)\| = \left\| \int_a^b \dot{f}(t) \, dt \right\| \leq \int_a^b \|\dot{f}(t)\| \, dt
\]
\[
\leq \int_a^b \|\dot{f}\|_\infty \, dt = (b-a) \cdot \|f\|_\infty.
\]

Corollary 12.22 (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 \( t \in (a,b) \) 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,
\footnote{Look up Cantor functions.}
\[
\frac{d}{ds} F(T(s)) = F'(T(s)) T'(s) = f(T(s)) T'(s).
\]
Integrating this equation on \( s \in [c,d] \) and using the chain rule again gives
\[
\int_c^d f(T(s)) T'(s) \, ds = F(T(d)) - F(T(c)) = \int_{T(c)}^{T(d)} f(t) \, dt.
\]

Corollary 12.23 (Integration by Parts). Suppose that \( g : [a,b] \to X \) and \( f : [a,b] \to \mathbb{R} \) are continuous functions such that \( \dot{f}(t) \) and \( \dot{g}(t) \) exists for \( t \in (a,b) \) and \( \dot{f} \) and \( \dot{g} \) extend to continuous function on \([a,b]\). Then
\[
\int_a^b f(t) \dot{g}(t) \, dt + \int_a^b \dot{f}(t) g(t) \, dt = f(t) g(t)|_a^b.
\]

Proof. By the product rule,
\[
\frac{d}{dt} (f(t) g(t)) = f(t) \dot{g}(t) + \dot{f}(t) g(t) \text{ for all } a \leq t \leq b.
\]
Integrating this equation, making use of the fundamental theorem of calculus, gives the desired result.

Exercise 12.11. Show, using Calculus, that for any \( x, \varepsilon > 0 \),
\[
u^\varepsilon e^{-\varepsilon u} \leq \left( \frac{x}{\varepsilon} \right)^x e^{-x} \text{ for all } u \geq 0.
\]
In other words,
\[
u^\varepsilon \leq \left[ \left( \frac{x}{\varepsilon} \right)^x e^{-x} \right] e^{\varepsilon u} \text{ for all } u \geq 0.
\]

Exercise 12.12 (The Gamma Function). Let \( 0 < a < b < \infty \) and set
\[
\Gamma_{a,b}(x) := \int_a^b u^{x-1} e^{-u} \, du
\]
and let
\[
\Gamma(x) := \lim_{b \to \infty, a \to 0} \Gamma_{a,b}(x)
\]
when these limits exists. [This is an example of an improper integral and as usual we will typically write, \( \Gamma(x) = \int_0^\infty u^{x-1} e^{-u} \, du \).

1. Show \( \Gamma_{a,b}(x+1) = x \Gamma_{a,b}(x) - u^\varepsilon e^{-\varepsilon u}|_a^b \). [Hint: integrate by parts.]
2. Show for \( x \geq 0 \) that
\[
\lim_{a \downarrow 0} \Gamma_{a,b}(x+1) = \int_0^b \Gamma_b(x) du =: \Gamma_b(x),
\]
and
\[
\lim_{b \to \infty} \Gamma_b(x+1) =: \Gamma(x+1) \text{ exists.}
\]

**Hint:** For the limit as \( b \to \infty \) notice that \( b \to \Gamma_b(x+1) \) is an increasing function and hence \( \lim_{b \to \infty} \Gamma_b(x+1) = \sup_{b<\infty} \Gamma_b(x+1) \) exist in \([0, \infty)\). Your goal is to show that the limit is finite and for this you may find Exercise [12.11](#) useful. This exercise implies, for example, that there exists \( C = C(x) < \infty \) such that \( u^x e^{-u} \leq Ce^{-u/2} \) for \( u \geq 0 \).

3. Making use items 1. and 2. show for all \( x > 0 \) that \( \Gamma(x) = \lim_{a \downarrow 0} \lim_{b \to \infty} \Gamma_{a,b}(x) \) exists and satisfies \( \Gamma(x+1) = x\Gamma(x) \).

4. Show \( \Gamma(1) = 1 \) and use this and item 3. to conclude that \( \Gamma(n+1) = n! \) for all \( n \in \mathbb{N}_0 \).

[The function, \( \Gamma : (0, \infty) \to \mathbb{R}_+ \) is called the Gamma function.]

**Remark 12.24.** Later we will show \( \Gamma(1/2) = \sqrt{\pi} \) from which it then follows that
\[
\Gamma\left(n + \frac{1}{2}\right) = \frac{(2n-1)!!}{2^n} \sqrt{\pi}
\]
where \((-1)!! := 1\) and \((2n-1)!! = (2n-1)(2n-3) \ldots 1\) for \( n \in \mathbb{N} \).

### 12.2 Taylor’s Theorem revisited

**Proposition 12.25 (Taylor’s Theorem 1).** Suppose that \( h : (-\varepsilon, 1+\varepsilon) \to Y \) is a \( C^N \) - differentiable function. Then
\[
h(1) = \sum_{l=0}^{N-1} \frac{h^{(l)}(0)}{l!} + \int_0^1 (1-t)^{N-1} \frac{h^{(N)}(t)}{(N-1)!} dt.
\]
(12.15)

**Exercise 12.13.** Prove Proposition [12.25](#). That is show by induction on \( N \) that if \( h : (-\varepsilon, 1+\varepsilon) \to Y \) is a \( C^N \) - differentiable function, then
\[
h(1) = \sum_{l=0}^{N-1} \frac{h^{(l)}(0)}{l!} + \int_0^1 (1-t)^{N-1} \frac{h^{(N)}(t)}{(N-1)!} dt.
\]
(12.16)

**Hints:** the case \( N = 1 \) is the fundamental theorem of calculus. The induction step uses integration by parts and the observation that
\[
(1-t)^N = \frac{d}{dt} \left( \frac{(1-t)^{N+1}}{N+1} \right).
\]

**Theorem 12.26 (Taylor’s Theorem with Integral Remainder).** Suppose that \( f : (a,b) \to Y \) is \( C^N \) - differentiable and \( x_0 \in (a,b) \). Then for all \( x \in (a,b) \),
\[
f(x) = \sum_{l=0}^{N-1} \frac{f^{(l)}(x_0)}{l!} (x-x_0)^l + \frac{(x-x_0)^N}{N!} R_N(x)
\]
where \( R_N(x) = R_N(f,x_0; x) \) is given by
\[
R_N(x) = \int_0^1 f^{(N)}(x_0 + t(x-x_0)) (1-t)^{N-1} dt
\]
\[
= \int_0^1 f^{(N)} ((1-t)x_0 + tx) (1-t)^{N-1} dt.
\]
[Observe that \( \int_0^1 (1-t)^{N-1} dt = \frac{1}{N} \).]

**Proof.** We apply Proposition [12.25](#) with \( h(t) := f(x_0+t(x-x_0)) \) using,
\[
h^{(k)}(t) = f^{(k)}(x_0 + t(x-x_0))(x-x_0)^k.
\]
Therefore,
\[
f(x) = h(1) = \sum_{l=0}^{N-1} \frac{h^{(l)}(0)}{l!} + \int_0^1 (1-t)^{N-1} \frac{h^{(N)}(t)}{(N-1)!} dt
\]
\[
= \sum_{l=0}^{N-1} \frac{f^{(l)}(x_0)}{l!} (x-x_0)^l + \frac{(x-x_0)^N}{N!} \int_0^1 f^{(N)}(x_0 + t(x-x_0)) (1-t)^{N-1} dt.
\]

**Example 12.27.** Let \( X = (-1,1) \subset \mathbb{R}, \beta \in \mathbb{R} \) and \( f(x) = (1-x)^\beta \). The reader should verify
\[
f^{(m)}(x) = (-1)^m \beta(\beta-1) \ldots (\beta-m+1)(1-x)^{\beta-m}
\]
and therefore by Taylor’s theorem (Eq. [12.15](#) with \( x = 0 \) and \( y = x \))
\[
(1-x)^\beta = 1 + \sum_{m=1}^{N-1} \frac{1}{m!} (-1)^m \beta(\beta-1) \ldots (\beta-m+1)x^m + R_N(x)
\]
(12.17)
where
\[
R_N(x) = \frac{x^N}{N!} \int_0^1 (1-sx)^{\beta-1} \ldots (\beta-N+1)(1-sx)^{\beta-N} ds.
\]
\[
= \frac{x^N}{N!} (-1)^\beta - (\beta-N+1) \int_0^1 \frac{(1-s)N-1}{(1-sx)^{N-\beta}} ds.
\]
Now for $x \in (-1, 1)$ and $N > \beta$, 
\[
0 \leq \int_0^1 \frac{N(1-s)^{N-1}}{(1-sx)^{N-\beta}} ds \leq \int_0^1 \frac{N(1-s)^{N-1}}{(1-s)^{N-\beta}} ds = \int_0^1 N(1-s)^{\beta-1} ds = \frac{N}{\beta}
\]
and therefore,
\[
|R_N(x)| \leq \frac{|x|^N}{(N-1)!} |(\beta-1) \ldots (\beta-N+1)| =: \rho_N.
\]
Since
\[
\limsup_{N \to \infty} \rho_{N+1} \rho_N = |x| \cdot \limsup_{N \to \infty} \frac{N-\beta}{N} = |x| < 1
\]
and so by the Ratio test, $|R_N(x)| \leq \rho_N \to 0$ (exponentially fast) as $N \to \infty$. Therefore by passing to the limit in Eq. (12.17) we have proved
\[
(1-x)^\beta = 1 + \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \beta(\beta-1) \ldots (\beta-m+1)x^m
\]
(12.18)
which is valid for $|x| < 1$ and $\beta \in \mathbb{R}$. An important special cases is $\beta = -1$ in which case, Eq. (12.18) 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. (12.18) becomes
\[
\sqrt{1-x} = 1 + \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \left\{ \frac{1}{2} \right\} \ldots (\frac{1}{2} - m + 1)x^m
\]
(12.19)
\[
= 1 - \sum_{m=1}^{\infty} \frac{(2m-3)!!}{2^m m!} x^m \quad \text{for all } |x| < 1.
\]

**Theorem 12.28 (Intermediate Value Theorem for Averages).** Suppose $\rho : [0,1] \to \mathbb{R}^+$ is continuous, $\rho(t) > 0$ on $(0,1)$, and \(f[0]^t \rho(\tau) d\tau = 1\). Then for every $f \in C([0,1], \mathbb{R})$, show there exists $c \in (0,1)$ such that $f(t) = f(c)$.

**Proof.** Let $m := \min_{0 \leq t \leq 1} f(t)$ and $M := \max_{0 \leq t \leq 1} f(t)$. Then $m \leq f(t) \leq M$ for all $t$ and therefore,
\[
m = \int_0^1 \rho(\tau) m d\tau \leq \int_0^1 \rho(\tau) f(\tau) d\tau \leq \int_0^1 \rho(\tau) M d\tau \leq M.
\]
It now follows by the intermediate value theorem that there exists a $c \in [0,1]$ such that $f(c) = \int_0^1 \rho(\tau) f(\tau) d\tau$. The only thing left to do is to show that we may choose $c \in (0,1)$. If this is not the case we will have either $f(t) > L$ for all $0 < t < 1$ or $f(t) < L$ for all $0 < t < 1$ — again use the intermediate value theorem. But in the former case we will have $\lim_{t \to 0} f(t) = L_1$ for some $\varepsilon > 0$ and hence
\[
L = \int_0^1 f(t) \rho(t) dt = \int_0^{1/3} f(t) \rho(t) dt + \int_{1/3}^{1/2} f(t) \rho(t) dt + \int_{2/3}^1 f(t) \rho(t) dt
\]
\[
\geq \int_0^{1/3} L \rho(t) dt + \int_{1/3}^{2/3} L \rho(t) dt + \int_{1/3}^{2/3} \rho(t) (L + \varepsilon) dt
\]
\[
= L + \frac{\varepsilon}{3} \int_{1/3}^{2/3} \rho(t) dt \geq L
\]
which is a contradiction. A similar argument works if $f(t) < L$ for all $0 < t < 1$.

**Corollary 12.29 (Real valued Taylor’s Theorem).** If $f : (a,b) \to \mathbb{R}$ is $C^N$ differentiable and $x_0, x \in (a,b)$, then there exist $c \in (x_0, x)$ s.t.
\[
f(x) = \sum_{i=0}^{N-1} \frac{f^{(i)}(x_0)}{i!} (x-x_0)^i + \frac{f^{(N)}(c)}{N!} (x-x_0)^N.
\]

**12.3 Interchanging Limits**

**Theorem 12.30 (Interchanging derivates with limits).** Suppose that $-\infty < a < b < \infty$ and $f_n : (a,b) \to Y$ is a sequence of continuously differentiable functions such that;

1. $y := \lim_{n \to \infty} f_n(c)$ exists in $Y$ for some $c \in (a,b)$,
2. $g(t) := \lim_{n \to \infty} f_n(t)$ exits uniformly in $t$, i.e. $g : (a,b) \to Y$ such that $\lim_{n \to \infty} \sup_{t \in (a,b)} \|g(t) - f_n(t)\|_Y = 0$.

Then $f(t) := \lim_{n \to \infty} f_n(t)$ exists for all $t \in (a,b)$, $f_n(t) \to f(t)$ uniformly in $t$, and $f$ is continuously differentiable with $f(t) = g(t) = \lim_{n \to \infty} f_n(t)$.

**Proof.** By the fundamental theorem of calculus, for all $t \in (a,b)$ and $n \in \mathbb{N}$,
\[
f_n(t) = f_n(c) + \int_c^t f_n(\tau) d\tau.
\]
Using the continuity of the integral under uniform limits, we may let $n \to \infty$ in the previous equation to learn, $f(t) := \lim_{n \to \infty} f_n(t)$ exists and
Suppose that

\[ f(t) = y + \int_c^t g(\tau) \, d\tau. \]

Since \( g \) is the uniform limit of continuous functions we know that \( g \) is continuous as well. Therefore by the fundamental theorem of calculus, \( f(t) \) exits and is equal to \( g(t) \) for all \( t \in (a, b) \).

Since \( (\text{Prove Theorem 12.31.}) \) Suppose that \( \{K \times [a,b]\} \rightarrow Y \) is a continuous function such that \( f_x(s,t) := \frac{\partial}{\partial s} f(s,t) \) exists for \( s \in (u,v) \) and \( f_x : (u,v) \times [a,b] \rightarrow Y \) is a continuous. Then

\[
\frac{d}{ds} \int_a^b f(s,t) \, d\mu(t) = \int_a^b f_x(s,t) \, d\mu(t) \quad \text{for all } s \in (u,v). \]

\textbf{Proof.} Let \( s \in (u,v) \) and choose \( \varepsilon > 0 \) sufficiently small so that \( |h| \leq \varepsilon \). Let \( F(s) := \int_a^b f(s,t) \, d\mu(t) \). Then for \( 0 < |h| \leq \varepsilon \)

\[
\frac{F(s+h) - F(s)}{h} = \int_a^b \frac{f(s+h,t) - f(s,t)}{h} \, d\mu(t) = \int_a^b g(h,t) \, d\mu(t)
\]

where

\[
g(h,t) = \begin{cases} 
\frac{f(s+h,t) - f(s,t)}{h} & \text{if } 0 < |h| \leq \varepsilon \\
0 & \text{if } h = 0
\end{cases}
\]

The function \( g(h,t) \) may also be written as,

\[
g(h,t) = \int_0^1 f_x(s + rh,t) \, dr
\]

as this formula agrees trivially for \( h = 0 \) and for \( h \neq 0 \) by the fundamental theorem of calculus. The function \( u(h,t,r) := f_x(s + rh,t) \) is continuous for on \([-\varepsilon,\varepsilon] \times [a,b] \times [0,1] \) and so by Theorem 12.31 it follows that \( g(h,t) = \int_0^1 u(h,t,r) \, dr \) is continuous in \((h,t)\). Therefore by another application of Theorem 12.31

\[
\lim_{h \to 0} \frac{F(s+h) - F(s)}{h} = \lim_{h \to 0} \int_a^b g(h,t) \, d\mu(t) = \int_a^b g(0,t) \, d\mu(t) = \int_a^b f_x(s,t) \, d\mu(t).
\]

\textbf{Example 12.33.} Let \( 0 < b < \infty \) and \( n \in \mathbb{N} \), then by repeated use of Theorem 12.32

\[
\int_0^b e^{-sx}x^n \, dx = \int_0^b \left(-\frac{d}{ds}\right)^n e^{-sx} \, dx
\]

\[
= \left(-\frac{d}{ds}\right)^n \int_0^b e^{-sx} \, dx = \left(-\frac{d}{ds}\right)^n \left(1 - e^{-sb}\right).
\]
A simple induction argument shows,
\[
\left( -\frac{d}{ds} \right)^n \frac{1}{s} (1 - e^{-sb}) = n! s^{-(n+1)} + e^{-bs} p_n \left( s^{-1} \right)
\]
for some polynomial function \( p_n \). Therefore we may conclude that
\[
\int_0^\infty e^{-sx} x^n dx := \lim_{b \to \infty} \int_0^b e^{-sx} x^n dx = \frac{n!}{s^{n+1}}
\]
and taking \( s = 1 \) we find,
\[
\Gamma(n+1) = \int_0^\infty e^{-x} x^n dx = n!
\]
as you already showed in Exercise 12.12.

**Exercise 12.15.** Explain why the following identities hold for \( t \neq 0 \):
\[
\int_{-\pi}^\pi x \sin t x \, dx = \frac{d}{dt} \int_{-\pi}^\pi \cos t x \, dx = -2 \frac{d}{dt} \sin t \pi \tag{12.20}
\]
and
\[
\int_{-\pi}^\pi x^2 \cos t x \, dx = \frac{d}{dt} \int_{-\pi}^\pi x \sin t x \, dx = -2 \frac{d^2}{dt^2} \sin t \pi \tag{12.21}
\]
Now use these formulas to show
\[
\int_{-\pi}^\pi x \sin n x \, dx = 2 \pi \frac{(-1)^{n+1}}{n}
\]
and
\[
\int_{-\pi}^\pi x^2 \cos n x \, dx = 4 \pi \frac{(-1)^n}{n^2}
\]
for all \( n \in \mathbb{Z} \setminus \{0\} \).

### 12.4 Stirling’s Formula

**Warning:** methods in this section are beyond what we have covered so far.

On occasion one is faced with estimating an integral of the form, \( \int_J e^{-G(t)} dt \), where \( J = (a,b) \subset \mathbb{R} \) and \( G(t) \) is a \( C^1 \) function with a unique (for simplicity) global minimum at some point \( t_0 \in J \). The idea is that the majority contribution of the integral will often come from some neighborhood, \((t_0 - \alpha, t_0 + \alpha)\), of \( t_0 \). Moreover, it may happen that \( G(t) \) can be well approximated on this neighborhood by its Taylor expansion to order 2;
\[
G(t) \approx G(t_0) + \frac{1}{2} \frac{d^2}{dt^2} G(t_0) (t - t_0)^2.
\]

Notice that the linear term is zero since \( G(t_0) = 0 \). We will further assume that \( \frac{d}{dt} G(t_0) \neq 0 \) and hence \( \frac{d^2}{dt^2} G(t_0) > 0 \). Under these hypothesis we will have,
\[
\int_J e^{-G(t)} dt \approx e^{-G(t_0)} \int_{|t - t_0| < \alpha} \exp \left( -\frac{1}{2} \frac{d}{dt} G(t_0) (t - t_0)^2 \right) dt.
\]
Making the change of variables, \( s = \sqrt{\frac{d}{dt} G(t_0)} (t - t_0), \) in the above integral then gives,
\[
\int_J e^{-\tilde{G}(t)} dt \approx e^{-G(t_0)} \frac{1}{\sqrt{\frac{d}{dt} G(t_0)}} \int_{|s| < \sqrt{\frac{d}{dt} G(t_0)} \cdot \alpha} e^{-\frac{1}{2} s^2} ds
\]
\[
= \frac{1}{\sqrt{\frac{d}{dt} G(t_0)}} \left[ \sqrt{2\pi} - \int_{\sqrt{\frac{d}{dt} G(t_0)} \cdot \alpha}^{\infty} e^{-\frac{1}{2} s^2} ds \right]
\]
\[
= \frac{1}{\sqrt{\frac{d}{dt} G(t_0)}} \left[ \sqrt{2\pi} - O \left( \frac{1}{\sqrt{\frac{d}{dt} G(t_0)} \cdot \alpha^2} \right) \right].
\]
If \( \alpha \) is sufficiently large, for example if \( \sqrt{\frac{d}{dt} G(t_0)} \cdot \alpha = 3 \), then the error term is about 0.0037 and we should be able to conclude that
\[
\int_J e^{-G(t)} dt \approx \sqrt{\frac{2\pi}{\frac{d}{dt} G(t_0)}} e^{-G(t_0)}. \tag{12.22}
\]
The proof of the next theorem (Stirling’s formula for the Gamma function) will illustrate these ideas and what one has to do to carry them out rigorously.

**Theorem 12.34 (Stirling’s formula).** The Gamma function (see Exercise 12.12), satisfies Stirling’s formula,
\[
\lim_{x \to \infty} \frac{\Gamma(x+1)}{\sqrt{2\pi} e^{-x} x^{x+1/2}} = 1. \tag{12.23}
\]
In particular, if \( n \in \mathbb{N} \), we have
\[
n! = \Gamma(n+1) \sim \sqrt{2\pi} e^{-n} n^{n+1/2}
\]
where we write \( a_n \sim b_n \) to mean, \( \lim_{n \to \infty} \frac{a_n}{b_n} = 1 \). (See Example 12.37 below for a slightly cruder but more elementary estimate of \( n! \))
Proof. (The following proof is an elaboration of the proof found on page 236-237 in Krantz’s Real Analysis and Foundations.) We begin with the formula for \( \Gamma(x+1) \):

\[
\Gamma(x+1) = \int_{0}^{\infty} e^{-t} t^{x} \, dt = \int_{0}^{\infty} e^{-s} s^{x} \, ds, \tag{12.24}
\]

where

\[
G_{x}(t) := t - x \ln t.
\]

Then \( \dot{G}_{x}(t) = 1 - x/t \), \( \ddot{G}_{x}(t) = x/t^{2} \), \( G_{x} \) has a global minimum (since \( \ddot{G}_{x} > 0 \)) at \( t_{0} = x \) where

\[
G_{x}(x) = x - x \ln x \text{ and } \ddot{G}_{x}(x) = 1/x.
\]

So if Eq. \([12.22]\) is valid in this case we should expect,

\[
\Gamma(x+1) \approx \sqrt{2\pi} e^{-(x - x \ln x)} = \sqrt{2\pi} e^{-x^{2}/2}
\]

which would give Stirling’s formula. The rest of the proof will be spent on rigorously justifying the approximations involved.

Let us begin by making the change of variables \( s = \sqrt{G(t_{0})}(t - t_{0}) = \frac{1}{x^{2}}(t - x) \) as suggested above. Then

\[
G_{x}(t) - G_{x}(x) = (t - x) - x \ln (t/x) = \sqrt{x} - x \ln \left( \frac{x + \sqrt{x}s}{x} \right)
\]

\[
= x \left[ \frac{s}{\sqrt{x}} - \ln \left( 1 + \frac{s}{\sqrt{x}} \right) \right] = s^{2} q \left( \frac{s}{\sqrt{x}} \right)
\]

where

\[
q(u) := \frac{1}{u^{2}} [u - \ln (1 + u)] \text{ for } u > -1 \text{ with } q(0) := \frac{1}{2}.
\]

Setting \( q(0) = 1/2 \) makes \( q \) a continuous and in fact smooth function on \((-1, \infty)\), see Figure \ref{fig:q(u)}. Using the power series expansion for \( \ln(1+u) \) we find,

\[
q(u) = \frac{1}{2} + \sum_{k=3}^{\infty} \frac{(-u)^{k-2}}{k} \text{ for } |u| < 1. \tag{12.25}
\]

Making the change of variables, \( t = x + \sqrt{x}s \) in the second integral in Eq. \([12.24]\) yields,

\[
\Gamma(x+1) = e^{-(x - x \ln x)} \sqrt{x} \int_{-\infty}^{\infty} e^{-q(u)} u^{2} \, ds = x^{x+1/2} e^{-x} \cdot I(x),
\]

where

\[
I(x) = \int_{-\infty}^{\infty} e^{-q(u)} u^{2} \, ds = \int_{-\infty}^{\infty} 1_{s \geq \sqrt{x}} e^{-q(u)} u^{2} \, ds. \tag{12.26}
\]

From Eq. \([12.25]\) it follows that \( \lim_{u \to 0} q(u) = 1/2 \) and therefore,

\[
\int_{-\infty}^{\infty} \lim_{x \to \infty} \left[ 1_{s \geq -\sqrt{x}} e^{-q(u)} u^{2} \right] ds = \int_{-\infty}^{\infty} e^{-x^{2}/2} ds = \sqrt{2\pi}. \tag{12.27}
\]

So if there exists a dominating function, \( F \in L^{1}(\mathbb{R}, m) \), such that

\[
1_{s \geq -\sqrt{x}} e^{-q(u)} u^{2} \leq F(s) \text{ for all } s \in \mathbb{R} \text{ and } x \geq 1,
\]

we can apply the DCT to learn that \( \lim_{x \to \infty} I(x) = \sqrt{2\pi} \) which will complete the proof of Stirling’s formula.

We now construct the desired function \( F \). From Eq. \([12.25]\) it follows that \( q(u) \geq 1/2 \) for \(-1 < u \leq 0 \). Since \( u - \ln (1 + u) > 0 \) for \( u \neq 0 \) \((u - \ln (1 + u)) \) is convex and has a minimum of 0 at \( u = 0 \) we may conclude that \( q(u) > 0 \) for all \( u > -1 \) therefore by compactness (on \([0, M]\), \( \min_{-1 < u < M} q(u) = \varepsilon(M) > 0 \) for all \( M \in (0, \infty) \), see Remark \([12.33]\) for more explicit estimates. Lastly, since \( \frac{1}{u} \ln (1 + u) \to 0 \) as \( u \to \infty \), then exists \( M < \infty \) \((M = 3 \text{ would due})\) such that \( \frac{1}{u} \ln (1 + u) \leq \frac{1}{2} \) for \( u \geq M \) and hence,

\[
q(u) = \frac{1}{u} \left[ 1 - \frac{1}{u} \ln (1 + u) \right] \geq \frac{1}{2u} \text{ for } u \geq M.
\]

So there exists \( \varepsilon > 0 \) and \( M < \infty \) such that (for all \( x \geq 1 \),

\[
\int_{-\infty}^{\infty} e^{-q(u)} u^{2} \, ds = \int_{-\infty}^{\infty} 1_{s \geq -\sqrt{x}} e^{-q(u)} u^{2} \, ds,
\]

Fig. 12.1. Plot of \( q(u) \).
\[ 1_{s \geq -\sqrt{2} e^{-\frac{x^2}{2}}} e^{-x^2} + 1_{s \geq 0} e^{-s^2} + 1_{s \geq 0} e^{-s^2} \leq e^{-x^2} + e^{-s^2} =: F(s) \in L^1(\mathbb{R}, ds). \]

We will sometimes use the following variant of Eq. (12.23):

\[
\lim_{x \to \infty} \frac{\Gamma(x)}{\sqrt{2\pi e^{-x^2} \cdot e \cdot (x-1)^{x-1/2}}} = 1
\]

To prove this let \( x \) go to \( x - 1 \) in Eq. (12.23) in order to find,

\[
1 = \lim_{x \to \infty} \frac{\Gamma(x)}{\sqrt{2\pi e^{-x^2} \cdot e \cdot (x-1)^{x-1/2}}} = \lim_{x \to \infty} \frac{\Gamma(x)}{\sqrt{2\pi e^{-x^2} \cdot e \cdot (x-1)^{x-1/2}}}
\]

which gives Eq. (12.28) since

\[
\lim_{x \to \infty} \sqrt{\frac{x}{x-1}} \cdot e \cdot \left(1 - \frac{1}{x}\right)^{x-1/2} = 1.
\]

Remark 12.35 (Estimating \( q(u) \) by Taylor’s Theorem). Another way to estimate \( q(u) \) is to use Taylor’s theorem with integral remainder. In general if \( h \) is \( C^2 \) function on \([0,1]\), then by the fundamental theorem of calculus and integration by parts,

\[
\begin{align*}
  h(1) - h(0) &= \int_0^1 \hat{h}(t) \, dt = -\hat{h}(0) \int_0^1 (1-t) \\
  &= -\hat{h}(0) + \int_0^1 \hat{h}(t) (1-t) \, dt \\
  &= \hat{h}(0) + \frac{1}{2} \int_0^1 \hat{h}(t) \, d\nu(t)
\end{align*}
\]

where \( d\nu(t) := 2(1-t) \, dt \) which is a probability measure on \([0,1]\). Applying this to \( h(t) = F(a + t(b-a)) \) for a \( C^2 \) function on an interval of points between \( a \) and \( b \) in \( \mathbb{R} \) then implies,

\[
F(b) - F(a) = (b-a) F'(a) + \frac{1}{2} (b-a)^2 \int_0^1 F'(a + t(b-a)) \, d\nu(t).
\]

(Similar formulas hold to any order.) Applying this result with \( F(x) = x - \ln(1 + x) \), \( a = 0 \), and \( b = u \in (-1, \infty) \) gives,

\[
u(1 + u) = \frac{1}{2} u^2 \int_0^1 \frac{1}{(1 + tu)^2} \, d\nu(t),
\]

i.e.

\[
q(u) = \frac{1}{2} \int_0^1 \frac{1}{(1 + tu)^2} \, d\nu(t).
\]

From this expression for \( q(u) \) it now easily follows that

\[
q(u) \geq \frac{1}{2} \int_0^1 \frac{1}{(1 + u)^2} \, d\nu(t) = \frac{1}{2} \text{ if } -1 < u \leq 0
\]

and

\[
q(u) \geq \frac{1}{2} \int_0^1 \frac{1}{(1 + u)^2} \, d\nu(t) = \frac{1}{2} \text{ if } -1 < u \leq 0
\]

So an explicit formula for \( \varepsilon(M) \) is \( \varepsilon(M) = (1 + M)^{-2}/2 \).

12.4.1 A primitive Stirling type approximation

Theorem 12.36. Suppose that \( f : (0, \infty) \to \mathbb{R} \) is an increasing concave down function (like \( f(x) = \ln x \)) and let \( s_n := \sum_{k=1}^{n} f(k) \), then

\[
s_n - \frac{1}{2} (f(n) + f(1)) \leq \int_1^n f(x) \, dx
\]

\[
\leq s_n - \frac{1}{2} [f(n+1) + 2f(1)] + \frac{1}{2} f(2)
\]

\[
\leq s_n - \frac{1}{2} [f(n) + 2f(1)] + \frac{1}{2} f(2).
\]
Proof. On the interval, \([k - 1, k]\), we have that \(f(x)\) is larger than the straight line segment joining \((k - 1, f(k - 1))\) and \((k, f(k))\) and thus

\[
\frac{1}{2} (f(k) + f(k - 1)) \leq \int_{k-1}^{k} f(x)\,dx.
\]

Summing this equation on \(k = 2, \ldots, n\), shows

\[
s_n - \frac{1}{2} (f(n) + f(1)) = \sum_{k=2}^{n} \frac{1}{2} (f(k) + f(k - 1)) \leq \sum_{k=2}^{n} \int_{k-1}^{k} f(x)\,dx = \int_{1}^{n} f(x)\,dx.
\]

For the upper bound on the integral we observe that \(f(x) \leq f(k) - f'(k)\,(x-k)\) for all \(x\) and therefore,

\[
\int_{k-1}^{k} f(x)\,dx \leq \int_{k-1}^{k} [f(k) - f'(k)\,(x-k)]\,dx = f(k) - \frac{1}{2} f'(k).
\]

Summing this equation on \(k = 2, \ldots, n\) then implies,

\[
\int_{1}^{n} f(x)\,dx \leq \sum_{k=2}^{n} f(k) - \frac{1}{2} \sum_{k=2}^{n} f'(k).
\]

Since \(f''(x) \leq 0\), \(f'(x)\) is decreasing and therefore \(f'(x) \leq f'(k - 1)\) for \(x \in [k - 1, k]\) and integrating this equation over \([k - 1, k]\) gives

\[
f(k) - f(k - 1) \leq f'(k - 1).
\]

Summing the result on \(k = 3, \ldots, n + 1\) then shows,

\[
f(n + 1) - f(2) \leq \sum_{k=2}^{n} f'(k)
\]

and thus it follows that

\[
\int_{1}^{n} f(x)\,dx \leq \sum_{k=2}^{n} f(k) - \frac{1}{2} (f(n + 1) - f(2))
\]

\[= s_n - \frac{1}{2} [f(n + 1) + 2f(1)] + \frac{1}{2} f(2)
\]

\[\leq s_n - \frac{1}{2} [f(n) + 2f(1)] + \frac{1}{2} f(2)
\]

\[= 2^n - 2(f(1) + f(2))
\]

Example 12.37 (Approximating \(n!\)). Let us take \(f(n) = \ln n\) and recall that

\[
\int_{1}^{n} \ln x\,dx = n \ln n - n + 1.
\]

Thus we may conclude that

\[
s_n - \frac{1}{2} \ln n \leq n \ln n - n + 1 \leq s_n - \frac{1}{2} \ln n + \frac{1}{2} \ln 2.
\]

Thus it follows that

\[
(n + \frac{1}{2}) \ln n + 1 - \ln \sqrt{2} \leq s_n \leq \left(n + \frac{1}{2}\right) \ln n - n + 1.
\]

Exponentiating this identity then implies,

\[
\frac{e}{\sqrt{2}} \cdot e^{-n} n^{n+1/2} \leq n! \leq e \cdot e^{-n} n^{n+1/2}
\]

which compares well with Strirling’s formula (Theorem 12.34) which states,

\[
n! \sim \sqrt{2\pi e^{-n} n^{n+1/2}}.
\]

Observe that

\[
\frac{e}{\sqrt{2}} \approx 1.9207 \leq \sqrt{2\pi} \approx 2.506 \leq e \approx 2.7183.
\]

12.4.2 Bone Yard

Exercise 12.16. If \(f(x) = \ln(1 + x)\) we know that

\[
f^{(n)}(x) = (-1)^{n+1} (n-1)! \cdot (1 + x)^{-n}.
\]

Therefore the integral remainder term in Taylor’s theorem is given by

\[
R_N(x) = \frac{1}{N!} \int_{0}^{x} f^{(N+1)}(y) (x-y)^N \,dy
\]

\[
= \frac{1}{N!} (-1)^N N! \int_{0}^{x} (1 + x)^{-(N+1)} (x-y)^N \,dy
\]

\[
= (-1)^N \int_{0}^{x} \left(\frac{x-y}{1+y}\right)^N \frac{1}{1+y} \,dy.
\]

and hence

\[
|R_N(x)| \leq \int_{0}^{x} \left|\frac{x-y}{1+y}\right|^N \frac{1}{1+y} \,dy.
\]
continuously differentiable, then

\begin{equation}
\left(\frac{x-y}{1+y}\right)^N = \left(\frac{|x|-|y|}{1-|y|}\right)^N
\end{equation}

and therefore,

\begin{equation}
|R_N(x)| \leq \left|\int_0^x \left[\frac{|x|-|y|}{1+y}\right]^N \frac{1}{1+y} dy \right| \to 0 \text{ as } N \to \infty
\end{equation}

wherein we have used, as \( f(t) := \frac{|x|-t}{1-t} \) satisfies

\begin{equation}
\dot{f}(t) = \frac{-(1-t) + (|x|-t)}{(1-t)^2} = \frac{|x|-1}{(1-t)^2} < 0
\end{equation}

so \( f(t) \) on \([0,|x|]\) has its maximum at \( t = 0 \) and so

\begin{equation}
\frac{|x|-|y|}{1-|y|} \leq \frac{|x|}{1} = |x|
\end{equation}

Exercise 12.17. For \(-1 < x \leq 0\), the best we can do with mean value theorem type remainder is

\begin{equation}
|R_N(x)| = \frac{1}{N+1} \left|\frac{x}{1+c_N}\right|^{N+1} \leq \frac{1}{N+1} \frac{|x|}{1-x} |x|^{N+1}.
\end{equation}

Theorem 12.38 (Taylor's at order 1 and 2). If \( f(t) \) for \(0 \leq t \leq 1\) is twice continuously differentiable, then

\begin{equation}
f(1) = f(0) + \int_0^1 \dot{f}(t) \, dt \quad \text{and}
\end{equation}

\begin{equation}
f(1) = f(0) + \frac{\dot{f}(0)}{1!} + \frac{1}{1!} \int_0^1 \ddot{f}(t) \, (1-t) \, dt.
\end{equation}

Proof. The first assertion follows by the fundamental theorem of calculus

\begin{equation}
f(1) - f(0) = \int_0^1 \dot{f}(t) \, dt.
\end{equation}

For the second we integrate by parts as follows;

\begin{equation}
\int_0^1 \dot{f}(t) \, dt = -\int_0^1 \ddot{f}(t) \, d(1-t) = -\ddot{f}(t) \, (1-t) \bigg|_0^1 + \int_0^1 \ddot{f}(t) \, (1-t) \, dt
\end{equation}

and therefore

\begin{equation}
f(1) = f(0) + \int_0^1 \dot{f}(t) \, dt = f(0) + \frac{\dot{f}(0)}{1!} + \frac{1}{1!} \int_0^1 \ddot{f}(t) \, (1-t) \, dt.
\end{equation}

\begin{equation}
\begin{aligned}
\int_0^1 \dot{f}(t) \, dt &= \int_0^1 \ddot{f}(t) \, d(1-t) = \ddot{f}(t) \, (1-t) \bigg|_0^1 + \int_0^1 \dddot{f}(t) \, (1-t)^2 \, dt \\
&= \ddot{f}(0) + \int_0^1 \dddot{f}(t) \, (1-t)^2 \, dt
\end{aligned}
\end{equation}

Exercise 12.18. If \( f : \mathbb{R} \to \mathbb{C} \) is a function which is differentiable to all orders, show for all \( N \in \mathbb{N}_0 = \{0, 1, 2, 3, \ldots\} \) that

\begin{equation}
f(1) = \sum_{k=0}^N \frac{1}{k!} f^{(k)}(0) + \frac{1}{N!} \int_0^1 f^{(N+1)}(t) \, (1-t)^N \, dt.
\end{equation}

Hint: use integration by parts and induction with Theorem 12.38 providing the case \( N = 0 \) and \( N = 1 \).

Exercise 12.19. If \( f : \mathbb{R} \to \mathbb{R} \) is a function which is differentiable to all orders, show there exists \( c \in (0, 1) \) such that

\begin{equation}
\frac{1}{N!} \int_0^1 f^{(N+1)}(t) \, (1-t)^N \, dt = \frac{f^{(N+1)}(c)}{(N+1)!}.
\end{equation}

Hint: apply Exercise 12.18 with \( \rho(t) = (N+1) \, (1-t)^N \).

Exercise 12.20. Recall the if we define \( e^z := e^x (\cos y + i \sin y) \) where \( z = x + iy \), then \( \frac{\partial}{\partial t} e^{iz} = ze^{iz} \). Use Exercise 12.18 with \( f(t) = e^{zt} \) to conclude;

\begin{equation}
e^z = f(1) = \sum_{k=0}^N \frac{z^k}{k!} + R_N(z)
\end{equation}
where
\[ R_N(z) = \frac{z^{N+1}}{N!} \int_0^1 e^{tz} (1 - t)^N \, dt. \] (12.32)

Then show that \( \lim_{N \to \infty} |R_N(z)| = 0 \) for all \( z \in \mathbb{C} \) and use this to conclude
\[ e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!} \] for all \( z \in \mathbb{C} \).
Part IV

A Little Functional Analysis
Weierstrass Approximation Theorem

In this chapter, \((X,d)\), will be a metric space. Recall that if \(Z\) is a Banach space then \(C(X,Z)\) and \(BC(X,Z)\) denotes the continuous and bounded continuous functions form \(X\) to \(Z\) respectively. Recall that \(BC(X,Z)\) is a Banach spaces when equipped with the supremum norm,

\[
\|f\|_* := \sup_{x \in X} \|f(x)\|_Z.
\]

(I will also often write \(\|f\|_\infty\) for \(\|f\|_*\)). We are most interested in the case where \(Z = \mathbb{R}\) or \(\mathbb{C}\). The main goal of this chapter is to find dense subsets of \(C(X,\mathbb{R})\) and \(C(X,\mathbb{C})\) when \((X,d)\) is compact.

Definition 13.1 (Support). Let \(f: X \to Z\) be a function from a metric space \((X,d)\) to a vector space \(Z\). The support of \(f\) is the closed subset, \(\text{supp}(f)\), of \(X\) defined by

\[
\text{supp}(f) := \{x \in X : f(x) \neq 0\}.
\]

Example 13.2. 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 \(\text{supp}(f) = [0,4\pi]\).

13.1 Classical Weierstrass Approximation Theorem

Are first priority is to consider the metric space \(X = J = [a,b]\) for some \(-\infty < a < b < \infty\). For the remainder of this section, \(Z\) will be used to denote a Banach space.

Definition 13.3 (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 13.4 (Approximate \(\delta -\) 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

\[
\begin{align*}
\int_{\mathbb{R}} q_n(x) \, dx &= 1 \quad \text{and} \\
\lim_{n \to \infty} \int_{|x| \geq \varepsilon} q_n(x) \, dx &= 0 \quad \text{for all } \varepsilon > 0.
\end{align*}
\]

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 \(\mathbb{R}\).

Proof. Let \(x \in \mathbb{R}\), then because of Eq. (13.1),

\[
\|q_n * f (x) - f (x)\| = \left\| \int_{\mathbb{R}} q_n(y)(f(x-y) - f (x)) \, dy \right\| \leq \int_{\mathbb{R}} q_n(y) \|f(x-y) - f (x)\| \, dy.
\]

Let \(M = \sup \{\|f(x)\| : x \in \mathbb{R}\}\). Then for any \(\varepsilon > 0\), using Eq. (13.1),

\[
\|q_n * f (x) - f (x)\| \leq \int_{|y| \leq \varepsilon} q_n(y) \|f(x-y) - f (x)\| \, dy + \int_{|y| > \varepsilon} q_n(y) \|f(x-y) - f (x)\| \, dy \\
\leq \sup_{|y| \leq \varepsilon} \|f(x + w) - f (x)\| + 2M \int_{|y| > \varepsilon} q_n(y)dy.
\]

So if \(K\) is a compact subset of \(\mathbb{R}\) (for example a large interval) we have

\[
\begin{align*}
\sup_{(x) \in K} \|q_n * f (x) - f (x)\| &\leq \sup_{|w| \leq \varepsilon, x \in K} \|f(x + w) - f (x)\| + 2M \int_{|y| > \varepsilon} q_n(y)dy.
\end{align*}
\]
and hence by Eq. (13.2),

\[ \limsup_{n \to \infty} \sup_{x \in K} \| q_n * f (x) - f (x) \| \leq \sup_{|x| \leq \varepsilon, \ x \in K} \| f (x + w) - f (x) \|. \]

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

\[ q_n (x) := \frac{1}{c_n} (1 - x^2)^n 1_{|x| \leq 1} \text{ where } c_n := \int_{-1}^{1} (1 - x^2)^n \, dx. \]  

(13.3)

Figure 13.1 displays the key features of the functions \( q_n \).

![Figure 13.1](image)

Fig. 13.1. 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 13.5.** The sequence \( \{ q_n \}_{n=1}^{\infty} \) is an approximate \( \delta \) - sequence, i.e. they satisfy Eqs. (13.1) and (13.2).

**Proof.** By construction, \( q_n \in C_c (\mathbb{R}, [0, \infty)) \) for each \( n \) and Eq. (13.1) holds. Since

\[ \int_{|x| \geq \varepsilon} q_n (x) \, dx = \frac{2}{\varepsilon} \int_{0}^{1} (1 - x^2)^n \, dx + \frac{2}{\varepsilon^2} \int_{1}^{\infty} (1 - x^2)^n \, dx \]

\[ \leq \int_{0}^{\varepsilon} \frac{2}{\varepsilon} (1 - x^2)^n \, dx + \int_{1}^{\varepsilon} \frac{2}{\varepsilon^2} (1 - x^2)^n \, dx = \frac{(1 - \varepsilon^2)^n + 1}{1 - (1 - \varepsilon^2)^n + 1} \to 0 \text{ as } n \to \infty, \]

the proof is complete.

**Theorem 13.6 (Classical Weierstrass Approximation Theorem).** Suppose \( -\infty < a < b < \infty, \ J = [a, b] \) and \( f \in C(J, Z) \). Then there exists polynomial functions, \( p_n : \mathbb{R} \to Z \) such that \( p_n \to f \) uniformly on \( J \).

**Proof.** By replacing \( f \) by \( F \) where

\[ F(t) := f (a + t (b - a)) - f (a) + t (f (b) - f (a)) \] 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. (13.3), let \( f_n (x) := (q_n * f) (x) \) and recall from Lemma 13.4 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 \( x \in [0, 1] \),

\[ f_n (x) = \int_{\mathbb{R}} q_n (x - y) f (y) \, dy \]

\[ = \frac{1}{c_n} \int_{[0, 1]} f (y) \left[ (1 - (x - y)^2)^n 1_{|x - y| \leq 1} \right] \, dy \]

\[ = \frac{1}{c_n} \int_{[0, 1]} f (y) (1 - (x - y)^2)^n \, dy \]

\[ = \frac{1}{c_n} \int_{[0, 1]} f (y) \sum_{k=0}^{2n} a_k (y) x^k \, dy = \sum_{k=0}^{2n} A_k x^k \]

where

\[ A_k = \frac{1}{c_n} \int_{[0, 1]} f (y) a_k (y) \, dy. \]
Corollary 13.7. Any continuous function \( f : J \to Z \) may be approximated uniformly by infinitely differentiable functions.

Exercise 13.1. Suppose \(-\infty < a < b < \infty \) and \( g \in C((a,b], \mathbb{R}) \) satisfies \( g(t) \geq 0 \) for all \( t \in J := [a, b] \) and \( \int_J g(t) \, dt = 0 \). Show \( g \equiv 0 \).

Exercise 13.2. Let \( J = [a, b] \) where \(-\infty < a < b < \infty \). Show if \( f \in C(J, \mathbb{C}) \) satisfies
\[
\int_J f(t) \, t^n \, dt = 0 \quad \text{for all } n \in \mathbb{N}_0,
\]
then \( f \equiv 0 \). Hint: use Theorem 13.6 to conclude that \( \int_J |f(t)|^2 \, dt = 0 \) and then use Exercise 13.1.

Example 13.8 (Probability densities are determined by their moments). Let \( J = [a, b) \) where \(-\infty < a < b < \infty \) and suppose that \( \rho_1 \) and \( \rho_2 \) are two non-negative continuous probability density functions on \( J \) having the same moments, then \( \rho_1 = \rho_2 \). Indeed by assumption that \( \rho_1 \) and \( \rho_2 \) are probability density functions with the same moments is exactly that
\[
\int_J t^n \rho_1(t) \, dt = \int_J t^n \rho_2(t) \, dt \quad \text{for all } n \in \mathbb{N}_0.
\]
This last equation is equivalent to
\[
\int_J t^n [\rho_1(t) - \rho_2(t)] \, dt = 0 \quad \text{for all } n \in \mathbb{N}_0
\]
and hence by Exercise 13.2 it follows that \( \rho_1(t) = \rho_2(t) \) for all \( t \in J \).

13.2 Stone-Weierstrass Theorem for Compact \( X \)

We now wish to generalize Theorem 13.6 to general metric spaces and to generalize to collections other than polynomial functions.

Definition 13.9. Let \( X \) be a metric space and \( A \subseteq C(X) = C(X, \mathbb{R}) \) or \( C(X, \mathbb{C}) \) be a collection of functions. Then
\begin{enumerate}
  \item \( A \) is said to be \textbf{separate points} if for all distinct points \( x, y \in X \) there exists \( f \in A \) such that \( f(x) \neq f(y) \).
  \item \( A \) is an \textbf{algebra} if \( A \) is a vector subspace of \( C(X) \) which is closed under pointwise multiplication. (\textbf{Note well:} we do not assume \( 1 \in A \).)
  \item \( A \subseteq C(X, \mathbb{R}) \) is called \textbf{a lattice} if \( f \lor g := \max(f, g) \) and \( f \land g = \min(f, g) \in A \) for all \( f, g \in A \).
  \item \( A \subseteq C(X, \mathbb{C}) \) is closed under conjugation if \( \overline{f} \in A \) whenever \( f \in A \).
\end{enumerate}

Lemma 13.10. If \( A \) is a closed (relative to the supremum norm) sub-algebra of \( BC(X, \mathbb{R}) \) then \(|f| \in A \) for all \( f \in A \) and \( A \) is a lattice.

\textbf{Proof.} Let \( f \in A \) and let \( M = \sup_{x \in X} |f(x)| \). By Theorem 13.6 there are polynomials \( p_n(t) \) such that
\[
\lim_{n \to \infty} \sup_{|t| \leq M} |t - p_n(t)| = 0.
\]
By replacing \( p_n \) by \( p_n - p_n(0) \) if necessary we may assume that \( p_n(0) = 0 \). Since \( A \) is an algebra, it follows that \( f_n = p_n(f) \in A \) and \(|f| \in A \), because \(|f| \) is the uniform limit of the \( f_n \)'s. Since
\[
f \lor g = \frac{1}{2} (f + g + |f - g|)
\]
and
\[
f \land g = \frac{1}{2} (f + g - |f - g|),
\]
we have shown \( A \) is a lattice. \( \blacksquare \)

Lemma 13.11. Let \( A \subseteq C(X, \mathbb{R}) \) be an algebra which separates points and suppose \( x \) and \( y \) are distinct points of \( X \). If there exists \( f, g \in A \) such that \( f(x) \neq 0 \) and \( g(y) \neq 0 \), then
\[
V := \{(f(x), g(y)) : f \in A\} = \mathbb{R}^2.
\]

\textbf{Proof.} It is clear that \( V \) is a non-zero subspace of \( \mathbb{R}^2 \). If \( \dim(V) = 2 \), then \( V = \text{span}(a, b) \) for some \((a, b) \in \mathbb{R}^2 \) which, necessarily by Eq. (13.5), satisfy \( a \neq b \). Since \((a, b) = (f(x), f(y)) \) for some \( f \in A \) and \( f^2 \in A \), 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
\[
0 = \det \begin{pmatrix} a & b \\ a^2 & b^2 \end{pmatrix} = ab^2 - a^2b = ab(b - a)
\]
which implies that \( a = b \). But this the implies that \( f(x) = f(y) \) for all \( f \in A \), violating the assumption that \( A \) separates points. Therefore we conclude that \( \dim(V) = 2 \), i.e. \( V = \mathbb{R}^2 \). \( \blacksquare \)

Theorem 13.12 (Stone-Weierstrass Theorem, Compact Case). Suppose \( X \) is a compact metric space and \( A \subseteq C(X, \mathbb{R}) \) is a closed subalgebra which separates points. For \( x \in X \) let
\[
A_x := \{f(x) : f \in A\}
\]
and
\[
I_x = \{f \in C(X, \mathbb{R}) : f(x) = 0\}.
\]
Then either one of the following two cases hold.
1. \( A = C(X, \mathbb{R}) \) or
2. there exists a unique point \( x_0 \in X \) such that \( A = I_{x_0} \).

Moreover, case 1. holds iff \( A_x = \mathbb{R} \) for all \( x \in X \) and case 2. holds iff there exists a point \( x_0 \in X \) such that \( A_{x_0} = \{0\} \).

**Proof.** If there exists \( x_0 \) such that \( A_{x_0} = \{0\} \) (\( x_0 \) is unique since \( A \) separates points) then \( A \subset I_{x_0} \). If such an \( x_0 \) exists let \( C = I_{x_0} \), and if \( A_x = \mathbb{R} \) for all \( x \), set \( C = C(X, \mathbb{R}) \). Let \( f \in C \) be given. By Lemma 13.11 for all \( x,y \in X \) such that \( x \neq y \), there exists \( g_{xy} \in A \) such that \( f = g_{xy} \) on \( \{x,y\} \). When \( X \) is compact the basic idea of the proof is contained in the following identity,

\[
f(z) = \inf_{x \in X} \sup_{y \in X} g_{xy}(z) \text{ for all } z \in X. \tag{13.7}
\]

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. (13.7) becoming only an approximate identity.

**Claim.** Given \( \epsilon > 0 \) and \( x \in X \) there exists \( g_x \in A \) such that \( g_x(x) = f(x) \) and \( f < g_x + \epsilon \) on \( X \).

To prove the claim, let \( V_y \) be an open neighborhood of \( y \) such that \( |f - g_{xy}| < \epsilon \) on \( V_y \) so in particular \( f < \epsilon + g_{xy} \) on \( V_y \). By compactness, there exists \( \Lambda \subset_f X \) such that \( X = \bigcup_{y \in \Lambda} V_y \). Set

\[ g_x(z) = \max\{g_{xy}(z) : y \in \Lambda\}, \]

then for any \( y \in \Lambda \), \( f < \epsilon + g_{xy} < \epsilon + g_x \) on \( V_y \) and therefore \( f < \epsilon + g_x \) on \( X \). Moreover, by construction \( f(x) = g_x(x) \), see Figure 13.2 below. This completes the proof of the claim and we now can complete the proof of the theorem.

For each \( x \in X \), let \( U_x \) be a neighborhood of \( x \) such that \( |f - g_x| < \epsilon \) on \( U_x \). Choose \( \Gamma \subset_f X \) such that \( X = \bigcup_{x \in \Gamma} U_x \) and define

\[ g = \min\{g_x : x \in \Gamma\} \in A. \]

Then \( f < g + \epsilon \) on \( X \) and for \( x \in \Gamma \),

\[ g \leq g_x < f + \epsilon \text{ on } U_x. \]

Since \( X = \bigcup_{x \in \Gamma} U_x \), we conclude

\[ f < g + \epsilon \text{ and } g < f + \epsilon \text{ on } X, \]

i.e. \( |f - g| < \epsilon \) on \( X \). Since \( \epsilon > 0 \) is arbitrary it follows that \( f \in \overline{A} = A \).

**Corollary 13.13 (Complex Stone-Weierstrass Theorem, Compact Case.).** Let \( X \) be a compact metric space. Suppose \( A \) is a subalgebra of \( C(X, \mathbb{C}) \) which is closed in the uniform topology, separates points, and is closed under complex conjugation. Then either \( A = C(X, \mathbb{C}) \) or \( A = I_{x_0}^C := \{f \in C(X, \mathbb{C}) : f(x_0) = 0\} \) for some \( x_0 \in X \).

**Proof.** Since

\[ \operatorname{Re} f = \frac{f + \bar{f}}{2} \text{ and } \operatorname{Im} f = \frac{f - \bar{f}}{2i}, \]

\( \operatorname{Re} f \) and \( \operatorname{Im} f \) are both in \( A \). Therefore

\[ A_\mathbb{R} = \{ \operatorname{Re} f, \operatorname{Im} f : f \in A \} \]

is a real sub-algebra of \( C(X, \mathbb{R}) \) which separates points. Therefore either \( A_\mathbb{R} = C(X, \mathbb{R}) \) or \( A_\mathbb{R} = I_{x_0}^C \cap C(X, \mathbb{R}) \) for some \( x_0 \) and hence \( A = C(X, \mathbb{C}) \) or \( I_{x_0}^C \) respectively.

Here are some applications of this theorem.

---

1 If \( A_{x_0} = \{0\} \) and \( x = x_0 \) or \( y = x_0 \), then \( g_{xy} \) exists merely by the fact that \( A \) separates points.
Example 13.14. 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 13.3. Let \((X, d)\) be a separable compact metric space. Show that \( C(X) \) is also separable. \textbf{Hint:} Let \( E \subseteq X \) be a countable dense set and then consider the algebra, \( \mathcal{A} \subseteq C(X) \), generated by \( \{d(x, \cdot)\}_{x \in E} \).

Theorem 13.15 (Weierstrass Approximation Theorem). Suppose that \( K \subseteq \mathbb{R}^d \) 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 \( K \).

Remark 13.16. 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
\[
q(z, \bar{z}) = p\left(\frac{z + \bar{z}}{2}, \frac{z - \bar{z}}{2i}\right).
\]
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 13.17 (Complex Weierstrass Approximation Theorem). Suppose that \( K \subseteq \mathbb{C}^d \) 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 \( n \to \infty \).

\textbf{Proof.} This is an immediate consequence of Theorem 13.15 and Remark 13.16. \( \square \)

Example 13.18. 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 \( C(S^1, \mathbb{C}) \). Since \( \bar{z} = z^{-1} \) on \( S^1 \), we have shown polynomials in \( z \) and \( z^{-1} \) are dense in \( C(S^1, \mathbb{C}) \). This example generalizes in an obvious way to \( K = (S^1)^d \subseteq \mathbb{C}^d \).

Exercise 13.4. Suppose \( f \in C(\mathbb{R}, \mathbb{C}) \) is a \( 2\pi \) - periodic function (i.e. \( f(x + 2\pi) = f(x) \) for all \( x \in \mathbb{R} \)) and
\[
\int_{0}^{2\pi} f(x) e^{inx} \, dx = 0 \text{ for all } n \in \mathbb{Z},
\]
show again that \( f \equiv 0 \).

\textbf{Exercise 13.5.} Given a continuous function \( f : \mathbb{R} \to \mathbb{C} \) which is \( 2\pi \) - 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 \( \theta \in \mathbb{R} \). \textbf{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 13.19. Exercise 13.15 generalizes to \( 2\pi \) - 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
\[
p(\theta) = \sum_{n \in \Gamma} \alpha_n e^{in\theta}
\]
where \( \Gamma \) is a finite subset of \( \mathbb{Z}^d \). The assertion is again that these trigonometric polynomials are dense in the \( 2\pi \) - periodic functions relative to the supremum norm.

13.3 Stone-Weierstrass Theorem for Locally Compact \( X \)

Definition 13.20. If \( A \) and \( B \) are two subsets of a metric space \((X, d)\) we let
\[
d(A, B) := \inf \{d(a, b) : a \in A \text{ and } b \in B\}
\]
be the distance between \( A \) and \( B \).

\textbf{Exercise 13.6.} Let \( K \) be a compact subset and \( F \) be a closed subset of a metric space, \((X, d)\). If \( K \cap F = \emptyset \), show \( d(K, F) > 0 \).

Definition 13.21. A metric space \((X, d)\) is \textbf{locally compact} if for all \( x \in X \) there exists \( V \subseteq X \) such that \( V \) is compact. \[\text{[We actually only need to use here the easier fact that } \bigcup_{x \in A} V_x \text{ is a compact set being the finite union of compact sets and that } V \text{ is a closed subset of } \bigcup_{x \in A} V_x.\]

Exercise 13.7. Let \((X, d)\) be a locally compact metric space and \( K \) be a compact subset of \( X \). Show there exists a precompact open set \( V \subseteq X \) such that \( K \subseteq V \). Use this result to show there exists an \( \varepsilon > 0 \) so that \( V_x(K) := \{x \in X : d_K(x) < \varepsilon\} \) is precompact.

\textsuperscript{2} Note that it is easy to extend \( f \in C(S^1) \) to a function \( F \in C(\mathbb{C}) \) by setting \( F(z) = |z| f\left(\frac{z}{|z|}\right) \) for \( z \neq 0 \) and \( F(0) = 0 \). So this special case does not require the Tietze extension theorem.

\textsuperscript{3} Recall that a subset \( V \) of a metric space is precompact if \( \bar{V} \) is compact.
Definition 13.22. If \((X, d)\) is a metric space and \(Z\) is a normed space, let \(C_c(X, Z)\) denote those \(f \in C(X, Z)\) with compact support. Recall that the support of \(f\) is defined to be \(\text{supp}(f) := \overline{f^{-1} \{ \{0\} \}}\).

The next corollary is a minor variant of Urysohn’s Lemma.

Corollary 13.23 (Locally Compact Urysohn’s Lemma). Let \((X, d)\) be a locally compact metric space and \(K\) be a compact subset of \(X\) and \(F\) be a closed subset of \(X\) such that \(K \cap F = \emptyset\). Then there exists \(f \in C_c(X, [0, 1])\) such that \(f = 1\) on \(K\) and \(f = 0\) on \(F\).

**Proof.** By Exercise 13.7, there exists \(0 < \varepsilon < \delta\) such that \(V_\varepsilon(K)\) is precompact. We then define

\[
\begin{align*}
f(x) &= \frac{d_{V_\varepsilon(K)^c}(x)}{d_K(x) + d_{V_\varepsilon(K)^c}(x)} \quad \text{for } x \in X.
\end{align*}
\]

This function satisfies \(f = 1\) on \(K\) and \(f = 0\) on \(V_\varepsilon(K)^c\) and hence \(f = 0\) on \(F \subset V_\varepsilon(K)\). Moreover we have \(\{f \neq 0\} \subset V_\varepsilon(K)\) and so \(\text{supp}(f)\) is a closed subset of the compact set, \(V_\varepsilon(K)\), and hence is itself compact. ■

Definition 13.24. If \((X, d)\) is a locally compact metric space we let \(C_0(X, Z)\) denote those \(f \in C(X, Z)\) satisfying \(K_x := \{x \in X : ||f(x)|| \geq \varepsilon\}\) is compact for all \(\varepsilon > 0\). If this condition holds we say \(f\) vanishes at infinity.

Theorem 13.25. \(C_0(X, Z)\) is a closed subspace of \(BC(X, Z)\) and \(C_c(X, Z)\) is a dense subspace of \(C_0(X, Z)\).

**Proof.** Suppose that \(\{f_n\} \subset C_0(X, Z)\) and \(f_n\) converges uniformly to \(f \in BC(X, Z)\). Let \(\varepsilon > 0\) be given and choose \(n\) so that \(||f - f_n||_\infty < \varepsilon/2\). Then choose a compact set \(K_x = \{x \in X : ||f_n(x)|| \geq \varepsilon/2\}\). It then follows that, for all \(x \in X \setminus K_x\),

\[
||f(x)|| \leq ||f_n(x)|| + ||f(x) - f_n(x)|| \\
\leq ||f_n(x)|| + ||f - f_n||_\infty < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\]

Therefore \(\{x \in X : ||f(x)|| \geq \varepsilon\} \subset K_x\) which shows \(\{x \in X : ||f(x)|| \geq \varepsilon\}\) is compact and hence \(f\) vanishes at infinity, i.e. \(f \in C_0(X, Z)\).

For the second assertion, let \(f \in C_0(X, Z)\) and suppose that \(\varepsilon > 0\) is given. Let \(K_x := \{x \in X : ||f(x)|| \geq \varepsilon\}\) and use Corollary 13.23 to find \(\varphi \in C_c(X, [0, 1])\) such that \(\varphi = 1\) on \(K_x\). Then the function, \(f_\varepsilon := \varphi \cdot f \in C_c(X, Z)\) and

\[
||f(x) - f_\varepsilon(x)|| = ||f(x)|| - ||f_\varepsilon(x)|| = \begin{cases} 0 & \text{if } x \in K_x \\ < \varepsilon & \text{if } x \notin K_x. \end{cases}
\]

Thus it follows that \(||f - f_\varepsilon||_\infty \leq \varepsilon\) and since \(\varepsilon > 0\) was arbitrary this suffices to show \(C_c(X, Z)\) is dense in \(C_0(X, Z)\).

Theorem 13.26 (Stone-Weierstrass Theorem, Locally Compact Case). Suppose \((X, d)\) is a locally compact metric space and \(A \subset C_0(X, \mathbb{R})\) is a closed subalgebra which separates points. For \(x \in X\) let

\[
A_x := \{f(x) : f \in A\}
\]

\[
\mathcal{I}_x = \{f \in C_0(X, \mathbb{R}) : f(x) = 0\}.
\]

Then either one of the following two cases hold.

1. \(A = C_0(X, \mathbb{R})\) or
2. there exists a unique point \(x_0 \in X\) such that \(A_{x_0} = \{0\}\).

**Proof.** If there exists \(x_0\) such that \(A_{x_0} = \{0\}\) \((x_0\) is unique since \(A\) separates points) then \(A \subset \mathcal{I}_{x_0}\). If such an \(x_0\) exists let \(C = \mathcal{I}_{x_0}\) and if \(A_{x_0} = \mathbb{R}\) for all \(x\), set \(C = C_0(X, \mathbb{R})\). Let \(f \in C\) be given. By Lemma 13.11 for all \(x, y \in X\) different \(x \neq y\), there exists \(g_{xy} \in A\) such that \(f = g_{xy} + \{x, y\}\). When \(X\) is compact the basic idea of the proof is contained in the following identity,

\[
f(z) = \inf_{x \in X} \sup_{y \in X} g_{xy}(z) \quad \text{for all } z \in X. \tag{13.8}
\]

To prove this identity, let \(g_z := \sup_{y \in X} g_{xy}\) and notice that \(g_z \geq f\) since \(g_{xy}(y) = f(y)\) for all \(y \in X\). Moreover, \(g_z(x) = f(x)\) for all \(x \in X\) since \(g_{xy}(x) = f(x)\) for all \(x\). Therefore,

\[
\inf_{x \in X} \sup_{y \in X} g_{xy} = \inf_{x \in X} g_z = f.
\]

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. \((13.8)\) becoming only an approximate identity. We also have to modify Eq. \((13.8)\) slightly to take care of the non-compact case.

Claim. Given \(\varepsilon > 0\) and \(x \in X\) there exists \(g_x \in A\) 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 \(g_{x,*}\) be any fixed element in \(A\) such that \(g_{x,*}(x) = f(x)\) and let

\footnote{The reason we suppose that \(X\) is locally compact is so that we know that \(C_0(X, \mathbb{R})\) is a non-trivial subspace of \(C(X, \mathbb{R})\).}

\footnote{If \(A_{x_0} = \{0\}\) and \(x = x_0\) or \(y = x_0\), then \(g_{xy}\) exists merely by the fact that \(A\) separates points.}
\[ K = \left\{ |f| \geq \frac{\varepsilon}{2} \right\} \cup \left\{ |g_{x,*}| \geq \frac{\varepsilon}{2} \right\}. \]  

(13.9)

Since \( K \) is compact, there exists \( A \subset f K \) such that \( K \subset \bigcup_{y \in A} V_y \). Define

\[ g_x(z) = \max\{g_{xy}(z) : y \in A \cup \{\ast\}\}. \]

Since

\[ f < \varepsilon + g_{xy} < \varepsilon + g_x \text{ on } V_y, \]

for any \( y \in A \), and

\[ f < \frac{\varepsilon}{2} < \varepsilon + g_{x,*} \leq g_x + \varepsilon \text{ on } K^c, \]

\( f < \varepsilon + g_x \) on \( X \) and by construction \( f(x) = g_x(x) \), see Figure 13.3. This completes the proof of the claim.

Fig. 13.3. 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 \( f < g_* + \varepsilon \) on \( X \); for example let \( g_* = g_{x_0} \in \mathcal{A} \) 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

\[ \Gamma \subset_F F := \left\{ |f| \geq \frac{\varepsilon}{2} \right\} \cup \left\{ |g_*| \geq \frac{\varepsilon}{2} \right\} \]

such that \( F \subset \bigcup_{x \in \Gamma} U_x \) (\( \Gamma \) exists since \( F \) is compact) and define

\[ g = \min\{g_x : x \in \Gamma \cup \{\ast\}\} \in \mathcal{A}. \]

Then, for \( x \in F \), \( g_x < f + \varepsilon \) on \( U_x \) and hence \( g < f + \varepsilon \) on \( \bigcup_{x \in F} U_x \supset F \). Likewise,

\[ g \leq g_* < \varepsilon/2 < f + \varepsilon \text{ on } F^c. \]

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 \( f \in \mathcal{A} = \mathcal{A} \) and so \( \mathcal{A} = \mathcal{C} \).

Corollary 13.27 (Complex Stone-Weierstrass Theorem, Locally Compact Case). Let \( X \) be a locally compact metric space. Suppose \( \mathcal{A} \) is a subalgebra of \( C_0(X, \mathbb{C}) \) which is closed in the uniform topology, separates points, and is closed under complex conjugation. Then either \( \mathcal{A} = C_0(X, \mathbb{C}) \) or

\[ \mathcal{A} = \mathcal{T}_{x_0}^C := \{ f \in C_0(X, \mathbb{C}) : f(x_0) = 0 \} \]

for some \( x_0 \in X \).

Proof. Since

\[ \text{Re } f = \frac{f + \bar{f}}{2} \text{ and } \text{Im } f = \frac{f - \bar{f}}{2i}, \]

\( \text{Re } f \) and \( \text{Im } f \) are both in \( \mathcal{A} \). Therefore

\[ \mathcal{A}_R = \{ \text{Re } f, \text{Im } f : f \in \mathcal{A} \} \]

is a real sub-algebra of \( C_0(X, \mathbb{R}) \) which separates points. Therefore either \( \mathcal{A}_R = C_0(X, \mathbb{R}) \) or \( \mathcal{A}_R = \mathcal{T}_{x_0} \cap C_0(X, \mathbb{R}) \) for some \( x_0 \) and hence \( \mathcal{A} = C_0(X, \mathbb{C}) \) or \( \mathcal{T}_{x_0}^C \) respectively.

Example 13.28. 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, \( \mathcal{A} = C_0(X, \mathbb{R}) \).

Exercise 13.8. Suppose that \( g \in C([0, \infty), \mathbb{R}) \) is a function such that

\[ \|g\|_1 := \int_0^\infty |g(x)| \, dx := \lim_{M \to \infty} \int_0^M |g(x)| \, dx < \infty. \]

Show:

1. \( \int_0^\infty f(x)g(x) \, dx := \lim_{M \to \infty} \int_0^M f(x)g(x) \, dx \) exists for all \( f \in BC([0, \infty)). \)
2. Also show
\[ \left| \int_0^\infty f(x) g(x) \, dx \right| \leq \|f\|_u \cdot \|g\|_1 \text{ for all } f \in BC \left([0, \infty)\right). \]

Thus it follows if \( f_n \in BC \left([0, \infty)\right) \) converges uniformly to \( f \in BC \left([0, \infty)\right) \) then
\[ \int_0^\infty g(x) f(x) \, dx = \lim_{n \to \infty} \int_0^\infty g(x) f_n(x) \, dx. \quad (13.10) \]

3. Observe (but do not turn in) that the map, \( BC \left([0, \infty)\right) \ni f \mapsto \int_0^\infty f(x) g(x) \, dx \in \mathbb{R} \) is linear.

**Definition 13.29 (Laplace Transform).** Suppose that \( g \in C \left([0, \infty), \mathbb{R}\right) \) is a function such that
\[ \int_0^\infty |g(t)| e^{-at} \, dt < \infty \text{ for some } a \in (0, \infty). \]

Then for \( \lambda \in (a, \infty) \) we define \( (\mathcal{L}g)(\lambda) := \int_0^\infty g(t) e^{-\lambda t} \, dt \) and refer to \( \mathcal{L}g \) as the **Laplace transform** of \( g \).

**Theorem 13.30 (Injectivity of the Laplace Transform).** Continuing the notation in **Definition 13.29**. If \( \mathcal{L}g(\lambda) = 0 \) for all \( \lambda > a \) then \( g(t) = 0 \) for all \( t \geq 0 \).

**Proof.** Let \( \lambda > 0 \) be fixed. From Example 13.28
\[ \mathcal{A} = \left\{ p\left(e^{-\lambda t}\right) : p \text{ is a polynomial w/o constant term} \right\} \]
\[ = \text{span} \left\{ e^{-n\lambda t} : n \in \mathbb{N} \right\} \]

is dense in \( C_0 \left([0, \infty)\right)\). By assumption,
\[ \int_0^\infty g(t) e^{-at} e^{-n\lambda t} \, dt = \int_0^\infty g(t) e^{-(n\lambda + a)t} \, dt = 0 \]
and so by the linearity of the integral it follows that
\[ \int_0^\infty g(t) e^{-at} f(t) \, dt = 0 \text{ for all } f \in \mathcal{A}. \]

Then form the continuity properties of the integral in Eq. (13.10) and the fact that \( \mathcal{A} = C_0 \left([0, \infty)\right) \) we may conclude that
\[ \int_0^\infty g(t) e^{-at} f(t) \, dt = 0 \text{ for all } f \in C_0 \left([0, \infty)\right). \]

**Exercise 13.9.** Suppose that \( g \in C \left([0, \infty)\right) \) such that
\[ \int_0^\infty |g(x)| \, dx := \lim_{M \to \infty} \int_0^M |g(x)| \, dx < \infty \]
and
\[ \int_0^\infty g(x) \left(\frac{1}{1+x}\right)^n \, dx := \lim_{M \to \infty} \int_0^M g(x) \left(\frac{1}{1+x}\right)^n \, dx = 0 \text{ for } n \in \mathbb{N}. \quad (13.11) \]

Show \( g(x) = 0 \) for all \( x \geq 0 \).
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. $\langle af + bg | h \rangle = a \langle f | h \rangle + b \langle g | h \rangle$ i.e. $f \rightarrow \langle f | h \rangle$ is linear.
2. $\langle f | g \rangle = \overline{\langle g | f \rangle}$.
3. $\|f\|^2 := \langle f | f \rangle \geq 0$ with equality $\|f\|^2 = 0$ iff $f = 0$.

Notice that combining properties (1) and (2) that $f \rightarrow \langle h | f \rangle$ is conjugate linear for fixed $h \in H$, i.e.

$$\langle h | af + bg \rangle = \overline{a} \langle h | f \rangle + \overline{b} \langle h | g \rangle.$$  

The following identity will be used frequently in the sequel without further mention,

$$\|f + g\|^2 = \langle f + g | f + g \rangle = \|f\|^2 + \|g\|^2 + \langle f | g \rangle + \langle g | f \rangle = \|f\|^2 + \|g\|^2 + 2 \text{Re} \langle f | g \rangle.$$  \hfill (14.1)

Theorem 14.2 (Schwarz Inequality). Let $(H, \langle \cdot | \cdot \rangle)$ be an inner product space, then for all $f, g \in H$

$$|\langle f | g \rangle| \leq \|f\| \|g\|$$

and equality holds iff $f$ and $g$ are linearly dependent.

Proof. If $g = 0$, the result holds trivially. So assume that $g \neq 0$ and observe; if $f = \alpha g$ for some $\alpha \in \mathbb{C}$, then $\langle f | g \rangle = |\alpha| \|g\|^2$ and hence

$$|\langle f | g \rangle| = |\alpha| \|g\|^2 = \|f\| \|g\|.$$  

Now suppose that $f \in H$ is arbitrary, let $h := f - \|f\|^{-2} \langle f | g \rangle g$. (So $h$ is the “orthogonal projection” of $f$ onto $g$, see Figure 14.1) Then

$$0 \leq \|h\|^2 = \left\| f - \frac{\langle f | g \rangle}{\|g\|^2} g \right\| = \|f\|^2 + \left\| \frac{\langle f | g \rangle}{\|g\|^2} g \right\|^2 - 2 \text{Re} \left( \frac{\langle f | g \rangle}{\|g\|^2} \right) \langle f | g \rangle g$$

$$= \|f\|^2 - \left\| \frac{\langle f | g \rangle}{\|g\|^2} g \right\|^2$$

from which it follows that $0 \leq \|g\|^2 \|f\|^2 - |\langle f | g \rangle|^2$ with equality iff $h = 0$ or equivalently iff $f = \|g\|^{-2} \langle f | g \rangle g$. \hfill $\blacksquare$

Corollary 14.3. Let $(H, \langle \cdot | \cdot \rangle)$ be an inner product space and $\|f\| = \|f\|_H := \sqrt{\langle f | f \rangle}$. Then the Hilbert 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 $f, g \in H$, then, using Schwarz’s inequality,

$$\|f + g\|^2 = \|f\|^2 + \|g\|^2 + 2 \text{Re} \langle f | g \rangle \leq \|f\|^2 + \|g\|^2 + 2 \|f\| \|g\| = (\|f\| + \|g\|)^2.$$  

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. The continuity of the inner product follows from Theorem 14.2 as in Exercise ???. For completeness, if $f, g, \Delta f, \Delta g \in H$, then

$$\left| \langle f + \Delta f | g + \Delta g \rangle - \langle f | g \rangle \right| = |\langle f | \Delta g \rangle + \langle \Delta f | g \rangle + \langle \Delta f | \Delta g \rangle|$$

$$\leq \|f\| \|\Delta g\| + \|g\| \|\Delta f\| + \|\Delta g\| \|\Delta f\|$$

with the latter expression clearly going to zero as $\Delta f$ and $\Delta g$ go to zero proving the inner product is continuous. \hfill $\blacksquare$

Definition 14.4. An inner product space $(H, \langle \cdot | \cdot \rangle)$ is a Hilbert space if $H$ is complete in the $\| \cdot \|_H$, i.e. $(H, \| \cdot \|_H)$ is a Banach space.

Unfortunately for the purposes of this course our inner product spaces will typically not be complete because we are using Riemann integration theory rather than Lebesgue integration theory. Here is our one and only example of a Hilbert space.
Example 14.5 (Space of square summable sequences). Let \( l^2 \) denote those sequences \( a = (a_1, a_2, a_3, \ldots) \) with \( a_i \in \mathbb{C} \) such that \( \|a\|_2^2 := \sum_{n=1}^{\infty} |a_n|^2 < \infty \). We then define \( (a|b) := \sum_{n=1}^{\infty} a_n b_n \) for all \( a, b \in l^2 \). Applying the Cauchy-Schwarz inequality for finite sequences we find,

\[
\sum_{n=1}^{N} |a_n b_n| \leq \sqrt{\sum_{n=1}^{N} |a_n|^2} \sqrt{\sum_{n=1}^{N} |b_n|^2} \leq \|a\|_2 \cdot \|b\|_2.
\]

Letting \( N \to \infty \) then shows

\[
\sum_{n=1}^{\infty} |a_n b_n| \leq \|a\|_2 \cdot \|b\|_2 < \infty \text{ for all } a, b \in l^2.
\]

Using this remark, one may show that \( (l^2, \langle \cdot | \cdot \rangle) \) is an inner product space.

The proof that \( l^2 \) is complete follows the familiar completeness argument which we now.

Let \( \{a(k)\}_{k=1}^{\infty} \subset l^2 \) be a Cauchy sequence, i.e.

\[
\|a(k) - a(l)\|_2^2 = \sum_{n=1}^{\infty} |a_n(k) - a_n(l)|^2 \to 0 \text{ as } k, l \to \infty.
\]

This implies that for each \( n \in \mathbb{N} \) that \( \{a(k)\}_{k=1}^{\infty} \in \mathbb{C} \) is a Cauchy sequence in \( \mathbb{C} \) and hence convergent. Let \( a_n := \lim_{k \to \infty} a_n(k) \in \mathbb{C} \) for all \( n \in \mathbb{N} \). Then for any \( N \in \mathbb{N} \),

\[
\sum_{n=1}^{N} |a_n(k) - a_n|^2 = \lim_{l \to \infty} \sum_{n=1}^{N} |a_n(k) - a_n(l)|^2 \\
\leq \limsup_{l \to \infty} \sum_{n=1}^{\infty} |a_n(k) - a_n(l)|^2 = \limsup_{l \to \infty} \|a(k) - a(l)\|_2^2.
\]

Letting \( N \to \infty \) in this inequality then shows,

\[
\|a(k) - a(l)\|_2^2 \leq \limsup_{l \to \infty} \|a(k) - a(l)\|_2^2 \to 0 \text{ as } k \to \infty.
\]

Definition 14.6. Let \( (H, \langle \cdot | \cdot \rangle) \) be an inner product space, we say \( f, g \in H \) are orthogonal and write \( f \perp g \) iff \( \langle f|g \rangle = 0 \). More generally if \( A \subset H \) is a set, \( f \in H \) is orthogonal to \( A \) (write \( f \perp A \)) iff \( \langle f|g \rangle = 0 \) for all \( g \in A \). Let \( A^\perp = \{f \in H : f \perp A\} \) be the set of vectors orthogonal to \( A \). A subset \( S \subset H \) is an orthogonal set if \( f \perp g \) for all distinct elements \( f, g \in S \). If \( S \) further satisfies, \( \|f\| = 1 \) for all \( f \in S \), then \( S \) is said to be an orthonormal set.

Remark 14.7. For any subset, \( V \) of \( H \) we have \( V \cap V^\perp \subset \{0\} \). Indeed if \( f \in V \cap V^\perp \), then \( \|f\|^2 = \langle f|f \rangle = 0 \).

Proposition 14.8. Let \( (H, \langle \cdot | \cdot \rangle) \) be an inner product space then

1. (Parallelogram Law)

\[
\|f + g\|^2 + \|f - g\|^2 = 2\|f\|^2 + 2\|g\|^2
\]

(14.2)

for all \( f, g \in H \).

2. (Pythagorean Theorem) If \( S \subset F \) is a finite orthogonal set, then

\[
\left\| \sum_{f \in S} f \right\|^2 = \sum_{f \in S} \|f\|^2.
\]

(14.3)

3. If \( A \subset H \) is a set, then \( A^\perp \) is a closed linear subspace of \( H \).

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;

\[
\|f + g\|^2 + \|f - g\|^2 = |\langle f|g \rangle|^2 + \|f\|^2 + \|g\|^2 - 2Re\langle f|g \rangle \\
= 2\|f\|^2 + 2\|g\|^2,
\]

and

\[
\left\| \sum_{f \in S} f \right\|^2 = \sum_{f \in S} \sum_{g \in S} \langle f|g \rangle = \sum_{f, g \in S} \langle f|g \rangle \\
= \sum_{f \in S} \|f\|^2.
\]

Item 3. is a consequence of the continuity of \( \langle \cdot | \cdot \rangle \) and the fact that

\[
A^\perp = \cap_{f \in A} \text{Nul}(\langle \cdot|f \rangle)
\]

where \( \text{Nul}(\langle \cdot|f \rangle) = \{g \in H : \langle g|f \rangle = 0\} \) – a closed subspace of \( H \).

Definition 14.9. If \( A \subset H \) is a finite orthonormal set, let \( P_A : H \to H \) be the linear operator defined by

\[
P_A f := \sum_{\varphi \in A} \langle f|\varphi \rangle \varphi \in V := \text{span} A.
\]

We refer to \( P_A \) to the orthogonal projection determined by \( A \).
Proposition 14.10 (Properties of $P_A$). Let $A$ be a finite orthonormal subset of $H$ and $V = \text{span} A$. For all $f, g \in H$ the following properties hold:

1. $\langle P_A f | P_A g \rangle = \langle f | P_A g \rangle = \sum_{\varphi \in A} \langle f | \varphi \rangle \langle \varphi | g \rangle$.
2. $\| P_A f \|^2 = \sum_{\varphi \in A} |\langle f | \varphi \rangle|^2$
3. $(f - P_A f) \perp V$
4. $\| f \|^2 = \| f - P_A f \|^2 + \| P_A f \|^2$.
5. $P_A f = f$ for $f \in V$ and $P_A f = 0$ for $f \in V^\perp$, and
6. $P_A^2 = P_A$.

Moreover if $A'$ is another orthonormal basis for $V$ then $P_{A'} = P_A$.

Proof. These follow by standard linear algebra direct calculations. We supply the proof for the reader convenience.

1. For $f, g \in H$ we have,

$$\langle P_A f | g \rangle = \left( \sum_{\varphi \in A} \langle f | \varphi \rangle \varphi \right) \langle \varphi | g \rangle = \sum_{\varphi \in A} \langle f | \varphi \rangle \langle \varphi | g \rangle$$

and similarly on shows $\langle f | P_A g \rangle = \sum_{\varphi \in A} \langle f | \varphi \rangle \langle \varphi | g \rangle$. Taking $f = \varphi_0 \in A$ in this equation shows $\langle \varphi_0 | P_A g \rangle = \langle \varphi_0 | g \rangle$ and using this result in Eq. (14.4) shows,

$$\sum_{\varphi \in A} \langle f | \varphi \rangle \langle \varphi | g \rangle = \sum_{\varphi \in A} \langle f | \varphi \rangle \langle \varphi | P_A g \rangle = \langle P_A f | P_A g \rangle$$

wherein we have used Eq. (14.4) again for the second equality.

2. Take $f = g$ in item 1.

3. Taking $g = \varphi \in A$ in Eq. (14.4) shows $\langle P_A f | \varphi \rangle = \langle f | \varphi \rangle$. This is equivalent to $\langle \varphi | f - P_A f \rangle = 0$ for all $\varphi \in A$ which is equivalent to $(f - P_A f) \perp V$.

4. This follows from Pythagorean’s theorem as $(P_A f) \perp (f - P_A f)$ since $P_A f \in V$ while $(f - P_A f) \in V$.

5. If $f \in V^\perp$ then $\langle \varphi | f \rangle = 0$ for all $\varphi \in A \subset V$ and therefore $P_A f = 0$. If $f \in V$, then $f - P_A f \in V \cap V^\perp$ and hence $f - P_A f = 0$, i.e. $f = P_A f$.

6. This follows from item 5. since for $f \in H$, $P_A f \in V$ and therefore $P_A (P_A f) = P_A f$.

For the last assertion, observe that $(P_A f - P_{A'} f) = (f - P_A f) - (f - P_{A'} f)$ from which it follows that $(P_A f - P_{A'} f) \in V \cap V^\perp = \{0\}$, i.e. $P_A f = P_{A'} f$.

Proposition 14.11 (Gram–Schmidt process). Every finite dimensional subspace, $V \subset H$, has an orthonormal basis $A$.

Proof. Let $\{v_1, \ldots, v_n\}$ be a basis for $V$ and set $u_1 := v_1 / \|v_1\|$. Then define $u_k$ inductively by setting

$$\tilde{v}_k = v_k - P_{\{u_1, \ldots, u_{k-1}\}} v_k$$

and then setting $u_k := \tilde{v}_k / \|\tilde{v}_k\|$.

Definition 14.12. If $V$ is a finite dimensional subspace of $H$ we let $P_V = P_A$ where $A$ is any orthonormal basis for $V$.

Corollary 14.13 (Best Approximation Theorem). If $V$ is a finite dimensional subspace of $H$, then

$$\| f - P_V f \| = \inf_{g \in V} \| f - g \| \text{ for all } f \in H.$$

In words, $P_V f \in V$ is the best approximation to $f$ by an element in $V$.

Proof. If $g \in V$,

$$f - g = (f - P_V f) + (P_V f - g)$$

where $(f - P_V f) \in V^\perp$ and $(P_V f - g) \in V$. Therefore by Pythagorean’s theorem,

$$\| f - g \|^2 = \| f - P_V f \|^2 + \| P_V f - g \|^2 \geq \| f - P_V f \|^2.$$

Proposition 14.14 (Bessel’s Inequality). If $T$ is an orthonormal subset of $H$, then

$$\sum_{\varphi \in T} |\langle f | \varphi \rangle|^2 \leq \sup_{A \subset \overline{T}} \sum_{\varphi \in A} |\langle f | \varphi \rangle|^2 \leq \|f\|^2 \text{ for all } f \in H \tag{14.5}$$

and

$$\sum_{\varphi \in T} |\langle f | \varphi \rangle| \leq \sup_{A \subset \overline{T}} \sum_{\varphi \in A} |\langle f | \varphi \rangle| \leq \|f\| \|g\| \text{ for all } f, g \in H. \tag{14.6}$$

In particular, the sum, $\sum_{\varphi \in T} |\langle f | \varphi \rangle|$, is absolutely convergent.

Proof. Let $A \subset f T$ be any finite set. Then from Proposition 14.10

$$\sum_{\varphi \in A} |\langle f | \varphi \rangle|^2 \leq \|P_A f\|^2 = \|f\|^2 - \|f - P_A f\|^2 \leq \|f\|^2.$$

This shows shows $\|f\|^2$ is an upper bound $\left\{ \sum_{\varphi \in A} |\langle f | \varphi \rangle|^2 : A \subset f T \right\}$ and proves Eq. (14.5). From the Cauchy Schwarz inequality and Eq. (14.5),
Definition 14.15. If $\beta \subset H$ is a countable (for simplicity) orthonormal set and $f \in H$ we write
\[
  f = \sum_{\varphi \in \beta} \langle f | \varphi \rangle \varphi
\]
iff for every choice of $A_n \subset f_\beta$ such that $A_n \uparrow \beta$, we have
\[
  \lim_{n \to \infty} \sum_{\varphi \in A_n} \langle f | \varphi \rangle \varphi = f \quad \text{in } H, \quad \text{i.e.} \quad \lim_{n \to \infty} P_{A_n} f = f \quad \text{in } H.
\]

Theorem 14.16 (Orthonormal Basis Theorem). Let $(H, \langle \cdot | \cdot \rangle)$ be a not necessarily complete inner product space and $\beta \subset H$ be a countable (for simplicity) orthonormal set. Then the following three conditions are equivalent:

1. $f = \sum_{\varphi \in \beta} \langle f | \varphi \rangle \varphi$ for all $f \in H$.
2. There exists $A_n \subset \beta$ such that $A_n \uparrow \beta$ and $\lim_{n \to \infty} \sum_{\varphi \in A_n} \langle f | \varphi \rangle \varphi = f$ for all $f \in H$.
3. $\langle f, g \rangle = \sum_{\varphi \in \beta} \langle f | \varphi \rangle \langle \varphi | g \rangle$ for all $f, g \in H$.
4. $|\langle f | \varphi \rangle|^2 = \sum_{\psi \in \beta} |\langle f | \psi \rangle|^2$ for all $f \in H$.

From Proposition 14.14 the sums in items 2. and 3. are absolutely convergent and so how we order the terms in the sum is not important.

Proof. Certainly 1. implies 2. If 2. holds then $P_{A_n} f \to f$ as $n \to \infty$ and therefore,
\[
  \langle f | g \rangle = \lim_{n \to \infty} \langle P_{A_n} f | g \rangle = \lim_{n \to \infty} \left( \sum_{\varphi \in A_n} \langle f | \varphi_n \rangle \varphi_n | g \rangle \right)
  = \lim_{n \to \infty} \sum_{\varphi \in A_n} \langle f | \varphi_n \rangle \langle \varphi_n | g \rangle = \sum_{\varphi \in \beta} \langle f | \varphi \rangle \langle \varphi | g \rangle
\]
which is item 3. Item 3. implies 4. by taking $g = f$. So it only remains to prove 4. implies 1.

Suppose that $A_n \subset_f \beta$ with $A_n \uparrow \beta$ is given. Then
\[
  \|f - P_{A_n} f\|^2 = \|f\|^2 - \|P_{A_n} f\|^2
  = \|f\|^2 - \sum_{\varphi \in A_n} |\langle f | \varphi \rangle|^2 \to \|f\|^2 - \sum_{\varphi \in \beta} |\langle f | \varphi \rangle|^2 = 0
\]
which verifies that 4. $\implies$ 1. \hfill \blacksquare

Definition 14.17. A countable orthonormal subset $\beta \subset H$ is an orthonormal basis for $H$ if any one of the equivalent conditions in Theorem 14.16 are satisfied.

Definition 14.18. A subset $T \subset H$ is total if span$(T)$ is a dense subspace of $H$.

Corollary 14.19. If $T = \{u_1, u_2, u_3, \ldots\}$ is a total subspace of $H$, then applying the Gram–Schmidt process to $T$, see the proof of Proposition 14.11, produces an orthonormal basis $\beta$ for $H$. Note the span $\beta = \text{span } T$.

Proof. Let us assume $\dim H = \infty$ which is the case of interest to us. The Gram-Schmidt orthogonalization process produces an orthonormal set $\beta = \{\varphi_1, \varphi_2, \ldots\}$ and as sequence $N_n \in \mathbb{N}$ such that $N_n \uparrow \infty$ and span $\{\varphi_1, \ldots, \varphi_n\} = \text{span } \{u_1, \ldots, u_{N_n}\}$. It is possible that that $N_n > n$ because the set $T$ is not assumed to be linearly independent. By assumption, for any $f \in H$ there exists an $N \in \mathbb{N}$ and $(a_1, \ldots, a_N) \in \mathbb{C}^N$ such that
\[
  \left\| f - \sum_{k=1}^{N} a_k u_k \right\| \leq \varepsilon.
\]
As $\text{span } \{u_1, \ldots, u_N\} \subset \text{span } \{\varphi_1, \ldots, \varphi_M\}$ for any $M \geq N$ we may conclude that
\[
  \left\| f - P_M f \right\| \leq \left\| f - \sum_{k=1}^{N} a_k u_k \right\| \leq \varepsilon
\]
where $A_M := \{\varphi_1, \ldots, \varphi_M\}$. This shows that $\lim_{M \to \infty} P_M f = f$ and hence $\beta$ is an orthonormal basis for $H$. \hfill \blacksquare

Theorem 14.20 (Orthogonal Polynomials). Let $-\infty < a < b < \infty$ and $\rho \in C([a,b];[0,\infty))$ such that $\rho$ has only finitely many zeros. Let $H = C([a,b], \mathbb{C})$ with
\[
  \langle f | g \rangle = \int_a^b f(x) \overline{g(x)} \rho(x) \, dx.
\]
Then $H$ is an inner product space and applying the Gram–Schmidt process to the functions, $T := \{1, x, x^2, x^3, \ldots\}$, produces an orthonormal basis $\beta = \{p_n\}_{n=0}^{\infty}$ for $H$. Here $p_n(x)$ is a polynomial with $\deg p_n = n$ for all $n$.

Proof. By the Weierstrass approximation Theorem 13.6 or by Corollary 13.13 we know that $T$ is a total subset of $C([a,b], \mathbb{C})$ when given the supremum norm. That is for any $f \in H$, there exists polynomials, $q_n(x)$, such that $\|f - q_n\|_{\infty} = |f(x) - q_n(x)| = 0$. Since
where \( \|\rho\|_1 = \int_a^b \rho(x) \, dx \), it follows that
\[
\|f - q_n\|_H^2 \leq \|\rho\|_1 \cdot \|f - q_n\|_\infty^2 \to 0 \text{ as } n \to \infty.
\]
This shows that \( T \) is total in \( H \) with the norm coming from the inner product. Since the the set \( T \) is linearly independent, the Gram-Schmidt process never produces any zeros. Moreover, it is easily seen by induction that \( T \) is an algebraic basis for the polynomials of degree \( n \) and hence \( p_{n+1} \) must be a polynomial of degree \( n + 1 \).

**Example 14.21 (Jacobi Polynomials).** Let \( a = -1, b = 1, \) and \( \rho(x) = (1 - x)^{\alpha}(1 + x)^{\beta} \) where \( \alpha, \beta > -1 \). The associated polynomial are the Jacobi polynomials, \( p_{n}(\alpha, \beta)(x) \), see the [Wiki Article](http://www.example.com). When \( \alpha \) or \( \beta \) are less than 0 it is not longer true that \( \rho \) is continuous at the end points. In this case one has to interpret the integrals as improper integrals. Never the less the condition \( \alpha, \beta > -1 \) insure that \( \|\rho\|_1 = \int_a^b \rho(x) \, dx < \infty \) which is what is really important in the proof of Theorem 14.20.

**Exercise 14.1 (Legendre Polynomials).** When \( \alpha = \beta = 0 \) (so that \( \rho \equiv 1 \)) in Example 14.21 the resulting polynomials are multiples of the Legendre Polynomials. Apply Gram-Schmidt process to find the first four Legendre polynomials, i.e. find \( p_0, p_1, p_2 \) and \( p_3 \).

**Example 14.22 (Hermite polynomials).** Let
\[
H = \left\{ f \in C(\mathbb{R}, \mathbb{C}) : \int_{-\infty}^{\infty} |f(x)|^2 e^{-\frac{1}{2}x^2} \, dx < \infty \right\}
\]
and for \( f, g \in H \) let
\[
\langle f | g \rangle = \int_{-\infty}^{\infty} f(x) \bar{g}(x) e^{-\frac{1}{2}x^2} \, dx.
\]
It can be shown that \( T := \{1, x, x^2, x^3 \ldots \} \) is total in \( H \). Applying the Gram-Schmidt procedure to \( T \) gives orthonormal basis, \( \beta \), for \( H \) of polynomial functions on \( \mathbb{R} \). The resulting polynomials are multiples of the so called "Hermite polynomials."

**Remark 14.23 (An Interesting Phenomena).** Let \( H = L^2([-1,1],dm) \) and \( S := \{1, x^3, x^6, x^9, \ldots \} \). Then again \( S \) is total in \( H \) by the same argument as in Theorem 14.20. This is true even though \( S \) is a proper subset of \( T \). Notice that \( T \) is an algebraic basis for the polynomials on \([-1,1]\) while \( S \) 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
\[
\int_{-1}^{1} |f(x)|^2 \, dx = \int_{-1}^{1} \left| f(y^{1/3}) \right|^2 \frac{1}{3} y^{-2/3} \, dy = \int_{-1}^{1} \left| f(y^{1/3}) \right|^2 \, d\mu(y) \tag{14.7}
\]
where \( d\mu(y) = \frac{1}{3} y^{-2/3} \, dy \). Since \( \mu([-1,1]) = m([-1,1]) = 2 \), \( \mu \) is a finite measure on \([-1,1]\) and hence \( T := \{1, x, x^2, x^3, \ldots \} \) is total 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. (14.7) 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
\[
\varepsilon > \|g - p\|_\infty = \sup \{|g(y) - p(y)| : y \in [-1,1]\} = \sup \{|g(x^3) - p(x^3)| : x \in [-1,1]\} = \sup \{|f(x) - p(x^3)| : x \in [-1,1]\}.
\]
This gives another proof the polynomials in \( x^3 \) are dense in \( C([-1,1]) \) and hence in \( L^2([-1,1]) \).

We will see another important example of Corollary 14.19 when we study Fourier series in the next chapter.
Fourier Series

In this chapter we let $H$ be the continuous functions on $[-\pi, \pi]$ with periodic boundary conditions. That is we let

$$H = C_p([-\pi, \pi], \mathbb{C}) := \{ f \in C([-\pi, \pi], \mathbb{C}) : f(\pi) = f(-\pi) \}.$$ 

We equip $H$ with the inner product,

$$\langle f | g \rangle := \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) \bar{g}(\theta) d\theta.$$ 

Let $C_p(\mathbb{R}, \mathbb{C})$ be the space of $2\pi$-periodic functions, i.e. those $F \in C(\mathbb{R}, \mathbb{C})$ such that $F(x + 2\pi) = F(x)$ for all $x \in \mathbb{R}$. The reader is asked to check that to each $f \in C_p([-\pi, \pi], \mathbb{C})$ there is a unique function $F \in C_p(\mathbb{R}, \mathbb{C})$ such that $f = F|_{[-\pi, \pi]}$. In the future we will simply identify $C_p([-\pi, \pi], \mathbb{C})$ with $C_p(\mathbb{R}, \mathbb{C})$.

15.1 Fourier Basis

**Lemma 15.1.** The functions, $\beta := \{ \chi_n(\theta) := e^{in\theta} : n \in \mathbb{Z} \}$ form an orthonormal subset of $H$.

**Proof.** This is simple computation,

$$\langle \chi_n | \chi_m \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{in\theta} e^{-im\theta} d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(n-m)\theta} d\theta = \begin{cases} 1 & \text{if } m = n \text{ and } n \neq 0, \\ \frac{1}{2\pi} e^{i(n-m)\theta} |_{-\pi} & \text{if } m \neq n \neq 0. \end{cases}$$

Let $S^1 := \{ z \in \mathbb{C} : |z| = 1 \}$ be the unit circle in $\mathbb{C}$.

**Theorem 15.2 (Fourier Basis).** The function $\beta := \{ \chi_n(\theta) := e^{in\theta} : n \in \mathbb{Z} \}$ form an orthonormal basis for $H$.

**Proof.** By the same method of proof of Theorem 14.20 it suffices to show that $\beta$ is total in $C_p([-\pi, \pi], \mathbb{C})$ when equipped with the supremum norm. By Proposition 15.5 below, every $f \in C_p([-\pi, \pi], \mathbb{C})$ may be written at $f(\theta) = F(e^{i\theta})$ where $F \in C(S^1, \mathbb{C})$. Let $\tilde{\beta} := \{ \tilde{\chi}_n(z) := z^n : n \in \mathbb{Z} \} \subset C(S^1, \mathbb{C})$. Then $A := \text{span}_{\mathbb{C}} \tilde{\beta}$ is a complex algebra inside of $C(S^1, \mathbb{C})$ which is stable under complex conjugation, contains the function 1, and separates points. [The function $\tilde{\chi}_1(z) = z$ already separates points.] Therefore by the Stone Weierstrass Theorem (see Corollary 13.17), we know that $A$ is dense in $C(S^1, \mathbb{C})$ equipped with the supremum norm. In particular if $\varepsilon > 0$ is given there exists

$$P(z) = \sum_{k=-N}^{N} a_k z^k \sup_{z \in S^1} |F(z) - P(z)| \leq \varepsilon.$$ 

Letting

$$p(\theta) := P(e^{i\theta}) = \sum_{k=-N}^{N} a_k e^{ik\theta} = \sum_{k=-N}^{N} a_k \chi_k(\theta)$$

(which we refer to as a trigonometric polynomial), we have

$$\sup_{\theta \in [-\pi, \pi]} |f(\theta) - p(\theta)| = \sup_{\theta \in [-\pi, \pi]} |F(e^{i\theta}) - P(e^{i\theta})| = \sup_{z \in S^1} |F(z) - P(z)| \leq \varepsilon.$$ 

This shows that $\beta$ is total in $C_p([-\pi, \pi], \mathbb{C})$ with the sup-norm and hence with the $L^2$ norm as well.

For the proof of Proposition 15.5 below we will need a couple of lemmas.

**Lemma 15.3.** If $\psi : [-\pi, \pi] \rightarrow S^1 \subset \mathbb{C}$ is defined by $\psi(\theta) = e^{i\theta}$ for $\theta \in [-\pi, \pi]$, then $\psi|_{(-\pi, \pi)} : (-\pi, \pi) \rightarrow S^1 \setminus \{ -1 \}$ is a homeomorphism.

**Proof.** The map $\psi$ is continuous and $\psi|_{(-\pi, \pi)} : (-\pi, \pi) \rightarrow S^1 \setminus \{ -1 \}$ is a bijection and hence we must show $\psi|_{(-\pi, \pi)}$ is continuous. To this end, suppose that $\theta_n \in (-\pi, \pi)$ and $z_n = e^{i\theta_n} \in S^1 \setminus \{ -1 \}$ and $z_n \rightarrow z = e^{i\theta} \in S^1 \setminus \{ -1 \}$ with $\theta \in (-\pi, \pi)$. We wish to show $\theta_n \rightarrow \theta$ as $n \rightarrow \infty$. If not there exists an $\varepsilon > 0$ and a subsequence $\alpha_k := \theta_n$ such that $|\alpha_k - \theta| \geq \varepsilon$ for all $k$. By passing to a further subsequence if necessary we may assume that $\lim_{k \rightarrow \infty} \alpha_k = \alpha$ exists in $[-\pi, \pi]$. It then follows that

$$e^{i\theta} = \lim_{k \rightarrow \infty} z_{\alpha_k} = \lim_{k \rightarrow \infty} e^{i\alpha_k} = e^{i\alpha}.$$
However this shows that $\alpha = \theta$ and therefore $\varepsilon \leq \lim_{k \to \infty} |\alpha_k - \theta| = 0$ which is a contradiction.

You are asked to give another proof of this Exercises [15.1 and 15.2]

**Lemma 15.4.** If $\theta_n \in (-\pi, \pi)$ and $z_n = e^{i\theta_n} \in S^1 \setminus \{-1\}$ and $z_n \to -1$, then $d(\theta_n, \{\pm \pi\}) := |\theta_n - \pi| \wedge |\theta_n + \pi| \to 0$ as $n \to \infty$.

**Proof.** If $d(\theta_n, \{\pm \pi\}) \to 0$ as $n \to \infty$ there exists an $\varepsilon > 0$ and a subsequence $\alpha_k := \theta_{n_k}$ such that $d(\alpha_k, \{\pm \pi\}) \geq \varepsilon$ for all $k$. By passing to a further subsequence if necessary we may assume that $\lim_{k \to \infty} \alpha_k = \alpha$ exists in $[-\pi, \pi]$. It then follows that

$$-1 = \lim_{k \to \infty} z_{n_k} = \lim_{k \to \infty} e^{i\alpha_k} = e^{i\alpha}$$

so that $\alpha \in \{\pm \pi\}$ and we would conclude that

$$\varepsilon \leq \lim_{k \to \infty} d(\alpha_k, \{\pm \pi\}) = d(\alpha, \{\pm \pi\}) = 0$$

which is a contradiction.

**Proposition 15.5.** To every $f \in C_p([-\pi, \pi], \mathbb{C})$ there exists a unique function $F \in C(S^1, \mathbb{C})$ such that $f(\theta) = F(e^{i\theta})$.

**Proof.** Given $f : [-\pi, \pi] \to \mathbb{C}$ such that $f(\pi) = f(-\pi)$, let $F(z) := f(\psi^{-1}(z))$ which is well defined even for $z = -1$ since $\psi^{-1}(-1) = \{\pm \pi\}$ and $f(\pi) = f(-\pi)$. Then $F$ is the unique function such that $f(\theta) = F(e^{i\theta})$ for all $\theta \in [-\pi, \pi]$. Let us further suppose that $f$ is continuous. Making use of Lemmas 15.3 and 15.4, if $z_n = e^{i\theta_n} \to z = e^{i\theta}$ in $S^1$ then either $\theta_n \to \theta$ as $n \to \infty$ if $\theta \in (-\pi, \pi)$ or $d(\theta_n, \{\pm \pi\}) \to 0$ if $\theta \in \{\pm \pi\}$. In either case we have,

$$\lim_{n \to \infty} F(z_n) = \lim_{n \to \infty} f(\theta_n) = f(\theta) = F(z)$$

which shows that $F$ is continuous.

**Exercise 15.2.** Continuing the notation in Exercise 15.1 let

$$f(\theta) := \sigma(\psi(\theta)) = \sigma(e^{i\theta}) = \frac{\sin \theta}{1 + \cos \theta}.$$  

Show $f : (-\pi, \pi) \to \mathbb{R}$ is a homeomorphism that is smooth with a smooth inverse. Hence it follows that $\psi = \sigma^{-1} \circ f : (-\pi, \pi) \to S^1 \setminus \{-1\}$ is a homeomorphism being the composition of two homeomorphisms.

**Notation 15.6** For $f \in C([-\pi, \pi], \mathbb{C})$ or more generally for $f$ Riemann integrable on $[-\pi, \pi]$, let

$$\tilde{f}(n) := \langle f, \chi_n \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-in\theta} d\theta.$$  

We say $\tilde{f}(n)$ is the $n$th – Fourier coefficient of $f$. Also for $N \in \mathbb{N}_0$, let

$$P_N f = \sum_{n=-N}^{N} \tilde{f}(n) \chi_n$$

be orthogonal projection onto span $\{\chi_n : |n| \leq N\}$.

**Theorem 15.7.** Suppose that $f : [-\pi, \pi] \to \mathbb{C}$ is Riemann integrable and there exists $f_n \in C_p([-\pi, \pi], \mathbb{C})$ such that $\|f - f_n\| \to 0$, then
\[ \lim_{N \to \infty} \| P_N f - f \|_2 = 0, \text{ and} \]
\[ \|f\|_2^2 = \sum_{n \in \mathbb{Z}} |\langle f | \chi_n \rangle|^2 = \sum_{n \in \mathbb{Z}} |\tilde{f}(n)|^2. \]

Moreover if \( f, g \) and two such functions, then
\[ \langle f | g \rangle = \sum_{n \in \mathbb{Z}} \langle f | \chi_n \rangle \langle \chi_n | g \rangle = \sum_{n \in \mathbb{Z}} \tilde{f}(n) \overline{g(n)}. \]

**Proof.** Using the triangle inequality,
\[ \| P_N f - f \|_2 \leq \| P_N f - P_N f_n \|_2 + \| P_N f_n - f_n \|_2 + \| f_n - f \|_2 \]
\[ \leq \| f - f_n \|_2 + \| P_N f_n - f_n \|_2 + \| f_n - f \|_2 \]

and therefore,
\[ \limsup_{N \to \infty} \| P_N f - f \|_2 \leq 2 \| f_n - f \|_2 \to 0 \text{ as } n \to \infty. \]

It follows that
\[ \langle f | g \rangle = \lim_{N \to \infty} \langle P_N f | P_N g \rangle = \lim_{N \to \infty} \sum_{|n| \leq N} \langle f | \chi_n \rangle \langle \chi_n | g \rangle = \sum_{n \in \mathbb{Z}} \langle f | \chi_n \rangle \langle \chi_n | g \rangle \]
where the sum is absolutely convergent. 

**Example 15.8.** Let \( f(\theta) := 1_{[0,\pi]}(\theta) \) and \( f_n(\theta) \) be the function which is one on \( \left[ \frac{1}{n}, \pi - \frac{1}{n} \right] \) and goes linearly to zero on the intervals \( [0, \frac{1}{n}] \) and \( [\pi - \frac{1}{n}, \pi] \) and is zero on \( [-\pi, 0] \). Then
\[ \| f_n - f \|_2^2 = 2 \int_0^{\frac{1}{n}} |1 - f_n(\theta)|^2 d\theta \leq 2 \cdot \frac{1}{n} \to 0 \text{ as } n \to \infty \]

and since \( f_n \in C_{[\pi, \pi]} \), Theorem 15.7 is applicable. The Fourier coefficients of \( f \) are
\[ \tilde{f}(0) = \langle f | \chi_0 \rangle = \frac{1}{2\pi} \int_0^\pi d\theta = \frac{1}{2} \]

and for \( n \neq 0 \),
\[ \tilde{f}(n) = \langle f | \chi_n \rangle = \frac{1}{2\pi} \int_0^\pi e^{-in\theta} d\theta = \frac{1}{2\pi} \int \left. e^{-in\theta} \right|_0^\pi \]
\[ = \frac{1}{2\pi n} (1 - e^{-in\pi}) = \frac{1}{\pi ni} 1_{n \text{ odd}}(n). \]

Thus we find,
\[ f(\theta) = \frac{1}{2} + \frac{1}{\pi} \sum_{n \in \mathbb{N}_{\text{odd}}} \frac{1}{n} (e^{in\theta} - e^{-in\theta}) \]
\[ = \frac{1}{2} + \frac{2}{\pi} \sum_{n=1}^\infty \frac{1}{n} \sin(n\theta) = s_N(\theta) \]

where
\[ s_N(\theta) := \frac{1}{2} + \frac{2}{\pi} \sum_{n=1}^\infty \frac{1}{2n+1} \sin((2n+1)\theta) \]

and we write \( \tilde{\sim} \) to emphasize that convergence takes place in the \( L^2 \) sense. See Figure 15.2 where the reader will likely be convinced that \( \lim_{N \to \infty} S_N(\theta) \neq f(\theta) \) for \( \theta \in \{0, \pm \pi\} \). It is only guaranteed that
\[ \lim_{N \to \infty} \int_0^\pi |f(\theta) - S_N(\theta)|^2 d\theta = 0. \]

We also have from Theorem 15.7 that
Exercise 15.5. Suppose that \( f \in C([\pi, \pi], \mathbb{C}) \). Show there exists \( f_n \in C_p([\pi, \pi], \mathbb{C}) \) such that \( \|f - f_n\|_2 \to 0 \) as \( n \to \infty \).

Exercise 15.4. Show that if \( f \in C([\pi, \pi], \mathbb{C}) \) and \( \tilde{f}(k) = 0 \) for all \( k \) then \( f \equiv 0 \).

Exercise 15.6 (Smoothness implies decay). Suppose \( m \in \mathbb{N} \) and \( f \in C^m_p(\mathbb{R}) \) - the 2\( \pi \) - periodic functions on \( \mathbb{R} \) such that \( f^{(l)} \) exists and is continuous for \( 0 \leq l \leq m \).

1. Use integration by parts to show \( \tilde{f}^{(l)}(k) = (f^{(l)})(\chi_k) = (ik)^l \tilde{f}(k) \) for all \( k \in \mathbb{Z} \).

Note: This equality implies \( \|\tilde{f}(k)\| \leq \frac{1}{k^l} \|f^{(l)}\|_H \leq \frac{1}{k^l} \|f^{(l)}\|_\infty \) for \( k \neq 0 \).

2. Show for \( f \in C^1_p(\mathbb{R}) \) that

\[
\sum_{k \in \mathbb{Z} \setminus \{0\}} \left| \tilde{f}(k) \right| = \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{1}{|k|} \left| \tilde{f}'(k) \right| \leq \frac{1}{\sqrt{3}} \pi \cdot \|f'\|_H < \infty. \tag{15.1}
\]

Hint: use the Cauchy–Schwarz inequality for sums, see Theorem 6.16 or Theorem 14.2 applied with \( \langle a|b \rangle = \sum_{n=1}^{\infty} a_n \bar{b}_n \) for sequence \( a \) and \( b \).

Exercise 15.7. Suppose \( f \in C_{per}([-\pi, \pi]) \) is a function such that \( \sum_{n \in \mathbb{Z}} \left| \hat{f}(n) \right| < \infty \) and set

\[
g(x) := \sum_{k \in \mathbb{Z}} \hat{f}(k)e^{ikx} \text{ (pointwise)}.\]

1. Show \( g \in C_{per}(\mathbb{R}) \).

2. Show \( g(x) = f(x) \) for all \( x \). Hint: Show \( \hat{g}(k) = \tilde{f}(k) \) and apply Exercise 15.3.

Corollary 15.9 (Pointwise Convergence). If \( f \in C^1_p(\mathbb{R}) \) then \( \sum_{k \in \mathbb{Z}} \hat{f}(k)e^{ik\theta} \) converges pointwise to \( f(\theta) \) for all \( \theta \in \mathbb{R} \). Moreover the sum is absolutely and uniformly convergent.

**Proof.** This result follows directly from Exercises 15.6 and 15.7.

Exercise 15.8 (Decay implies smoothness). Suppose \( s \in \mathbb{R} \) and \( \{c_k \in \mathbb{C} : k \in \mathbb{Z}\} \) are coefficients such that

\[
\sum_{k \in \mathbb{Z}} |c_k|^2 (1 + |k|)^s < \infty.
\]

Show if \( s > \frac{1}{2} + m \), the function \( f \) defined by

\[
f(x) = \sum_{k \in \mathbb{Z}} c_ke^{ikx}
\]

is in \( C^m_{per}(\mathbb{R}) \). Hint: First show

\[
\sum_{k \in \mathbb{Z}} |c_k||k|^l < \infty \text{ for all } l \leq m.
\]

This is perhaps the simplest version of a “Sobolev Imbedding Theorem.”

Exercise 15.9 (Poisson Summation Formula). Let \( F : \mathbb{R} \to \mathbb{C} \) be a continuous function such that there exists \( C < \infty \) and \( s > 1 \) such that

\[
|F(x)| \leq C(1 + |x|)^{-s} \text{ for all } x \in \mathbb{R}.
\]

Also define the Fourier transform, \( \hat{F} : \mathbb{R} \to \mathbb{C} \), of \( F \) by

\[
\hat{F}(\xi) := \frac{1}{2\pi} \int_{\mathbb{R}} F(x)e^{-i\xi x}dx = \lim_{M \to \infty} \frac{1}{2\pi} \int_{-M}^{M} F(x)e^{-i\xi x}dx.
\]

We further assume that \( \sum_{k \in \mathbb{Z}} |\hat{F}(k)| < \infty \). [This can be achieved by assuming \( F \) is twice continuously differentiable with the derivatives decaying fast enough so that the resulting functions are integrable.]

1. Show \( \sum_{k \in \mathbb{Z}} F(x + 2\pi k) \) is absolutely convergent for all \( x \in \mathbb{R} \) and the function,

\[
f(x) := \sum_{k \in \mathbb{Z}} F(x + 2\pi k)
\]

is in \( C^1_p(\mathbb{R}, \mathbb{C}) \).

2. Show \( \hat{f}(k) = \hat{F}(k) \) for all \( k \in \mathbb{Z} \).
Exercise 15.10 (Heat Equation 1.). Let \((t, x) \in [0, \infty) \times \mathbb{R} \to u(t, x)\) be a continuous function such that \(u(t, \cdot) \in C_{\text{per}}(\mathbb{R})\) for all \(t \geq 0\). \(\dot{u} := u_t, u_{xx}\) and \(u_{xxx}\) exists and are continuous when \(t > 0\). Further assume that \(u\) satisfies the heat equation \(\dot{u} = \frac{1}{2} u_{xx}\). Let \(\tilde{u}(t, k) := \langle u(t, \cdot) | \chi_k \rangle\) 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 \(u(t, x) = \sum_{k \in \mathbb{Z}} e^{-\frac{i}{2}k^2 t} \tilde{u}(k) e^{ikx}\) \(\tag{15.4}\)

where \(f(x) := u(0, x)\) and as above

\[
\tilde{f}(k) = \langle f | \chi_k \rangle = \int_{-\pi}^{\pi} f(y) e^{-iky} dy.
\]

Notice from Eq. \(15.4\) that \((t, x) \to u(t, x)\) is \(C^\infty\) for \(t > 0\).

Exercise 15.11 (Heat Equation 2.). Let \(q_t(x) := \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} e^{-\frac{1}{2}k^2 t} e^{ikx}\). Show that Eq. \(15.4\) 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 \(p_t(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t x^2}\). Also show \(u(t, x)\) may be written as

\[
u(t, x) = p_t * f(x) := \int_\mathbb{R} p_t(x-y) f(y) dy.
\]

Hint: To show \(q_t(x) = \sum_{k \in \mathbb{Z}} p_t(x + k2\pi)\), use the Poisson summation formula (Exercise 15.9) and the Gaussian integration identity,

\[
p_t(\omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} p_t(x) e^{i\omega x} dx = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \omega^2}.
\]

Equation \(15.5\) will be discussed later in the notes.

Exercise 15.12 (Wave Equation). Let \(u \in C^2(\mathbb{R} \times \mathbb{R})\) be such that \(u(t, \cdot) \in C_{\text{per}}(\mathbb{R})\) 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 \(\tilde{u}(t, k) := \langle u(t, \cdot), \chi_k \rangle\) for \(k \in \mathbb{Z}\) is twice continuously differentiable in \(t\) and \(\frac{d}{dt} \tilde{u}(t, k) = -k^2 \tilde{u}(t, k)/2\). Use this result to show

\[
u(t, x) = \sum_{k \in \mathbb{Z}} \left( \frac{1}{k} \cos(kt) + \frac{\sin kt}{k} \right) e^{ikx}
\]

with the sum converging absolutely. Also show that \(u(t, x)\) may be written as

\[
u(t, x) = \frac{1}{2} [f(x + t) + f(x - t)] + \frac{1}{2} \int_{-t}^{t} g(x + \tau) d\tau.
\]

Hint: To show Eq. \(15.6\) implies \(15.7\) use

\[
\cos kt = \frac{e^{ikt} + e^{-ikt}}{2},
\]

\[
\sin kt = \frac{e^{ikt} - e^{-ikt}}{2i},
\]

and

\[
\frac{e^{ikt} - e^{-ikt}}{ik} = \int_{-t}^{t} e^{ik(x+\tau)} d\tau.
\]

15.3 Dirichlet, Fejér and Kernels

For a Riemann integrable function, \(f : [-\pi, \pi] \to \mathbb{C}\), the sum

\[
f = \sum_{k \in \mathbb{Z}} \langle f | \chi_k \rangle \chi_k
\]

is guaranteed to converge relative to the Hilbertian norm on \(H\). It turns out that it need not converge pointwise even if \(f \in C_{\text{per}}(\mathbb{R})\). Nevertheless as we saw in Corollary 15.9 if \(f \in C^1_{\text{per}}(\mathbb{R}, \mathbb{C})\), then the sum in Eq. \(15.8\) will converge pointwise. We are going to discuss in pointwise convergence in a more detail here. In the process we will give a direct and constructive proof of the result in Exercise 15.5 see Theorem \(15.11\) below.

For \(n \in \mathbb{N}\) we will be considering,

\[
f_n(\theta) = \sum_{|k| \leq n} \tilde{f}(k) \chi_k(\theta) = \sum_{|k| \leq n} \frac{1}{2\pi} \int_{[-\pi, \pi]} f(x) e^{-ikx} dx \chi_k(\theta)
\]

\[
= \frac{1}{2\pi} \int_{[-\pi, \pi]} f(x) \sum_{|k| \leq n} e^{ik(\theta-x)} dx
\]

\[
= \frac{1}{2\pi} \int_{[-\pi, \pi]} f(x) D_n(\theta - x) dx
\]

\(\tag{15.9}\)
where
\[ D_n(\theta) := \sum_{k=-n}^{n} e^{ik\theta} \]
is called the **Dirichlet kernel**. Letting \( \alpha = e^{i\theta/2} \), we have
\[ D_n(\theta) = \sum_{k=-n}^{n} \alpha^k = \frac{\alpha^{2(n+1)} - \alpha^{-2n}}{\alpha^2 - 1} = \frac{\alpha^{2n+1} - \alpha^{-(2n+1)}}{\alpha - \alpha^{-1}} \]
\[ = \frac{2i\sin(n + \frac{1}{2})\theta}{2i\sin\frac{1}{2}\theta} = \frac{\sin(n + \frac{1}{2})\theta}{\sin\frac{1}{2}\theta}. \]
and therefore
\[ D_n(\theta) := \sum_{k=-n}^{n} e^{ik\theta} = \frac{\sin(n + \frac{1}{2})\theta}{\sin\frac{1}{2}\theta}, \]see Figure [15.3](image).

This is a plot of \( D_1 \) and \( D_{10} \).

with the understanding that the right side of this equation is \( 2n + 1 \) whenever \( \theta \in 2\pi\mathbb{Z} \).

**Theorem 15.10.** Suppose \( f \in C_p(\mathbb{R}, \mathbb{C}) \) (or just Riemann integrable) 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. [15.9].

**Proof.** Observe that
\[ \frac{1}{2\pi} \int_{[-\pi, \pi]} D_n(\theta - x)dx = \frac{1}{2\pi} \int_{[-\pi, \pi]} \sum_{|k| \leq n} e^{ik(\theta-x)}dx = 1 \]
and therefore,
\[ f_n(\theta) - f(\theta) = \frac{1}{2\pi} \int_{[-\pi, \pi]} [f(x) - f(\theta)] D_n(\theta - x)dx \]
\[ = \frac{1}{2\pi} \int_{[-\pi, \pi]} [f(x) - f(\theta - x)] D_n(x)dx \]
\[ = \frac{1}{2\pi} \int_{[-\pi, \pi]} \left[ \frac{f(\theta - x) - f(\theta)}{\sin \frac{x}{2}} \right] \sin(n + \frac{1}{2})x dx \]
(15.11)
where for \( x \in [-\pi, \pi] \),
\[ g(x) = \begin{cases} \frac{f(\theta-x) - f(\theta)}{\sin \frac{x}{2}} & \text{if } x \neq 0 \\ -2f'(\theta) & \text{if } x = 0 \end{cases} \]
The function \( g \) is continuous (Riemann integrable) on \([-\pi, \pi]\) and
\[ \frac{1}{2\pi} \int_{[-\pi, \pi]} g(x) \sin(n + \frac{1}{2})x dx \]
\[ = \frac{1}{4\pi i} \int_{[-\pi, \pi]} g(x) e^{ix/2} e^{ixn} dx - \frac{1}{4\pi i} \int_{[-\pi, \pi]} g(x) e^{-ix/2} e^{-ixn} dx \]
\[ = \frac{1}{2i} \left[ (x \to g(x) e^{ix/2}) \cdot (-n) - (x \to g(x) e^{-ix/2}) \cdot (n) \right]. \]
The last term goes to zero as \( n \to \infty \) since, by Bessel’s equality,
\[ \sum_{n \in \mathbb{Z}} \left| (x \to g(x) e^{-ix/2}) \cdot (n) \right|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |g(x)|^2 dx < \infty. \]

Despite the Dirichlet kernel not being positive, it still satisfies the approximate \( \delta - \)sequence property, \( \frac{1}{2\pi} D_n \to \delta_0 \) as \( n \to \infty \), when acting on \( C^1 \) - periodic functions in \( \theta \). 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,
\[ F_N(\theta) = \frac{1}{N+1} \sum_{n=0}^{N} f_n(\theta) = \frac{1}{N+1} \sum_{n=0}^{N} \frac{1}{2\pi} \int_{[-\pi, \pi]} f(x) \sum_{|k| \leq n} e^{ik(\theta-x)}dx \]
\[ = \frac{1}{2\pi} \int_{[-\pi, \pi]} K_N(\theta - x) f(x)dx \]
where
\[ K_N(\theta) := \frac{1}{N+1} \sum_{n=0}^{N} \sum_{|k| \leq n} e^{ik\theta} \] (15.12)
is the **Fejér kernel**.
Theorem 15.11. The Fejér kernel $K_N$ in Eq. (15.12) satisfies:

1. 

$$K_N(\theta) = \sum_{n=-N}^{N} \left[ 1 - \frac{|n|}{N+1} \right] e^{i n \theta} = \frac{1}{N+1} \frac{\sin^{2} \left( \frac{N+1}{2} \theta \right)}{\sin^{2} \left( \frac{\theta}{2} \right)}.$$  

(15.13)

2. $K_N(\theta) \geq 0$.

3. $\frac{1}{2\pi} \int_{-\pi}^{\pi} K_N(\theta)d\theta = 1$

4. $\sup_{\theta \in [-\pi, \pi]} K_N(\theta) \to 0$ as $N \to \infty$ for all $\varepsilon > 0$, see Figure 15.3.

5. For any continuous $2\pi$-periodic function $f$ on $\mathbb{R}$, $K_N * f(\theta) \to f(\theta)$ uniformly in $\theta$ as $N \to \infty$, where

$$K_N * f(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} K_N(\theta - \alpha)f(\alpha)d\alpha = \sum_{n=-N}^{N} \left[ 1 - \frac{|n|}{N+1} \right] \hat{f}(n) e^{i n \theta}. \quad (15.15)$$

![Figure 15.3. Plots of $K_N(\theta)$ for $N = 2, 7$ and 13.](image)

Proof. 1. Equation (15.13) is a consequence of the identity,

$$\sum_{n=0}^{N} \sum_{|k| \leq n} e^{i k \theta} = \sum_{|k| \leq n} (N+1 - |k|) e^{i k \theta}.$$ 

Moreover, letting $\alpha = e^{i \theta/2}$ and using Eq. (??) shows

$$K_N(\theta) = \frac{1}{N+1} \sum_{n=0}^{N} \sum_{|k| \leq n} \alpha^{2k} = \frac{1}{N+1} \sum_{n=0}^{N} \frac{\alpha^{2n+2} - \alpha^{-2n}}{\alpha^2 - 1}$$

$$= \frac{1}{(N+1)(\alpha - 1)} \sum_{n=0}^{N} \left[ \alpha^{2n+1} - \alpha^{-2n-1} \right]$$

$$= \frac{1}{(N+1)(\alpha - 1)} \sum_{n=0}^{N} \left[ \alpha^{2n+2} - 1 - \alpha^{-1} \frac{\alpha^{2N-2} - 1}{\alpha^2 - 1} \right]$$

$$= \frac{1}{(N+1)(\alpha - 1)^2} \left[ \alpha^{2(N+1)} - 1 + \alpha^{-2(N+1)} - 1 \right]$$

$$= \frac{1}{(N+1)(\alpha - 1)^2} \left[ \alpha^{(N+1)} - \alpha^{-(N+1)} \right]^2$$

$$= \frac{1}{N+1} \frac{\sin^2 \left( (N+1) \theta/2 \right)}{\sin^2 (\theta/2)}.$$
15.4 The Dirichlet Problems 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 15.12 (Dirichlet problem for \( D \)).** To every continuous function \( g \in C(\partial(D)) \) there exists a unique function \( u \in C(\bar{D}) \cap C^2(D) \) solving

\[
\Delta u(z) = 0 \quad \text{for } z \in D \quad \text{and} \quad u|_{\partial D} = g.
\]

Moreover for \( r < 1 \), \( u \) is given by,

\[
u(re^{i\theta}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_r(\theta - \alpha) u(e^{i\alpha}) d\alpha =: P_r * u(e^{i\theta})
\]

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. (15.16) is called the Dirichlet problem for \( D \).)

**Proof.** In this proof, we are going to be identifying \( S^1 = \partial(D) := \{ z \in D : |z| = 1 \} \) 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 \( \Delta \) 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 \( \partial_r u \left( re^{i\theta} \right) = \frac{\partial}{\partial r} u \left( re^{i\theta} \right) \) 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. (15.16) and let

\[
g(k) := \frac{1}{2\pi} \int_{-\pi}^{\pi} g(e^{ik\theta}) e^{-ik\theta} d\theta
\]

and

\[
\tilde{u}(r,k) := \frac{1}{2\pi} \int_{-\pi}^{\pi} u(re^{i\theta}) e^{-ik\theta} d\theta
\]

be the Fourier coefficients of \( g(\theta) \) and \( \theta \to u \left( re^{i\theta} \right) \) respectively. Then for \( r \in (0, 1) \),

\[
r^{-1} \partial_r \left( r \partial_r \tilde{u}(r,k) \right) = \frac{1}{2\pi} \int_{-\pi}^{\pi} r^{-1} \partial_r \left( r^{-1} \partial_r u \right) (re^{i\theta}) e^{-ik\theta} d\theta
\]

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. (15.20) 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. (15.21) 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 \( r \in (0, 1) \),

\[
A_k r^{|k|} = \tilde{u}(r,k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} u(re^{i\theta}) e^{-ik\theta} d\theta.
\]

Since all terms of this equation are continuous for \( r \in [0, 1] \), Eq. (15.22) 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. (15.16) then \( u \) must be given by

\[
u(re^{i\theta}) = \sum_{k \in \mathbb{Z}} \tilde{g}(k) r^{|k|} e^{ik\theta} \quad \text{for } r < 1.
\]

or equivalently,

\[
u(z) = \sum_{k \in \mathbb{N}_0} \tilde{g}(k) z^k + \sum_{k \in \mathbb{N}} \tilde{g}(-k) z^k.
\]

Notice that the theory of the Fourier series implies Eq. (15.23) is valid in the \( L^2(d\theta) \) - sense. However more is true, since for \( r < 1 \), the series in Eq. (15.23) is absolutely convergent and in fact defines a \( C^\infty \) – function (see Exercise ??)
or Corollary ??) which must agree with the continuous function, \( \theta \to u(re^{i\theta}) \), for almost every \( \theta \) and hence for all \( \theta \). This completes the proof of uniqueness.

**Existence.** Given \( g \in C \left( \text{bd}(D) \right) \), let \( u \) be defined as in Eq. \((15.23)\). Then, again by Exercise ?? or Corollary ??, \( u \in C^\infty(D) \). So to finish the proof it suffices to show \( \lim_{x \to y} u(x) = g(y) \) for all \( y \in \text{bd}(D) \). Inserting the formula for \( \tilde{g}(k) \) into Eq. \((15.23)\) gives

\[
 u(re^{i\theta}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_r(\theta - \alpha) u(e^{i\alpha})d\alpha \text{ for all } r < 1
\]

where

\[
P_r(\delta) = \sum_{k \in \mathbb{Z}} r^{|k|} e^{ik\delta} = \sum_{k=0}^{\infty} r^k e^{ik\delta} + \sum_{k=-\infty}^{-1} r^k e^{-ik\delta} - 1 =
\]

\[
= \text{Re} \left[ \frac{1}{1 - re^{i\delta}} - 1 \right] = \text{Re} \left[ \frac{1 + re^{i\delta}}{1 - re^{i\delta}} \right]
\]

\[
= \text{Re} \left[ \frac{(1 + re^{i\delta})(1 - re^{-i\delta})}{|1 - re^{i\delta}|^2} \right] = \text{Re} \left[ \frac{1 - r^2 + 2ir \sin \delta}{1 - 2r \cos \delta + r^2} \right]
\]

\[(15.24)\]

The Poisson kernel again solves the usual approximate \( \delta \)–function properties (see Figure ??), namely:

1. \( P_r(\delta) > 0 \) and

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} P_r(\theta - \alpha) d\alpha = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{k \in \mathbb{Z}} r^{|k|} e^{ik(\theta - \alpha)} d\alpha = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} r^{|k|} \int_{-\pi}^{\pi} e^{ik(\theta - \alpha)} d\alpha = 1
\]

and

2. \( \sup_{\varepsilon \leq |\theta| \leq \pi} P_r(\theta) \leq \frac{1 - r^2}{1 - 2r \cos \varepsilon + r^2} \to 0 \text{ as } r \uparrow 1. \)

A plot of \( P_r(\delta) \) for \( r = 0.2, 0.5 \text{ and } 0.7 \).

Therefore by the same argument used in the proof of Theorem \((15.11)\)

\[
\lim_{r \uparrow 1} \sup_{\theta} |u(re^{i\theta}) - g(e^{i\theta})| = \lim_{r \uparrow 1} \sup_{\theta} |(P_r \ast g)(e^{i\theta}) - g(e^{i\theta})| = 0
\]

which certainly implies \( \lim_{x \to y} u(x) = g(y) \) for all \( y \in \text{bd}(D) \).

\[\Box\]

**Remark 15.13 (Harmonic Conjugate).** Writing \( z = re^{i\theta} \), Eq. \((15.18)\) may be rewritten as

\[
u(z) = \frac{1}{2\pi} \text{Re} \int_{-\pi}^{\pi} \frac{1 + ze^{-i\alpha}}{1 - ze^{-i\alpha}} u(e^{i\alpha})d\alpha
\]

which shows \( u = \text{Re} F \) 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. \((15.24)\) that

\[
\text{Im} F(re^{i\theta}) = \frac{1}{\pi} \text{Im} \int_{-\pi}^{\pi} \frac{r \sin(\theta - \alpha)}{1 - 2r \cos(\theta - \alpha) + r^2} g(e^{i\alpha})d\alpha =: (Q_r \ast u)(e^{i\theta})
\]

where

\[
Q_r(\delta) := \frac{r \sin(\delta)}{1 - 2r \cos(\delta) + r^2}.
\]

From these remarks it follows that \( v =: (Q_r \ast g)(e^{i\theta}) \) is the harmonic conjugate of \( u \) and \( P_r = Q_r \). For more on this point see Section ?? below.
Arzela-Ascoli Theorem (Compactness in \( C(X) \))

In this section, let \((X, d)\) be a metric space, \(\tau \subset 2^X\) be the collection of open sets, and for \(x \in X\) let \(\tau_x = \{V \in \tau : x \in V\}\) be the collection of open neighborhoods of \(x\).

**Definition 16.1.** Suppose \(F \subset C(X)\), i.e. \(F\) is a collection of continuous functions.

1. \(F\) 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 \(f \in F\).
2. \(F\) is **equicontinuous** if \(F\) is equicontinuous at all points \(x \in X\).
3. \(F\) is **pointwise bounded** if \(\sup\{|f(x)| : f \in F\} < \infty\) for all \(x \in X\).

**Theorem 16.2 (Ascoli-Arzelà Theorem).** Let \((X, \tau)\) be a compact metric space and \(F \subset C(X)\). Then \(F\) is precompact in \(C(X)\) iff \(F\) is equicontinuous and pointwise bounded.

**Proof.** \((\Leftarrow)\) Let \(m \in \mathbb{N}\) and use the equicontinuity of \(F\) to conclude for all \(x \in X\), there exists \(V_x \in \tau_x\) such that \(|f(y) - f(x)| < \frac{1}{m}\) if \(y \in V_x\) and \(f \in F\).

Since \(X\) is compact we may choose \(A_m \subset F\) such that \(X = \bigcup_{x \in A_m} V_x\). Let \(\{x_k : k \in \mathbb{N}\}\) be an enumeration of the set, \(A := \bigcup_{m=1}^{\infty} A_m\).

Because of the pointwise boundedness assumption, we know \(\{f_m(x_k) : m \in \mathbb{N}\}\) has a convergent subsequence for every \(k \in \mathbb{N}\). By the Cantor diagonalization technique, we may find a subsequence \(g_l = f_{m_l}\) of \(\{f_m : m \in \mathbb{N}\}\) such that \(\lim_{l \to \infty} g_l(x)\) exists for all \(x \in A\).

The proof will be finished by showing \(\{g_l : l \in \mathbb{N}\}\) is in fact a uniformly convergent sequence. To prove this let \(y \in X\) and then choose \(x \in A_m\) such that \(y \in V_x\), with \(V_x\) as above. Then

\[
|g_l(y) - g_k(y)| \leq |g_l(y) - g_l(x)| + |g_l(x) - g_k(x)| + |g_k(x) - g_k(y)| \\
\leq \frac{2}{m} + |g_l(x) - g_k(x)| \leq \frac{2}{m} + \sup_{x \in A_m} |g_l(x) - g_k(x)| .
\]

Therefore,

\[
\lim_{l,k \to \infty} \sup_{y \in X} |g_l(y) - g_k(y)| \leq \frac{2}{m} + \lim_{l,k \to \infty} \sup_{y \in X} \sup_{x \in A_m} |g_l(x) - g_k(x)| \\
= \frac{2}{m} \to 0 \text{ as } m \to \infty.
\]

\((\Rightarrow)\) *(The rest of this proof may safely be skipped.) Let me first give the argument under the added restriction that \(\tau = \tau_d\) for some metric, \(d\), and \(X\). Since \(\|\cdot\|_\infty : C(X) \to [0, \infty)\) is a continuous function on \(C(X)\) it is bounded on any compact subset \(F \subset C(X)\). This shows that \(\sup\{|f| : f \in F\} < \infty\) which clearly implies that \(F\) is pointwise bounded.\(\square\)

Suppose \(F\) were not equicontinuous at some point \(x \in X\), i.e. there exists \(\varepsilon > 0\) such that for all \(V \in \tau_x\), sup \(\|f(x) - f(y)\| > \varepsilon\) for all \(f \in F\). By the assumption that \(F\) is not equicontinuous at \(x\), there exist \(f_n \in F\) and \(x_n \in V_n\) such that \(|f_n(x) - f_n(x_n)| \geq \varepsilon\) for all \(n\). Since \(F\) is a compact metric space by passing to a subsequence if necessary we may assume that \(f_n\) converges uniformly to some \(f \in F\). Because \(x_n \to x\) as \(n \to \infty\) we learn that

\[
\varepsilon \leq |f_n(x) - f_n(x_n)| \leq |f_n(x) - f(x)| + |f(x) - f(x_n)| + |f(x_n) - f_n(x_n)| \\
\leq 2\|f_n - f\|_\infty + |f(x) - f(x_n)| \to 0 \text{ as } n \to \infty
\]

which is a contradiction.

**Exercise 16.1.** 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 \((C([0, 1], \mathbb{R}), \|\cdot\|_\infty)\) which by definition means that \(\{Kf : \|f\|_\infty \leq 1\}\) is compact in \(C([0, 1], \mathbb{R})\).

**Exercise 16.2 (Continuation of Exercise 16.1).** Suppose \(k \in C\left([0, 1]^2, \mathbb{R}\right)\) but now show that one can use the formula for \(K\) in Eq. [16.1] in order to define a linear operator from \(L^2([0, 1], \mathbb{R})\) to \(C([0, 1], \mathbb{R})\) and that this operator is still compact, i.e. \(\{Kf : \|f\|_{L^2} \leq 1\}\) is compact in \(C([0, 1], \mathbb{R})\). Conclude that \(K\) is also a compact operator when viewed as an operator from \(L^2([0, 1], \mathbb{R})\) to \(L^2([0, 1], \mathbb{R})\).

\(^1\) One could also prove that \(F\) 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 \(x \in X\).
The following result is a corollary of Theorem \ref{16.2}.

**Corollary 16.3 (Locally Compact Ascoli-Arzelà Theorem).** Let \((X,d)\) be a metric space with the Heine Borel property, i.e., closed and bounded sets are compact. If \(\{f_m\} \subset C(X)\) is 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 uniformly convergent on all compact subsets of \(X\).

**Proof.** Let \(x \in X\) and \(\{K_n = C_x(n))\}_{n=1}^{\infty} \subset \{X\}\), so that \(K_n\) is a nested sequence of compact sets such that \(K_n \uparrow X\) as \(n \to \infty\). We may now apply Theorem \ref{16.2} repeatedly to find a nested family of subsequences

\[
\{f_m\} \supset \{g^1_m\} \supset \{g^2_m\} \supset \{g^3_m\} \supset \ldots
\]

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 \(K \subset K^l \subset K^l\) and therefore \(\{h_n\}\) is uniformly convergent on \(K\) as well. \(\blacksquare\)

### 16.0.1 Arzelà-Ascoli Theorem Problems

**Exercise 16.3.** Let \((X,\tau)\) be a compact metric space and \(\mathcal{F} := \{f_n\}_{n=1}^{\infty} \subset C(X)\) be a sequence of functions which are equicontinuous and pointwise convergent. Show \(f(x) := \lim_{n \to \infty} f_n(x)\) is continuous and that \(\lim_{n \to \infty} \|f - f_n\|_\infty = 0\), i.e., \(f_n \to f\) uniformly as \(n \to \infty\).

**Exercise 16.4.** Let \(T \subset (0,\infty)\) and \(\mathcal{F} \subset C([0,T],\mathbb{R})\) be a family of functions such that:

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 \(-\varepsilon \leq t \leq 0\) and

\[
x_\varepsilon(t) = x_0 + \int_0^t Z(\tau, x_\varepsilon(\tau - \varepsilon))d\tau \quad \text{for all } t \geq 0.
\]

Here

\[
\int_0^t Z(\tau, x_\varepsilon(\tau - \varepsilon))d\tau = \left(\int_0^t Z_1(\tau, x_\varepsilon(\tau - \varepsilon))d\tau, \ldots, \int_0^t Z_d(\tau, x_\varepsilon(\tau - \varepsilon))d\tau\right)
\]

where \(Z = (Z_1,\ldots,Z_d)\) 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. \ref{16.3} 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 \(t \in [-2\varepsilon,\varepsilon]\). The process can be repeated.

2. Then use Exercise 16.4 to show there exists \(\{\varepsilon_k\}_{k=1}^{\infty} \subset (0,\infty)\) such that \(\lim_{k \to \infty} \varepsilon_k = 0\) and \(x_{\varepsilon_k}\) 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

\[
\lim_{k \to \infty} \sup_{\tau \leq T} |x_{\varepsilon_k}(\tau - \varepsilon_k) - x(\tau)| = 0.
\]

3. Pass to the limit (with justification) in Eq. \ref{16.3} with \(\varepsilon\) replaced by \(\varepsilon_k\) to show \(x\) satisfies

\[
x(t) = x_0 + \int_0^t Z(\tau, x(\tau))d\tau \quad \forall \ t \in [0,T].
\]

4. Conclude from this that \(x(t)\) exists for \(t \in (0,T)\) and that \(x\) solves Eq. \ref{16.2}.

5. Apply what you have just proved to the ODE,

\[
y(t) = -Z(-t, y(t)) \quad \text{for } 0 \leq t < T \text{ with } y(0) = x_0.
\]

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. \ref{16.2} for \(t \in (-T,T)\).
Appendix: Notation and Logic

The following abbreviations along with their negations are used throughout these notes.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Negation</th>
</tr>
</thead>
<tbody>
<tr>
<td>∀</td>
<td>for all</td>
<td>∃</td>
</tr>
<tr>
<td>∃</td>
<td>there exists</td>
<td>∀</td>
</tr>
<tr>
<td>, or &quot;space&quot; then</td>
<td>∃</td>
<td>∀</td>
</tr>
<tr>
<td>∈</td>
<td>such that</td>
<td>∋</td>
</tr>
<tr>
<td>a.a.</td>
<td>almost always</td>
<td>i.o.</td>
</tr>
<tr>
<td>i.o.</td>
<td>infinitely often</td>
<td>a.a.</td>
</tr>
<tr>
<td>=</td>
<td>equals</td>
<td>≠</td>
</tr>
<tr>
<td>≠</td>
<td>not equals</td>
<td>=</td>
</tr>
<tr>
<td>≤</td>
<td>less than or equal</td>
<td>&gt;</td>
</tr>
<tr>
<td>&gt;</td>
<td>greater than</td>
<td>≤</td>
</tr>
</tbody>
</table>

Here are some examples.

1. \( a_n = b_n \) i.o. \( n \) \( \iff \) \( \# \{ n : a_n = b_n \} = \infty \). The negation of \( \# \{ n : a_n = b_n \} = \infty \) is \( \# \{ n : a_n = b_n \} < \infty \iff a_n \neq b_n \) for a.a. \( n \).

2. \( \lim_{n \to \infty} a_n = L \) is by definition the statement;
\[
\forall \varepsilon > 0 \ \exists N \in \mathbb{N} \ \forall n \geq N, \ |L - a_n| \leq \varepsilon.
\]
This may also be written as
\[
\forall \varepsilon > 0, \ |L - a_n| \leq \varepsilon \text{ for a.a. } n.
\]

3. The negation of the previous statement is \( \lim_{n \to \infty} a_n \neq L \) which translates to
\[
\exists \varepsilon > 0 \ \exists N \in \mathbb{N} \ \exists n \geq N \ \exists |L - a_n| > \varepsilon.
\]
This last statement is also equivalent to;
\[
\exists \varepsilon > 0 \ \exists |L - a_n| > \varepsilon \text{ i.o. } n.
\]
It is sometimes useful to reformulate this last statement as; there exists \( \varepsilon > 0 \) and an increasing function \( \mathbb{N} \ni k \to n_k \in \mathbb{N} \) such that
\[
|L - a_{n_k}| > \varepsilon \text{ for all } k \in \mathbb{N}.
\]
Appendix: More Set Theoretic Properties (highly optional)

B.1 Appendix: Zorn’s Lemma and the Hausdorff Maximal Principle (optional)

Definition B.1. A partial order \( \leq \) on \( X \) is a relation with following properties:

1. If \( x \leq y \) and \( y \leq z \) then \( x \leq z \).
2. If \( x \leq y \) and \( y \leq x \) then \( x = y \).
3. \( x \leq x \) 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, \( A \leq B \) if \( A \subset B \) 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 or 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

\[
X = \{ a = \{1\}, b = \{1, 2\}, c = \{3\}, d = \{2, 4\}, e = \{2\} \}
\]

ordered by set inclusion. Then \( b \) and \( d \) are maximal elements despite that fact that \( b \not\leq d \) and \( d \not\leq b \). We also have:

1. If \( E = \{a, c, e\} \), then \( E \) has no upper bound.
2. If \( E = \{a, e\} \), then \( b \) is an upper bound.
3. If \( E = \{e\} \), then \( b \) and \( d \) are upper bounds.

Theorem B.6. 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,” \( x : A \to \prod_{\alpha \in A} X_\alpha \) 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.

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 \( y = x \).

(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 \( g : D(g) \rightarrow \prod_{\alpha \in A} X_\alpha \) 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 \( x_0 \in X_{\alpha_0} \) 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 \( D(\Phi) = \bigcup_{g \in \Phi} D(g) \) and for \( \alpha \in D(\Phi) \) let \( h(\alpha) = g(\alpha) \). Then \( h \in \mathcal{G} \) is an upper bound for \( \Phi \). So by Zorn’s Lemma as a “jazzed” up version of induction.

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 \( E = \bigcup_{n=1}^{\infty} E_n \) 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 \( E_m \) 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 \( M \subset \mathcal{F} \) is linearly ordered, let \( E = \bigcup M \). If \( x, y \in E \) then \( x \in A \) and \( y \in B \) for some \( A, B \subset M \). Now \( M \) 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 ordered 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.
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 \( D(h) = D(h) \cup \{ \alpha_0 \} \) and

\[
\tilde{h}(\alpha) = \begin{cases} h(\alpha) & \text{if } \alpha \in D(h) \\ x_0 & \text{if } \alpha = \alpha_0. \end{cases}
\]

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 all 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 without loss of generality that \( \mathcal{F} = \{ P \in \mathcal{F} : P \text{ is not maximal} \} \) is a non-empty set.

For \( P \in \mathcal{F} \), let \( P^* = \{ x \in X : P \cup \{x\} \in \mathcal{F} \} \). As the above argument shows, \( P^* \neq \emptyset \) for all \( P \in \mathcal{F} \). Using the axiom of choice, there exists \( f \in \prod_{P \in \mathcal{F}} P^* \). We now define \( g : \mathcal{F} \to \mathcal{F} \) by

\[
g(P) = \begin{cases} P & \text{if } P \text{ is maximal} \\ P \cup \{f(x)\} & \text{if } P \text{ is not maximal.} \end{cases} \tag{B.1}
\]

The proof is completed by Lemma B.7 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.7.** 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 \( P_1 \in \mathcal{F} \) be chosen arbitrarily. Notice that \( \Phi = \{ g^n(P_0) \}_{n=0}^\infty \subset \mathcal{F} \) 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 \( P_\infty = \bigcup_{n=0}^\infty P_n \) and start again with \( P_0 \) replaced by \( P_\infty \). 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 \( \bigcup \Phi \in \mathcal{F}_1 \).
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

\[
\mathcal{F}_0 = \cap \{ \mathcal{F}' : \mathcal{F}' \text{ is a tower} \}.
\]

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 construct 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 \( P \equiv \bigcup \mathcal{F}_0 \) 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 \( \Gamma \) is a tower. First off it is clear that \( P_0 \in \Gamma \) so in particular \( \Gamma \) 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 \( A_\Phi := \bigcup \Phi \in \mathcal{F}_0 \). We must show \( A_\Phi \subset \Phi_C \), that is that \( A_\Phi \subset C \text{ or } C \subset A_\Phi \). Now if \( C \subset \Phi \) for all \( A \in \Phi \), then \( A_\Phi \subset C \text{ and hence } A_\Phi \subset \Phi_C \). 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 \( A_\Phi \subset \Phi_C \). 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. \[B.3\] holds, so \( \Phi_C \) is a tower. It is now easy to show \( \Gamma \) is a tower. It is again clear that \( P_0 \in \Gamma \) and Property 2. may be checked for \( \Gamma \) in the same way as it was done for \( \Phi_C \) 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 \( \Gamma \) is a tower and we are done.

Here is an example of using Zorn’s lemma.

**Proposition B.8.** Suppose that \( X \) and \( Y \) are non-empty sets, then either there exists an injective function, \( f : X \rightarrow Y \), or an injective function \( g : Y \rightarrow X \). In other words, either \( \text{card} (X) \leq \text{card} (Y) \) or \( \text{card} (Y) \leq \text{card} (X) \).

**Proof.** Let \( \mathcal{F} \) be the collection of injective functions, \( u : \mathcal{D}(u) \rightarrow Y \) where \( \mathcal{D}(u) \) is a non-empty subset of \( X \). We say that \( u \leq v \) for \( u,v \in \mathcal{F} \) provided \( \mathcal{D}(u) \subset \mathcal{D}(v) \) and \( u = v|_{\mathcal{D}(u)} \). One now checks that \( (\mathcal{F},\leq) \) is a partially ordered set such that every linearly ordered subset of \( \mathcal{F} \) has an upper bound. Therefore, by an application of Zorn’s lemma, \( \mathcal{F} \) has a maximal element, \( U \).

If \( \mathcal{D}(U) = X \), we take \( f = U \) and we have constructed an injective map from \( X \) to \( Y \). If \( \mathcal{D}(U) \neq X \), then \( \text{ Ran}(U) := U(\mathcal{D}(U)) = Y \). Indeed, if not we could find \( x \in X \setminus \mathcal{D}(U) \) and \( y \in Y \setminus \text{ Ran}(U) \) and then extend \( U \) to \( \mathcal{D}(U) \cup \{x\} \) by setting \( U(x) = y \). The extended \( U \) is still injective and hence would violate the maximality of \( U \). In this case we take \( g := U^{-1} : Y \rightarrow \mathcal{D}(U) \subset X \).

### B.2 Cardinality

In mathematics, the essence of counting a set and finding a result \( n \), is that it establishes a one to one correspondence (or bijection) of the set with the set of numbers \( \{ 1, 2, \ldots, n \} \). A fundamental fact, which can be proved by mathematical induction, is that no bijection can exist between \( \{ 1, 2, \ldots, n \} \) and \( \{ 1, 2, \ldots, m \} \) unless \( n = m \); this fact (together with the fact that two bijections can be composed to give another bijection) ensures that counting the same set in different ways can never result in different numbers (unless an error is made). This is the fundamental mathematical theorem that gives counting its purpose; however you count a (finite) set, the answer is the same. In a broader context, the theorem is an example of a theorem in the mathematical field of (finite) combinatorics—hence (finite) combinatorics is sometimes referred to as “the mathematics of counting.”

Many sets that arise in mathematics do not allow a bijection to be established with \( \{ 1, 2, \ldots, n \} \) for any natural number \( n \); these are called infinite sets, while those sets for which such a bijection does exist (for some \( n \)) are called finite sets. Infinite sets cannot be counted in the usual sense; for one thing, the mathematical theorems which underlie this usual sense for finite sets are false for infinite sets. Furthermore, different definitions of the concepts in terms of which these theorems are stated, while equivalent for finite sets, are inequivalent in the context of infinite sets.

The notion of counting may be extended to them in the sense of establishing (the existence of) a bijection with some well understood set. For instance, if a set can be brought into bijection with the set of all natural numbers, then it is called “countably infinite.” This kind of counting differs in a fundamental way from counting of finite sets, in that adding new elements to a set does not necessarily increase its size, because the possibility of a bijection with the original set is not excluded. For instance, the set of all integers (including negative numbers) can be brought into bijection with the set of natural numbers, and even seemingly much larger sets like that of all finite sequences of rational numbers are still (only) countably infinite. Nevertheless there are sets, such as the set of real numbers, that can be shown to be “too large” to admit a bijection with the natural numbers, and these sets are called “uncountable.” Sets for which there exists a bijection between them are said to have the same cardinality, and in the most general sense counting a set can be taken to mean determining its cardinality. Beyond the cardinalities given by each of the natural numbers, there is an infinite hierarchy of infinite cardinalities, although only very few such cardinalities occur in ordinary mathematics (that is, outside set theory that explicitly studies possible cardinalities).

Counting, mostly of finite sets, has various applications in mathematics. One important principle is that if two sets \( X \) and \( Y \) have the same finite number of elements, and a function \( f : X \rightarrow Y \) is known to be injective, then it is also surjective, and vice versa. A related fact is known as the pigeonhole principle, which states that if two sets \( X \) and \( Y \) have finite numbers of elements \( n \) and \( m \) with \( n > m \), then any map \( f : X \rightarrow Y \) is not injective (there exist two distinct elements of \( X \) that \( f \) sends to the same element of \( Y \)); this follows from the former principle, since if \( f \) were injective, then so would its restriction to a strict subset \( S \) of \( X \) with \( m \) elements, which restriction would then be surjective, contradicting the fact that for \( x \) in \( X \) outside \( S \), \( f(x) \) cannot be in the image of the restriction. Similar counting arguments can prove the existence of certain objects without explicitly providing an example. In the case of infinite sets this can even apply in situations where it is impossible to give an example; for instance there must exist real numbers that are not computable numbers, because the latter set is only countably infinite, but by definition a non-computable number cannot be precisely specified.

The domain of enumerative combinatorics deals with computing the number of elements of finite sets, without actually counting them; the latter usually being impossible because infinite families of finite sets are considered at once, such as the set of permutations of \( \{ 1, 2, \ldots, n \} \) for any natural number \( n \).
B.3 Formalities of Counting

**Definition B.9.** We say \( \text{card}(X) \leq \text{card}(Y) \) if there exists an injective map, \( f : X \rightarrow Y \) and \( \text{card}(Y) \geq \text{card}(X) \) if there exists a surjective map \( g : Y \rightarrow X \). We say \( \text{card}(X) = \text{card}(Y) \) if there exists bijections, \( f : X \rightarrow Y \).

**Proposition B.10.** We have \( \text{card}(X) \leq \text{card}(Y) \) iff \( \text{card}(Y) \geq \text{card}(X) \).

**Proof.** If \( f : X \rightarrow Y \) is an injective map, define \( g : Y \rightarrow X \) by \( g(y) = f^{-1}(y) \) for \( y \in Y \). Then \( g \) is a bijection. Moreover if \( g : Y \rightarrow X \) is a surjective map, then \( f = g^{-1}(\{x\}) \) is the inverse map for all \( x \in X \).

We continue this process of inverse iterates as long as it is possible, i.e. we can construct \( y_n \) if \( x_n \in Y \) and \( x_{n+1} \in f(X) \). There are now three possibilities:

1. ancestor \((x)\) has infinite length so the process never gets stuck in which case we say \( x \in X_\infty \), read as start in \( X \) and end never get stuck.
2. ancestor \((x)\) is finite and the last term in the sequence is in \( Y \), in which case we say \( x \in X_Y \), read as start in \( X \) and end in \( Y \) (get stuck in \( X \)).
3. ancestor \((x)\) is finite and the last term in the sequence is in \( X \), in which case we say \( x \in Y_X \), read as start in \( X \) and end in \( Y \) (get stuck in \( X \)).

In this way we partition \( X \) into three disjoint sets, \( X_\infty, X_Y, \) and \( Y_X \). Similarly we may partition \( Y \) into \( Y_\infty, Y_X, \) and \( X_Y \). Let us now observe that, if \( f(X_\infty) = Y_\infty \), then \( f(x) = x \) for all \( x \in X_\infty \). If \( f(Y_\infty) = Y_\infty \), then \( f(x) = y \) for all \( x \in Y_\infty \) and \( y \in f(X_\infty) \). Thus we have shown \( f : X_\infty \rightarrow Y_\infty \) is a bijection, i.e. \( \text{card}(X_\infty) = \text{card}(Y_\infty) \).

2. \( f(X_Y) = Y_X \). Indeed if \( x \in X_Y \), then \( f(x) = x \). Moreover if \( y \in Y_X \), then \( f(y) = y \). Thus we have shown \( f : X_X \rightarrow Y_Y \) is a bijection, i.e. \( \text{card}(X_X) = \text{card}(Y_Y) \).

3. By the same argument as in item 2. it follows that \( g : Y_Y \rightarrow X_X \) is a bijection.

The last three statements implies \( \text{card}(X) = \text{card}(Y) \). We may in fact define a bijection, \( h : X \rightarrow Y \), by \( h(x) = \begin{cases} f(x) & \text{if } x \in X_\infty \cup X_Y \\ g^{-1}(x) & \text{if } x \in Y_Y \end{cases} \).
Proof. If \( f : A_1 \to A_2 \), then either \( f(1) = 1 \) or \( f(1) = 2 \). In either case \( f \) is injective but not bijective so that \( \text{card}(A_2) < \text{card}(A_1) \). Let \( S_n \) be the statement that \( \text{card}(A_k) < \text{card}(A_l) \) for all \( 1 \leq k < l \leq n \) and for any proper subset \( X \subset A_n \) we have \( \text{card}(X) = \text{card}(A_m) \) for some \( m < n \). Then we have just shown that \( S_2 \) is true. So suppose that \( S_n \) is now true. As \( f : A_k \to A_l \) defined by \( f(m) = m \) for all \( m \in A_k \) is a bijection when \( k < l \) we always have \( \text{card}(A_k) \leq \text{card}(A_l) \). Now suppose that \( \text{card}(A_k) = \text{card}(A_{k+1}) \) for some \( k \leq n \). Then there exists a bijection, \( f : A_{n+1} \to A_k \). In this case \( f(A_n) \) is a proper subset of \( A_k \) and therefore \( \text{card}(f(A_n)) < \text{card}(A_k) \) but on the other hand card \( (f(A_n)) = \text{card}(A_n) \geq \text{card}(A_k) \) which is a contradiction. So no such bijection can exists and we have shown \( \text{card}(A_k) < \text{card}(A_{n+1}) \) for all \( k \leq n \). Finally suppose that \( X \subset A_{n+1} \) is proper subset. If \( X \subset A_n \) then \( \text{card}(X) = \text{card}(A_k) \) for some \( k \leq n \) by the induction hypothesis. On the other hand if \( n+1 \in X \), let \( X' = X \setminus \{n+1\} \supseteq A_n \). Therefore by the induction hypothesis \( \text{card}(X') = \text{card}(A_k) \) for some \( k < n \). It is then clear that \( \text{card}(X) = \text{card}(A_{k+1}) \) where \( k+1 < n \), indeed we map \( X := X' \cup \{n+1\} \to A_k \cup \{k+1\} = A_{k+1} \).

Example B.17. \( \text{card}(A_n \setminus \{k\}) = n-1 \) for \( k \in A_n \). Indeed, let \( f : A_{n-1} \to A_n \setminus \{k\} \) be defined by

\[
 f(x) = \begin{cases} 
 x & \text{if } x < k \\
 x+1 & \text{if } x \geq k 
\end{cases}
\]

Then \( f \) is the desired bijection. More generally if \( X \subset Y \) and \( \text{card}(X) = m < n = \text{card}(Y) \), then \( \text{card}(Y \setminus X) = n-m \) and if \( X \) and \( Y \) are finite disjoint sets then \( \text{card}(X \cup Y) = \text{card}(X) + \text{card}(Y) \). Similarly, \( \text{card}(X \times Y) = \text{card}(X) \cdot \text{card}(Y) \).

Proposition B.18. If \( f : A_n \to A_n \) is a map, then the following are equivalent,

1. \( f \) is injective,
2. \( f \) is surjective,
3. \( f \) is bijective.

Moreover \( \text{card}(\text{Bijec}(A_n)) = n! \).

Proof. If \( n = 1 \), the only map \( f : A_1 \to A_1 \) is \( f(1) = 1 \). So in this case there is nothing to prove. So now suppose the proposition holds for level \( n \) and \( f : A_{n+1} \to A_{n+1} \) is a given map.

If \( f : A_{n+1} \to A_{n+1} \) is an injective map and \( f(A_{n+1}) \) is a proper subset of \( A_{n+1} \), then \( \text{card}(A_{n+1}) < \text{card}(f(A_{n+1})) = \text{card}(A_{n+1}) \) which is absurd. Thus \( f \) is injective implies \( f \) is surjective.

Conversely suppose that \( f : A_{n+1} \to A_{n+1} \) is surjective. Let \( g : A_{n+1} \to A_{n+1} \) be a right inverse, i.e. \( f \circ g = \text{id} \), which is necessarily injective, see the proof of Proposition B.10. By the previous paragraph we know that \( g \) is necessarily surjective and therefore \( f = g^{-1} \) is a bijection.

It now only remains to prove \( \text{card}(\text{Bijec}(A_n)) = n! \) which we again do by induction. For \( n = 1 \) the result is clear. So suppose it holds at level \( n \). If \( f : A_{n+1} \to A_{n+1} \) is a bijection. Given \( 1 \leq k \leq n+1 \) let

\[
 \text{Bij}_k(A_{n+1}) := \{ f \in \text{Bij}(A_{n+1}) : f(n+1) = k \} .
\]

For \( f \in \text{Bij}_k(A_{n+1}) \), we have \( f : A_n \to A_{n+1} \setminus \{k\} \cong A_n \) is a bijection. Thus \( \text{Bij}_k(A_{n+1}) \cong \text{Bij}(A_n) \) and

\[
 \text{Bij}(A_{n+1}) = \sum_{k=1}^{n+1} \text{Bij}_k(A_{n+1})
\]

we have

\[
 \text{card}(\text{Bij}(A_{n+1})) = \sum_{k=1}^{n+1} \text{card}(\text{Bij}_k(A_{n+1}))
\]

\[
 = \sum_{k=1}^{n+1} \text{card}(\text{Bij}(A_n)) = \sum_{k=1}^{n+1} n!
\]

\[
 = (n+1)! = (n+1)! .
\]

Theorem B.19. Suppose that \( X \) is a set. Then \( \text{card}(J_n) \leq \text{card}(X) \) for all \( n \in \mathbb{N} \) if \( \text{card}(J_n) \leq \text{card}(X) \) for all \( n \in \mathbb{N} \).

Proof. Since \( \text{card}(J_n) \leq \text{card}(\mathbb{N}) \) for all \( n \in \mathbb{N} \) it suffices to prove \( \text{card}(J_n) \leq \text{card}(X) \) for all \( n \in \mathbb{N} \).

The intuitive idea is as follows.

Suppose we have constructed \( f_n : J_n \to X \) which is injective. If \( f_n \) were bijective we would have \( \text{card}(J_n) = \text{card}(X) \) and in particular \( \text{card}(J_m) > \text{card}(J_n) = \text{card}(X) \) for all \( m > n \). Thus there exists \( x \in X \setminus f_n(J_n) \) and we then define \( f_{n+1} : J_{n+1} \to X \) so that \( f_{n+1}(n+1) = x \) and \( f_{n+1}|_{J_n} = f_n \). This process continues indefinitely and so we may construct injective maps \( f_n : J_n \to X \) such that \( f_m = f_n|_{J_m} \) for all \( m \leq n \). We then define \( f(m) := f_m(m) \) where \( n \in \mathbb{N} \) is any integer such that \( n \geq m \). In this way we construct a function, \( f : \mathbb{N} \to X \) such that \( f|_{J_n} = f_n \) for all \( n \). This function is easily seen to be injective.

Formalities Version 1. Consider the collection of injective maps \( f : D(f) \subset \mathbb{N} \to X \), where \( D(f) \) is either \( J_n \) for some \( n \in \mathbb{N} \) or is \( \mathbb{N} \). We say \( f \leq g \) if \( D(f) \subset D(g) \) and \( f = g|_{D(f)} \). It is easy to see that every linearly ordered collection of such maps has an upper bound and so by Zorn’s lemma (see
Theorem B.6), there exists a maximal element, \( f \). If \( D(f) \neq \mathbb{N} \) then \( D(f) = J_n \) for some \( n \). By the last paragraph we could extend \( f \) to injective map on \( J_{n+1} \) violating the maximality of \( f \). Thus \( D(f) = \mathbb{N} \) and we have found an injective map from \( \mathbb{N} \) to \( X \).

**Formalities Version 2.** (This argument will avoid the use of Zorn’s Lemma.) By assumption, for each \( n \in \mathbb{N} \) there exists an injective map, \( f_n : J_n \to X \). We now let \( Y := \bigcup_{n \in \mathbb{N}} f_n(J_n) \subset X \). We may construct a surjective map (but not necessarily injective map) \( F : \mathbb{N} \to Y \). From this map we then define \( \psi : Y \to \mathbb{N} \) by \( \psi(y) := \min F^{-1}(\{y\}) \) so that \( \psi : Y \to \mathbb{N} \) is now injective. Suppose for the sake of contradiction that \( \psi(Y) \subset J_N \) for some \( N \in \mathbb{N} \), i.e. \( \psi(Y) \) is a bounded set. Then using our above arguments, we know that \( \text{card}(\psi(Y)) = \text{card}(J_k) \) for some \( k \leq N \). On the other hand, \( f_n : J_n \to Y \) being injective implies \( \text{card}(\psi(Y)) \geq \text{card}(J_n) \) for all \( n \in \mathbb{N} \). As both of these statements cannot be correct at the same time we conclude that \( \psi(Y) \) is unbounded. We may now apply Lemma 5.26 in order to see that \( \text{card}(Y) = \text{card}(\psi(Y)) = \text{card}(\mathbb{N}) \). From this it follows that \( \text{card}(\mathbb{N}) \leq \text{card}(X) \).

**Alternate Proof.** By assumption, there exists and injective map, \( f_n : J_n \to X \) for each \( n \in \mathbb{N} \). By replacing \( X \) by \( X_0 := \bigcup_{n \in \mathbb{N}} f_n(J_n) \) we may assume that \( X = \bigcup_{n \in \mathbb{N}} f_n(J_n) \). As \( X \) is the countable union of finite sets it follows that there exists a surjective map, \( f : \mathbb{N} \to X \) by item 2 of Theorem 5.27. Let \( g : X \to \mathbb{N} \) be defined by \( g(x) := \min f^{-1}(\{x\}) \) for all \( x \in X \) and let \( S := g(\mathbb{N}) \). To finish the proof we need only show that \( S \) is unbounded. If \( S \) were bounded, then we would find \( k \in \mathbb{N} \) such that \( J_k \sim X \). However this is impossible since \( \text{card}(J_n) \leq \text{card}(X) = \text{card}(J_k) \) would imply \( n \leq k \).
Appendix: Aspects of General Topological Spaces

C.1 Compactness

**Definition C.1.** The subset \( A \) of a topological space \((X, \tau)\) is said to be **compact** if every open cover (Definition 9.42) of \( A \) has a finite sub-cover, i.e. if \( \mathcal{U} \) is an open cover of \( A \) there exists \( \mathcal{U}_0 \subset \mathcal{U} \) such that \( \mathcal{U}_0 \) is a cover of \( A \). (We will write \( A \subseteq X \) to denote that \( A \subset X \) and \( A \) is compact.) A subset \( A \subset X \) is **precompact** if \( A \) is compact.

**Proposition C.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 \( K = \bigcup_{i=1}^n K_i \) 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 \( F^c \subset \mathcal{U}_0 \subset \mathcal{U} \). 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 \( \mathcal{U}_i \subset \mathcal{U} \) for each \( i \) such that \( K_i \subset \mathcal{U}_i \). Then \( \mathcal{U}_0 := \bigcup_{i=1}^n \mathcal{U}_i \) is a finite cover of \( K \). \( \square \)

**Exercise C.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 ?? below.)

**Exercise C.2.** Suppose \( f : X \to Y \) is continuous and \( K \subset X \) is compact, then \( f(K) \) is a compact subset of \( Y \). \( \square \)

**Exercise C.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 C.3.** A collection \( \mathcal{F} \) of closed subsets of a topological space \((X, \tau)\) has the **finite intersection property** if \( \cap \mathcal{F}_0 \neq \emptyset \) for all \( \mathcal{F}_0 \subset \mathcal{F} \).

The notion of compactness may be expressed in terms of closed sets as follows.

**Proposition C.4.** A topological space \( X \) is compact iff every family of closed sets \( \mathcal{F} \subset 2^X \) having the finite intersection property satisfies \( \cap \mathcal{F} \neq \emptyset \).

**Proof.** The basic point here is that complementation interchanges open and closed sets and open covers go over to collections of closed sets with empty intersection. Here are the details.

\( (\Rightarrow) \) Suppose that \( X \) is compact and \( \mathcal{F} \subset 2^X \) is a collection of closed sets such that \( \cap \mathcal{F} = \emptyset \). 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 \( \mathcal{F}_0 = \mathcal{U}_0^c \subset \mathcal{F} \), 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 \( \cap \mathcal{F} = \emptyset \). \( \square \)

**Exercise C.4.** Let \((X, \tau)\) be a topological space. Show that \( A \subset X \) is compact iff \((A, \tau_A)\) is a compact topological space.

Let \((X, d)\) be a metric space and for \( x \in X \) and \( \varepsilon > 0 \) let

\[ B_x(\varepsilon) := B_x(x) \setminus \{x\} \]

be the ball centered at \( x \) of radius \( \varepsilon > 0 \) with \( x \) deleted. Recall from Definition 9.15 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 C.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 C.6. A metric space \((X, d)\) is \(\varepsilon\)-\textit{bounded} \((\varepsilon > 0)\) if there exists a finite cover of \(X\) by balls of radius \(\varepsilon\) and it is \textit{totally bounded} if it is \(\varepsilon\)-bounded for all \(\varepsilon > 0\).

Theorem C.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.

\textbf{Proof.} The proof will consist of showing that \(a \Rightarrow b \Rightarrow c \Rightarrow d \Rightarrow a\).

\((a \Rightarrow b)\) We will show that \textbf{not} \(b \Rightarrow \textbf{not} a\). 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 \(V_x \nsubseteq \{x\}\cap E = \emptyset\). 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 \(A \subset X, \cup_{x \in A} V_x\) can only contain a finite number of points from \(E\) (namely \(A \cap E\)). Thus for any \(A \subset X, E \nsubseteq \cup_{x \in A} V_x\) and in particular \(X \neq \cup_{x \in A} V_x\). (See Figure C.1)

\(b \Rightarrow c\) 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 C.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 \(d(x_3, x_1) \geq \varepsilon, \ldots, d(x_{n+1}, x_n) \geq \varepsilon\) and continue inductively choosing points \(\{x_j\}_{j=1}^{n} \subset X\) such that \(d(x_1, x_{j-1}) \geq \varepsilon\). (See Figure C.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 \(\mathcal{V} = \{V_x\}_{x \in A}\) of \(X\) with no finite subcover. Since \(X\) is totally bounded for each \(n \in \mathbb{N}\) there exists \(A_n \subset X\) such that \(X = \bigcup_{x \in A_n} B_x(1/n) \subset \bigcup_{x \in A_n} C_x(1/n)\).

Choose \(x_1 \in A_1\) such that no finite subset of \(\mathcal{V}\) covers \(K_1 := C_{x_1}(1)\). Since \(K_1 = \cup_{x \in A_1} K_1 \cap C_x(1/2)\), there exists \(x_2 \in A_2\) such that \(K_2 := K_1 \cap C_{x_2}(1/2)\) cannot be covered by a finite subset of \(\mathcal{V}\), see Figure C.3. Continuing this way inductively, we construct sets \(K_n = K_{n-1} \cap C_{x_n}(1/n)\) with \(x_n \in A_n\) 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 \(\text{diam}(K_n) \leq 2/n\), 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 \(K_n \nsubseteq \{y\}\) and \(\text{diam}(K_n) \to 0\), it now follows that \(K_n \subset V\) for some \(n\) large. But this violates the assertion that \(K_n\) cannot be covered by a finite subset of \(\mathcal{V}\).

\textbf{Corollary C.8.} Any compact metric space \((X, d)\) is second countable and hence also separable by Exercise ?? (See Example ?? below for an example of a compact topological space which is not separable.)

\textbf{Proof.} To each integer \(n\), there exists \(A_n \subset X\) such that \(X = \cup_{x \in A_n} B(x, 1/n)\). The collection of open balls,
We will show, by choosing $K$ for $i$ and subset of a complete space) and $K$ forms a countable basis for the metric topology on $X$. This is a consequence of Theorem B. Hence if choose $\delta > 0$ 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 $y \in B(x, 1/n)$,

$$d(y, x_0) \leq d(y, x) + d(x, x_0) < 2/n < \varepsilon.$$

**Corollary C.9.** The compact subsets of $\mathbb{R}^n$ are the closed and bounded sets.

**Proof.** This is a consequence of Theorem ?? and Theorem [C.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

$$A_\delta = \delta \mathbb{Z}^n \cap [-M, M]^n := \{\delta x : x \in \mathbb{Z}^n \text{ and } |x_i| \leq M \text{ for } i = 1, 2, \ldots, n\}.$$ 

We will show, by choosing $\delta > 0$ sufficiently small, that

$$K \subset [-M, M]^n \subset \bigcup_{x \in A_\delta} B(x, \varepsilon) \quad (C.1)$$

which shows that $K$ is totally bounded. Hence by Theorem [C.7] $K$ is compact. Suppose that $y \in [-M, M]^n$, then there exists $x \in A_\delta$ such that $|y_i - x_i| \leq \delta$ for $i = 1, 2, \ldots, n$. Hence

$$d^2(x, y) = \sum_{i=1}^n (y_i - x_i)^2 \leq n\delta^2$$

which shows that $d(x, y) \leq \sqrt{n}\delta$. Hence if choose $\delta < \varepsilon/\sqrt{n}$ we have shows that $d(x, y) < \varepsilon$, i.e. Eq. (C.1) holds.

**Example C.10.** Let $X = \ell^p(\mathbb{N})$ with $p \in [1, \infty)$ and $\mu \in \ell^p(\mathbb{N})$ such that $\mu(k) \geq 0$ for all $k \in \mathbb{N}$. The set

$$K := \{x \in X : |x(k)| \leq \mu(k) \text{ for all } k \in \mathbb{N}\}$$

is compact. To prove this, let $\{x_n\}_{n=1}^\infty \subset K$ be a sequence. By compactness of closed bounded sets in $C$, 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 $k \in \mathbb{N}$. Since $|y_n(k)| \leq \mu(k)$ for all $n$ it follows that $|y(k)| \leq \mu(k)$, i.e. $y \in K$.

Finally

$$\lim_{n \to \infty} \|y - y_n\|^p_p = \lim_{n \to \infty} \sum_{k=1}^\infty |y(k) - y_n(k)|^p = \sum_{k=1}^\infty \lim_{n \to \infty} |y(k) - y_n(k)|^p = 0$$

wherin we have used the Dominated convergence theorem. (Note

$$|y(k) - y_n(k)|^p \leq 2^p \mu^p(k)$$

and $\mu^p$ 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

$$|x(k)| \leq \lim_{n \to \infty} |x_n(k)| \leq \mu(k) \quad \forall k \in \mathbb{N}.$$ 

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 $(\sum_{k=N+1}^\infty \mu(k))^1/p < \varepsilon$. 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

$$\prod_{k=1}^N C(\mu(k))(0) \subset \bigcup_{x \in \Lambda} B_x^N(\varepsilon).$$

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 $m_j := n_j^k$ 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 $y_j := x_{m_j}$.
where \( B^N_\varepsilon \) is the open ball centered at \( z \in \mathbb{C}^N \) relative to the \( \ell^p(\{1, 2, 3, \ldots, N\}) \) norm. For each \( z \in A \), let \( \tilde{z} \in X \) be defined by \( \tilde{z}(k) = z(k) \) if \( k \leq N \) and \( \tilde{z}(k) = 0 \) for \( k \geq N + 1 \). I now claim that

\[
K \subset \bigcup_{z \in A} B_{\tilde{z}}(2\varepsilon)
\]

(C.2)

which, when verified, shows \( K \) is totally bounded. To verify Eq. (C.2), let \( x \in K \) and write \( x = u + v \) where \( u(k) = x(k) \) for \( k \leq N \) and \( u(k) = 0 \) for \( k < N \). Then by construction \( u \in B_{\tilde{z}}(\varepsilon) \) for some \( \tilde{z} \in A \) and

\[
\|v\|_{\ell^p} \leq \left( \sum_{k=N+1}^{\infty} |\mu(k)|^p \right)^{1/p} < \varepsilon.
\]

So we have

\[
\|x - \tilde{z}\| = \|u + v - \tilde{z}\| \leq \|u - \tilde{z}\| + \|v\|_{\ell^p} < 2\varepsilon.
\]

**Exercise C.5 (Extreme Value Theorem).** Let \((X, \tau)\) be a compact topological space and \( f : X \to \mathbb{R} \) be a continuous function. Show \(-\infty < \inf f \leq \sup f < \infty\) and there exists \( a, b \in X \) such that \( f(a) = \inf f \) and \( f(b) = \sup f \). Hint: use Exercise C.2 and Corollary C.9.

**Exercise C.6 (Uniform Continuity).** Let \((X, d)\) be a compact metric space, \((Y, \rho)\) 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 12.

**Theorem C.11.** Suppose that \((X, \|\cdot\|)\) is a normed vector in which the unit ball, \( V := B_0(1) \), is precompact. Then \( \dim X < \infty \).

An alternate proof is given in Proposition C.13. Since \( V \) is compact, we may choose \( A \subset X \) such that

\[
\tilde{V} \subset \bigcup_{x \in A} \left( x + \frac{1}{2} V \right)
\]

(C.3)

where, for any \( \delta > 0 \),

\[
\delta V := \{ \delta x : x \in V \} = B_0(\delta).
\]

Let \( Y := \text{span}(A) \), then Eq. (C.3) implies,

\[
V \subset \tilde{V} \subset Y + \frac{1}{2} V.
\]

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

\[
V \subset Y + \frac{1}{2} V \subset Y + \frac{1}{4} V = Y + \frac{1}{4} V.
\]

Continuing this way inductively then shows that

\[
V \subset Y + \frac{1}{2^n} V \text{ for all } n \in \mathbb{N}.
\]

(C.4)

Indeed, if Eq. (C.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 \( z_n \in B_0(2^{-n}) \) such that \( y_n + z_n \to x \). Since \( \lim_{n \to \infty} z_n = 0 \), it follows that \( x = \lim_{n \to \infty} y_n \in \tilde{Y} \). Since \( \dim Y \leq \#(A) < \infty \), Corollary 9.11 implies \( Y = \tilde{Y} \) and so we have shown that \( V \subset Y \).

Since for any \( x \in X, \frac{1}{2^n} x \in V \subset Y \), we have \( x \in Y \) for all \( x \in X \), i.e. \( X = Y \).

**Lemma C.12.** Let \( H \) be a normed linear space and \( H_0 \) a closed proper subspace. For any \( \varepsilon > 0 \), there exists \( x_0 \in H \) such that \( \|x_0\| = 1 \) and \( \|x - x_0\| \geq 1 - \varepsilon \) whenever \( x \in H_0 \).

**Proof.** Can assume \( \varepsilon < 1 \). Take any \( z_0 \notin H_0 \). Let \( d = \inf_{x \in H_0} \|x - z_0\| \). For any \( \delta > 0 \), there exists \( z \in H_0 \), such that \( \|z - z_0\| \leq d + \delta \). Take \( \delta = \frac{d}{1 - \varepsilon} \). Let \( x_0 = (z - z_0)/\|z - z_0\| \), where \( z \) is determined for this \( \delta \). Then \( \|x_0\| = 1 \), and if \( x \in H_0 \),

\[
\|x - x_0\| = \frac{\|z - z_0\|}{\|z - z_0\|} \geq \frac{d}{\|z - z_0\|} \geq \frac{d}{d + \delta} = 1 - \varepsilon.
\]

Here is the proof again at a higher level. Choose \( h \in H_0 \) such that \( d := \text{dist} (z_0, H_0) \equiv \|h - z_0\| \) and then take \( x_0 := (h - z_0)/\|h - z_0\| \) as above. Then

\[
\text{dist} (x_0, H_0) = \frac{1}{\|h - z_0\|} \text{dist} (h - z_0, H_0)
\]

\[= \frac{1}{\|h - z_0\|} \text{dist} (z_0, H_0) \equiv 1,
\]

where we have used the easily verified fact that \( \text{dist}(ax + h, H) = |a| \text{dist}(x, H) \) for all \( a \in \mathbb{R} \) and \( h \in H \).
Proposition C.13. A locally compact Banach space is finite dimensional.

Proof. We prove that an infinite dimensional Banach space is not locally compact. We construct a sequence \( x_1, x_2, \ldots, x_n, \ldots \) such that \( \| x_n \| = 1, \| x_i - x_j \| \geq 1/2, i \neq j \). Take \( x_1 \) to be any unit vector. Suppose vectors \( x_1, \ldots, x_n \) are constructed. Let \( H_0 \) be the linear span of \( x_1, \ldots, x_n \). By Corollary 9.41, \( H_0 \) is closed. By Lemma C.12 there exists \( x_{n+1} \) such that \( \| x_i - x_{n+1} \| \geq 1/2, i = 1, \ldots, n \). Now the sequence just constructed has no Cauchy subsequence. Hence the closed unit ball is not compact. Similarly the closed ball of radius \( r > 0 \) is also not compact.

Exercise C.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.

C.2 Exercises

C.2.1 General Topological Space Problems

Exercise C.8. 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 \( n = 1, 2, \ldots < N \) with \( N = \infty \) possible.

Exercise C.9. Let \((X, \tau) \) and \((Y, \tau')\) be topological spaces, \( f : X \to Y \) be a function, \( U \) 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 \((\tau, \tau')\) continuous then \( f|_A : A \to Y \) is \((\tau_A, \tau')\) continuous.
2. Show \( f : X \to Y \) is \((\tau, \tau')\) continuous iff \( f|_U : U \to Y \) is \((\tau_U, \tau')\) continuous for all \( U \in \mathcal{U} \).
3. Show \( f : X \to Y \) is \((\tau, \tau')\) continuous iff \( f|_{F_j} : F_j \to Y \) is \((\tau_{F_j}, \tau')\) continuous for all \( j = 1, 2, \ldots, n \).

Exercise C.10. 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 \( S_\alpha \subset \tau_\alpha \) is a sub-base for the topology \( \tau_\alpha \) for each \( \alpha \in A \), show \( S := \cup_{\alpha \in A} f_\alpha^{-1}(S_\alpha) \) is a sub-base for the topology \( \tau := \tau(f_\alpha : \alpha \in A) \).

C.2.2 Metric Spaces as Topological Spaces

Definition C.14. Two metrics \( d \) and \( \rho \) on a set \( X \) are said to be equivalent if there exists a constant \( c \in (0, \infty) \) such that \( c^{-1} \rho \leq d \leq c \rho \).

Exercise C.11. Suppose that \( d \) and \( \rho \) are two metrics on \( X \).
1. Show \( \tau_d = \tau_\rho \) if \( d \) and \( \rho \) are equivalent.
2. Show by example that it is possible for \( \tau_d = \tau_\rho \) even though \( d \) and \( \rho \) are inequivalent.

Exercise C.12. Let \((X_i, d_i) \) for \( i = 1, \ldots, n \) be a finite collection of metric spaces and for \( 1 \leq p \leq \infty \) 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
\[
\rho_p(x, y) = \left( \sum_{i=1}^n [d_i(x_i, y_i)]^p \right)^{1/p} \quad \text{if} \quad p \neq \infty.
\]
\[
\rho_\infty(x, y) = \max_{i=1}^n d_i(x_i, y_i) \quad \text{if} \quad p = \infty.
\]
1. Show \((X, \rho_p)\) is a metric space for \( p \in [1, \infty] \). Hint: Minkowski’s inequality.
2. Show for any \( p, q \in [1, \infty] \), the metrics \( \rho_p \) and \( \rho_q \) are equivalent. Hint: This can be done with explicit estimates or you could use Theorem 9.39 below.

Notation C.15 Let \( X \) be a set and \( p : = \{ p_n \}_{n=0}^\infty \) 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 \( p_n(x, y) \leq p_{n+1}(x, y) \) 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 \( \tau(p) \) 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. \( \tau(p) := \tau(p_n(x, \cdot) : n \in \mathbb{N} \) and \( x \in X \).

Exercise C.13. Using Notation C.15 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 \( \tau(p) \). Hint: Use Exercise C.10 to show \( \mathcal{B} \) is a sub-base for the topology \( \tau(p) \) and then use Exercise ?? to show \( \mathcal{B} \) is in fact a base for the topology \( \tau(p) \).

Exercise C.14 (A minor variant of Exercise 9.21). Let \( p_n \) be as in Notation C.15 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 \( \tau_d = \tau(p) \). 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 C.15. 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 9.21.) 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 \( X \).

\section*{C.2.3 Compactness Problems}

Exercise C.16 (Tychonoff’s Theorem for Compact Metric Spaces). Let us continue the Notation used in Exercise 9.21. Further assume that the spaces \( X_n \) are compact for all \( n \). Show (without using Theorem ?? below) that \( (X, d) \) 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 C.10 and see Theorem ?? below for the general version of this theorem.)

\section*{C.3 Connectedness in General Topological Spaces}

(You may safely skip this section on first reading.) As a warm up suppose that \( (X, d) \) is a metric space and \( Y \) be a non-empty subset of \( X \). If \( A \subset Y \), the closure of \( A \) in the metric space \( (Y, d) \) is denote by \( \overline{A}^Y \) and is defined by
\[
\overline{A}^Y := \{ y \in Y : y = \lim_{n \to \infty} a_n \text{ for some } \{a_n\} \subset A \} = \overline{A} \cap Y.
\]

Lemma C.16. A set \( A \subset Y \) is closed in \( Y \) iff \( A = \overline{A}^Y = \overline{A} \cap Y \) iff there exists \( C \subset X \) closed such that \( A = C \cap Y \). Similarly, \( A \) is open in \( Y \) iff there exists \( V \) open in \( X \) such that \( A = V \cap Y \).

Proof. The assertion that \( A \subset Y \) is closed in \( Y \) iff \( A = \overline{A}^Y = \overline{A} \cap Y \) follows directly from the definition. Hence if \( A \) is closed in \( Y \), then \( A = \overline{A} \cap Y \) where we may take \( C = \overline{A} \cap Y \) - a closed subset of \( X \). Conversely, if \( A = C \cap Y \), then \( \overline{A} \subset C \) and hence \( A \subset \overline{A}^Y = \overline{A} \cap Y \subset C \cap Y = A \) which shows that \( A = \overline{A}^Y \) and so \( A \) is closed in \( Y \).

If \( A \subset Y \) is open, then \( Y \setminus A \) is closed and hence \( Y \setminus A = Y \cap C \) for some closed subset, \( C \subset X \). Since,
\[
A = Y \setminus (Y \setminus A) = Y \cap \overline{Y \setminus C} = Y \cap (Y \cup C) = Y \cap C
\]
we see that \( A = Y \cap V \) where \( V := C \) is open in \( X \). Conversely if \( A = Y \cap V \), then
\[
Y \setminus A = Y \setminus (Y \cap V) = Y \cap (Y \setminus V) = Y \cap V
\]
which shows that \( Y \setminus A \) is closed in \( Y \).

Proposition C.17. Let \( E \) be a subset of a metric space, \( (X, d) \). Then \( E \) is disconnected iff \( E \) is the disjoint union of two non-empty relatively open subset of \( E \), i.e. there exists open subset \( U \) and \( V \) of \( X \) such that;
1. \( U \cap E \neq \emptyset \neq V \cap E \) while \( U \cap V \cap E = \emptyset \),
2. and \( E = (U \cap E) \cup (V \cap E) \).

Proof. Suppose that \( E \) is a disconnected and write \( E = A \cup B \) where where \( A \) and \( B \) are two non-empty separated sets. Let \( U := \overline{B} \) and \( V := \overline{A} \). Then \( A \subset U \) and \( B \subset V \). Moreover, \( E \cap U = (A \cup B) \cap \overline{B} = A \setminus B = A \neq \emptyset \) and similarly \( E \cap V = B \neq \emptyset \) while \( U \cap V \cap E = A \cap B = \emptyset \). Therefore \( E \) is the disjoint union of two non-empty relatively open sets.

Conversely, suppose that \( E = A \cup B \) where \( A \) and \( B \) are relatively open non-empty disjoint open subsets of \( E \). Then both \( A \) and \( B \) are relatively closed in \( E \) and hence,
\[
A = \overline{A}^E := \{ x \in E : \exists \{a_n\} \subset A \ni \lim_{n \to \infty} a_n = x \} = \overline{A} \cap E
\]
and similarly \( B = \overline{B}^E = \overline{B} \cap E \).

\[
A = \overline{A} \cap E = \overline{A} \cap [A \cup B] = A \cup [\overline{A} \cap B]
\]
and \( A \cap B = \emptyset \) we conclude,
\[
\emptyset = A \cap B = (A \cup [\overline{A} \cap B]) \cap B = \overline{A} \cap B.
\]
Similarly we may show \( \overline{A} \cap B = \emptyset \) and hence \( E \) is written as the union of two non-empty separated sets.

Definition C.18. \((X, \tau)\) is disconnected if there exist 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 \((X, \tau)\) 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 \((A, \tau_A)\) is connected where \( \tau_A \) is the relative topology on \( A \). Explicitly, \( A \) is disconnected in \((X, \tau)\) 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 \((X, \tau)\) is connected iff the only sets \(A \subsetneq X\) which are both open and closed are the sets \(X\) and \(\emptyset\). This version of the definition is often used in practice.

Remark C.19. Let \(A \subsetneq Y \subsetneq X\). Then \(A\) is connected in \(X\) iff \(A\) is connected in \(Y\).

Proof. Since

\[
\tau_A := \{ V \cap A : V \subset X \} = \{ V \cap A \cap Y : V \subset X \} = \{ U \cap A : U \subset Y \},
\]

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 \(Y\). \(\blacksquare\)

Theorem C.20 (The Connected Subsets of \(\mathbb{R}\)). The connected subsets of \(\mathbb{R}\) are intervals.

Proof. Suppose that \(A \subsetneq \mathbb{R}\) 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 \((a, b) \subsetneq A\). Let

\[
\alpha := \inf(A) \text{ and } \beta := \sup(A)
\]

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 \((\alpha, \beta) = \bigcup_{n=1}^{\infty} (\alpha_n, \beta_n)\) \(\subset A\). From this it follows that \(A = (\alpha, \beta), (\alpha, \beta)\) or \([\alpha, \beta]\), i.e. \(A\) is an interval.

Conversely suppose that \(A\) is a sub-interval of \(\mathbb{R}\). For the 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 \([a, b] \subsetneq A\). Let \(p = \sup([a, b] \cap U)\), then because \(U\) and \(V\) are open, \(a < p < b\). Now \(p\) can not be in \(U\) for otherwise \(sup([a, b] \cap U) > p\) and \(p\) can not be in \(V\) for otherwise \(p < sup([a, b] \cap U)\). 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 of \(A\).

Alternative proof of the converse. Because of Theorem C.22 below, it suffices to assume that \(A\) is an open interval. For the 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 \(a < b\). Let \(J_b = (\alpha, \beta)\) be the maximal open interval in \(U\) which contains \(a\). (See Exercise C.8 or Remark 9.67 for the structure of open subsets of \(\mathbb{R}\).) If \(\beta \in U\) we could extend \(J_b\) to the right and still be in \(U\) violating the definition of \(\beta\). Moreover we can not have \(\beta \in V\) because in this case \(J_b\) would not be in \(U\). Therefore \(\beta \notin U \cup V = A\) while on the other hand \(a < \beta < b\) and so \(\beta \in A\) as \(A\) is an interval and we have reached the desired contradiction. \(\blacksquare\)

Here is a typical way these connectedness ideas are used.

Example C.21. Suppose that \(f : X \to Y\) is a continuous map between two topological spaces, the space \(X\) is connected and the space \(Y\) is \(\mathbb{T}_1\) “i.e. \(\{y\}\) is a closed set for all \(y \in Y\) in as Definition 7? 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 C.22 (Properties of Connected Sets). Let \((X, \tau)\) 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 \(\widehat{A}\) is connected.

   b) More generally, if \(A\) is connected and \(B \subset \text{acc}(A)\) or \(B \subset \text{bd}(A)\), then \(A \cup B\) is connected as well. (Recall that \(\text{acc}(A)\) – the set of accumulation points of \(A\) was defined in Definition 9.17 above. Moreover by Exercise ??, we know that \(\text{acc}(A) \setminus A = \text{bd}(A) \setminus A\). What we are really showing here is that for any \(B\) such that \(A \subset B \subset \widehat{A}\), then \(B\) is connected.)

3. If \(\{E_\alpha\}_{\alpha \in A}\) is a collection of connected sets such that \(E_\alpha \cap E_\beta \neq \emptyset\) for all \(\alpha, \beta \in A\) 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 \(\widehat{A} \cap B \neq \emptyset\), 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 \(x\).2

One may assume much less here. What we really need is for any \(\alpha, \beta \in A\) there exists \((\alpha_i)_{i=0}^n\) in \(A\) such that \(\alpha_0 = \alpha, \alpha_n = \beta, \text{ and } E_{\alpha_i} \cap E_{\alpha_{i+1}} \neq \emptyset\) for all \(0 \leq i < n\). Moreover if we make use of item 4. it suffices to assume that \(E_{\alpha_i} \cup E_{\alpha_{i+1}} \cup E_{\alpha_i} \cap E_{\alpha_{i+1}} \neq \emptyset\) for all \(0 \leq i < n\).
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 $Y = A$ be equipped with the relative topology from $X$. Suppose that $U, V \subset A$ form a disconnection of $Y = A$. Then by 1. either $A \subset U$ or $A \subset V$. Say that $A \subset U$. Since $U$ is both open and closed in $Y$, it follows that $Y = A \subset U$. Therefore $V = \emptyset$ and we have a contradiction to the assumption that $\{ U, V \}$ is a disconnection of $Y = A$. Hence we must conclude that $Y = A$ is connected as well.

b) Now let $Y = A \cup B$ with $B \subset acc(A)$, then

$$\tilde{A}^Y = \tilde{A} \cap Y = (A \cup acc(A)) \cap Y = A \cup B.$$  

Because $A$ is connected in $Y$, by (2a) $Y = A \cup B = \tilde{A}^Y$ is also connected.

3. Let $Y := \bigcup_{\alpha \in A} E_\alpha$. By Remark C.19 we know that $E_\alpha$ 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 $\alpha$. Let $A = \{ \alpha \in A : E_\alpha \subset U \}$ then $U = \bigcup_{\alpha \in A} E_\alpha$ and $V = \bigcup_{\alpha \in A \setminus A} E_\alpha$. (Notice that neither $A$ or $A \setminus A$ can be empty since $U$ and $V$ are not empty.) Since

$$\emptyset = U \cap V = \bigcup_{\alpha \in A, \beta \in A^c} (E_\alpha \cap E_\beta) \neq \emptyset.$$  

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

$$A = \tilde{A}^Y = \tilde{A} \cap Y = \tilde{A} \cap (A \cup B) = A \cup (\tilde{A} \cap B)$$  

and therefore

$$\emptyset = A \cap B = [A \cup (\tilde{A} \cap B)] \cap B = \tilde{A} \cap B \neq \emptyset$$  

which gives us the desired contradiction.

Alternative proof. Let $A' := A \cup [B \cap \tilde{A}]$ so that $A \cup B = A' \cup B$. By item 2b, we know $A'$ is still connected and since $A' \cap B \neq \emptyset$ we may now apply item 3. to finish the proof.

5. Let $C$ denote the collection of connected subsets $C \subset X$ such that $x \in C$. Then by item 3., the set $C_x := \cup C$ is also a connected subset of $X$ which contains $x$ and clearly this is the unique maximal connected set containing $x$. Since $C_x$ is also connected by item (2) and $C_x$ is maximal, $C_x = \tilde{C}_x$, i.e. $C_x$ is closed.

Proposition C.23 (Stability of Connectedness Under Products). Let $(X_\alpha, \tau_\alpha)$ 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 9.57 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 C.4. By induction the theorem holds whenever $A$ is a finite set, i.e. for products of a finite number of connected spaces.

![Figure C.4](image_url)

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
Exercise C.17 (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 \((\bar{X}, \bar{d})\) and an isometric map \(i : X \to \bar{X}\) such that \(i(X)\) is dense in \(\bar{X}\), see Definition 9.13.

1. Let \(C\) denote the collection of Cauchy sequences \(a = \{a_n\}_{n=1}^{\infty} \subset X\). Given two elements \(a, b \in C\) show \(d_C(a, b) := \lim_{n \to \infty} d(a_n, b_n)\) exists, \(d_C(a, b) \geq 0\) for all \(a, b \in C\) and \(d_C\) satisfies the triangle inequality,

\[
d_C(a, c) \leq d_C(a, b) + d_C(b, c)
\]

Thus \((C, d_C)\) would be a metric space if it were true that \(d_C(a, b) = 0\) iff \(a = b\). This however is false, for example if \(a_n = b_n\) for all \(n \geq 100\), then \(d_C(a, b) = 0\) while \(a\) need not equal \(b\).

2. Define two elements \(a, b \in C\) to be equivalent (write \(a \sim b\)) whenever \(d_C(a, b) = 0\). Show \(\sim\) is an equivalence relation on \(C\) and that \(d_C(a', b') = d_C(a, b)\) if \(a \sim a'\) and \(b \sim b'\). (Hint: see Corollary 6.46)

3. Given \(a \in C\) let \(\bar{a} := \{b \in \bar{C} : b \sim a\}\) denote the equivalence class containing \(a\) and let \(\bar{X} := \{\bar{a} : a \in C\}\) denote the collection of such equivalence classes. Show that \(\bar{d}(\bar{a}, \bar{b}) := d_C(a, b)\) is well defined on \(\bar{X} \times \bar{X}\) and verify \((\bar{X}, \bar{d})\) 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 \((\bar{X}, \bar{d})\) is complete. Hint: if \(\{\bar{a}(m)\}_{m=1}^{\infty}\) is a Cauchy sequence in \(\bar{X}\) choose \(b_m \in X\) such that \(\bar{d}(\bar{a}(m), \bar{a}(m)) \leq 1/m\). Then show \(\bar{a}(m) \to \bar{b}\) where \(b = \{b_m\}_{m=1}^{\infty}\).

C.5 Supplementary Remarks

C.5.1 Riemannian Metrics

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

Lemma C.24. 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 \(\sigma\). We define \(\ell(\sigma) = \infty\) if \(\sigma\) is not piecewise smooth.

Then

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

where the boundary operation, \(\text{bd}(\cdot)\) is defined in Definition 9.15 below.

Proof. Let \(C := C_x(\varepsilon) \subset B_x(\varepsilon) =: \bar{B}\). 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 \(B_y(\delta) \cap \bar{B}(\varepsilon) \neq \emptyset\). 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
Fig. C.5. An almost length minimizing curve joining $x$ to $y$.

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

and choose a curve $\sigma$ 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

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

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

Remark C.25. 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 $\sigma$ be a curve from $x$ to $y$ and let $\varepsilon = \ell(\sigma) - d(x, y)$. Then for all $0 \leq u < v \leq 1$,

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

$\geq d(x, \sigma(u)) + \ell(\sigma|_{[u, v]}) + d(\sigma(v), y)$

and therefore, using the triangle inequality,

$\ell(\sigma|_{[u, v]}) \leq d(x, y) + \varepsilon - d(x, \sigma(u)) - d(\sigma(v), y)$

$\leq d(\sigma(u), \sigma(v)) + \varepsilon$.

This leads to the following conclusions. If $\sigma$ is within $\varepsilon$ of a length minimizing curve from $x$ to $y$ then $\sigma|_{[u, v]}$ is within $\varepsilon$ of a length minimizing curve from $\sigma(u)$ to $\sigma(v)$. In particular if $\sigma$ is a length minimizing curve from $x$ to $y$ then $\sigma|_{[u, v]}$ is a length minimizing curve from $\sigma(u)$ to $\sigma(v)$. Appendix: Limsup and liminf of functions
Appendix: Limsup and liminf of functions

[Thanks go to Ali Behzadan from providing this appendix.] Our goal in this appendix is to introduce the notion of limit superior and limit inferior for functions. The definitions and properties will be very similar to those that were previously discussed for sequences. Here we will assume the domain is a metric space and the destination set is the set of extended real numbers. One may easily generalize the definitions to the case where domain and codomain are topological spaces and of course the codomain must be an ordered set. By introducing a proper topology on the set of extended real numbers one can consider the limit superior and limit inferior of sequences as a special case of what will be discussed here.

First let's review the definition of limit points in a metric space.

Definition D.1. Let \((X, d)\) be a metric space and \(E \subseteq X\). A point \(p \in X\) is a limit point of the set \(E\) if every open ball about \(p\) contains a point \(q \neq p\) such that \(q \in E\).

Exercise D.1. Prove that \(p\) is a limit point of \(E\) iff there exists a sequence \(\{x_n\}_{n=1}^\infty \in E - \{p\}\) such that \(\lim_{n \to \infty} x_n = p\).

Example D.2. Consider \(\mathbb{R}\) (with the Euclidean metric \(d(x, y) = |x - y|\)). Let \(E = \{2, 3\} \cup \{5\}\). \(p = 2\) is a limit point of \(E\) because \(2 = \lim_{n \to \infty}(2 + 1/n)\). \(p = 5\) is not a limit point of \(E\) because the open ball of radius 1/2 centered at 5 does not intersect \(E - \{5\}\).

Definition D.3. Let \((X, d)\) be a metric space. Suppose \(E \subseteq X\) and \(p\) is a limit point of \(E\). Let \(f\) be a real (or extended real) valued function on \(E\), i.e., \(f : E \to \mathbb{R}\) (or \(f : E \to \hat{\mathbb{R}}\)). Then

\[
\liminf_{x \to p} f(x) = \liminf_{\delta \downarrow 0} \{f(x) : x \in E \cap B_p(\delta) - \{p\}\},
\]

\[
\limsup_{x \to p} f(x) = \limsup_{\delta \downarrow 0} \{f(x) : x \in E \cap B_p(\delta) - \{p\}\}.
\]

Remark D.4. Clearly if \(\nu > 0\) is such that \(B_p(\nu) \subseteq E\) and \(\hat{f} := f|_{B_p(\nu)}\) then

\[
\liminf_{x \to p} f(x) = \liminf_{x \to p} \hat{f}(x), \quad \limsup_{x \to p} f(x) = \limsup_{x \to p} \hat{f}(x).
\]

So in fact if \(f_1\) and \(f_2\) agree on an open ball about \(p\) in \(X\) then

\[
\liminf_{x \to p} f_1(x) = \liminf_{x \to p} f_2(x), \quad \limsup_{x \to p} f_1(x) = \limsup_{x \to p} f_2(x).
\]

Remark D.5. Notice that if we set,

\[
g(\delta) := \inf\{f(x) : x \in E \cap B_p(\delta) - \{p\}\},
\]

\[
h(\delta) := \sup\{f(x) : x \in E \cap B_p(\delta) - \{p\}\},
\]

then as \(\delta\) shrinks (as \(\delta \downarrow 0\)) \(g(\delta)\) increases and \(h(\delta)\) decreases. i.e., if \(\delta_1 > \delta_2\) then \(g(\delta_1) \leq g(\delta_2)\) and \(h(\delta_1) \geq h(\delta_2)\). Therefore the limits in the definition of \(\limsup\) and \(\liminf\) always exist in the set of extended real numbers and we have:

\[
\liminf_{x \to p} f(x) = \sup_{\delta > 0}\{f(x) : x \in E \cap B_p(\delta) - \{p\}\},
\]

\[
\limsup_{x \to p} f(x) = \inf_{\delta > 0}\{f(x) : x \in E \cap B_p(\delta) - \{p\}\}.
\]

Notation D.6. \(S_{p, \delta} := E \cap B_p(\delta) - \{p\}\).

Remark D.7. In the special case where \(X = \mathbb{R}\) and \(E\) is an open interval in \(\mathbb{R}\) containing \(p\) we have:

\[
\liminf_{x \to p} f(x) = \lim_{\delta \downarrow 0}\inf_{0 < |x - p| < \delta} f(x),
\]

\[
\limsup_{x \to p} f(x) = \lim_{\delta \downarrow 0}\sup_{0 < |x - p| < \delta} f(x).
\]

In the following two propositions we list some of the most important properties of limit superior and limit inferior of functions. The reader may check that these properties are completely analogous to those of sequences (Proposition 3.28).

Proposition D.8. Let \((X, d)\) be a metric space. Suppose \(E \subseteq X\), \(f : E \to \hat{\mathbb{R}}\), \(g : E \to \hat{\mathbb{R}}\) and \(p\) is a limit point of \(E\).

1. \(\liminf_{x \to p} (-f(x)) = -\limsup_{x \to p} f(x)\).
2. \(\liminf_{x \to p} f(x) \leq \limsup_{x \to p} f(x)\).
3. If \(f(x) \leq g(x)\) on the intersection of an open ball about \(p\) with \(E\) (or on \(S_{p, \nu}\) for some \(\nu > 0\)) then

\[
\liminf_{x \to p} f(x) \leq \liminf_{x \to p} g(x), \quad \limsup_{x \to p} f(x) \leq \limsup_{x \to p} g(x).
\]
Proof.

1. From Proposition 8.8, one can easily conclude that if $S$ is a nonempty subset of the extended real numbers then $\inf(S) = -\sup(-S)$. Thus

$$\liminf_{x \to p} (-f(x)) = \lim \inf_{\delta > 0} \inf_{x \in S_{p,\delta}} (-f(x)) = \lim(-\sup_{x \in S_{p,\delta}} f(x))$$

$$= -\lim_{\delta > 0} \sup_{x \in S_{p,\delta}} (f(x)) = -\limsup_{x \to p} f(x).$$

2. We have

$$\forall \delta > 0 \quad \inf_{x \in S_{p,\delta}} f(x) \leq \sup_{x \in S_{p,\delta}} f(x).$$

Since the limit of both sides as $\delta \to 0$ exists (in the extended real numbers), we get

$$\liminf_{x \to p} f(x) \leq \limsup_{x \to p} f(x),$$

therefore, $\liminf_{x \to p} f(x) \leq \limsup_{x \to p} f(x)$.

3. By assumption there exists $\nu > 0$ such that $f(x) \leq g(x)$ on $S_{p,\nu}$. Thus if $0 < \delta \leq \nu$ then $f(x) \leq \sup_{y \in S_{p,\delta}} g(y)$ for all $x \in S_{p,\delta}$. Therefore

$$\sup_{x \in S_{p,\delta}} f(x) \leq \sup_{y \in S_{p,\delta}} g(y).$$

Similarily, if $0 < \delta \leq \nu$ then $\inf_{x \in S_{p,\delta}} f(x) \leq g(y)$ for all $y \in S_{p,\delta}$ and hence

$$\inf_{x \in S_{p,\delta}} f(x) \leq \inf_{y \in S_{p,\delta}} g(y).$$

Passing to the limit (as $\delta \to 0$) in each of these inequalities gives the desired inequalities.

\[
\text{Proposition D.9. Let } (X, d) \text{ be a metric space. Suppose } E \subseteq X, f : E \to \mathbb{R}, g : E \to \mathbb{R} \text{ and } p \text{ is a limit point of } E.
\]

1. $\lim_{x \to p} f(x)$ exists in $\overline{\mathbb{R}}$ iff $\liminf_{x \to p} f(x) = \limsup_{x \to p} f(x) \in \overline{\mathbb{R}}$. In this case $\lim_{x \to p} f(x) = \liminf_{x \to p} f(x) = \limsup_{x \to p} f(x)$.

2. $\lim_{x \to p} (f(x) + g(x)) = \liminf_{x \to p} f(x) + \limsup_{x \to p} g(x)$, provided the right side of this equation is not of the form $\pm \infty$ or $-\infty + \infty$.

3. $\lim_{x \to p} (f(x) + g(x)) = \liminf_{x \to p} f(x) + \limsup_{x \to p} g(x)$, provided the left side of this equation is not of the form $\infty - \infty$ or $-\infty + \infty$.

4. $\lim_{x \to p} f(x) = \limsup_{x \to p} f(x)$, provided the left side of this equation is not of the form $\infty - \infty$ or $-\infty + \infty$.

5. $\lim_{x \to p} (f(x) + g(x)) = \liminf_{x \to p} f(x) + \limsup_{x \to p} g(x)$, provided the right side of this equation is not of the form $\pm \infty$ or $-\infty + \infty$.

6. $\lim_{x \to p} f(x)$ exists in $\mathbb{R}$ then,

$$\lim_{x \to p} (f(x) + g(x)) = \lim_{x \to p} f(x) + \lim_{x \to p} g(x),$$

$$\lim_{x \to p} (f(x) + g(x)) = \lim_{x \to p} f(x) + \lim_{x \to p} g(x).$$

7. If $f(x), g(x) \geq 0$ on the intersection of an open ball with $E$ (or on $S_{p,\nu}$ for some $\nu > 0$) then

$$\lim_{x \to p} (f(x) + g(x)) \leq (\lim_{x \to p} f(x))(\lim_{x \to p} g(x)),$$

provided the right hand side of this equation is not of the form $0 \cdot \infty$ or $\infty \cdot 0$.

8. If $f(x), g(x) \geq 0$ on the intersection of an open ball with $E$ (or on $S_{p,\nu}$ for some $\nu > 0$) and $A = \lim_{x \to p} f(x)$ exists in $[0, \infty)$ then

$$\lim_{x \to p} (f(x) + g(x)) = (\lim_{x \to p} f(x))(\lim_{x \to p} g(x)).$$

Proof.

1. ($\Leftarrow$) Let $A := \liminf_{x \to p} f(x) = \limsup_{x \to p} f(x) \in \overline{\mathbb{R}}$. We may consider 3 cases:

   a) $A \in \mathbb{R}$. Let $\{x_n\}$ be a sequence in $E \setminus \{p\}$ such that $x_n \to p$ as $n \to \infty$. In what follows we will prove that $\lim_{n \to \infty} f(x_n) = A$ and hence by Theorem 6.33 we can conclude that $\lim_{x \to p} f(x) = A$. Let $\varepsilon > 0$ be given. By assumption there exists $\delta_1 > 0$ such that $A - \varepsilon \leq \inf_{x \in S_{p,\delta_1}} f(x)$ and there exists $\delta_2 > 0$ such that $\sup_{x \in S_{p,\delta_2}} f(x) \leq A + \varepsilon$. Let $\delta := \min\{\delta_1, \delta_2\}$. Since $x_n \to p$ and $\{x_n\} \subseteq E \setminus \{p\}$ we can conclude that $x_n \in S_{p,\delta}$ a.a.n. Therefore,

$$\forall \varepsilon > 0 \exists \delta > 0 \quad A - \varepsilon \leq \inf_{x \in S_{p,\delta}} f(x) \leq f(x_n) \leq \sup_{x \in S_{p,\delta}} f(x) \leq A + \varepsilon \quad \text{a.a.n}$$

This precisely means that $\lim_{n \to \infty} f(x_n) = A$.

b) $A = \infty$. Let $M > 0$ be given. Since $\lim_{x \to p} f(x) = \infty$ we may conclude that there exists $\nu > 0$ such that $\inf_{x \in S_{p,\nu}} f(x) > M$ provided $0 < \delta \leq \nu$. Therefore

$$\forall M > 0 \quad \exists \nu > 0 \quad \text{s.t.} \quad \forall x \in S_{p,\nu} \quad f(x) > M.$$ 

This precisely means that $\lim_{x \to p} f(x) = \infty$.

c) $A = -\infty$. Let $M < 0$ be given. Since $\lim_{x \to p} f(x) = -\infty$ we may conclude that there exists $\nu > 0$ such that $\sup_{x \in S_{p,\nu}} f(x) < M$ provided $0 < \delta \leq \nu$. Therefore

$$\forall M < 0 \quad \exists \nu > 0 \quad \text{s.t.} \quad \forall x \in S_{p,\nu} \quad f(x) < M.$$ 

This precisely means that $\lim_{x \to p} f(x) = -\infty$. 

Let $\delta > 0$ be arbitrary. It follows that $\limsup_{x \to p} f(x) \leq A + \epsilon$.

Passing to the limit (as $\delta \downarrow 0$) we get
\[ A - \epsilon \leq \liminf_{x \to p} f(x) \leq \limsup_{x \to p} f(x) \leq A + \epsilon. \]

Since $\epsilon > 0$ is arbitrary, it follows that
\[ A \leq \liminf_{x \to p} f(x) \leq \limsup_{x \to p} f(x) \leq A, \]

i.e. that $A = \liminf_{x \to p} f(x) = \limsup_{x \to p} f(x)$.

b) $A = -\infty$. Let $M > 0$ be given. There exists $\nu > 0$ such that $f(x) > M$ provided $x \in S_{p,\nu}$. So for all $0 < \delta \leq \nu$ we have $\inf_{x \in S_{p,\delta}} f(x) \geq M$. Passing to the limit (as $\delta \downarrow 0$) we get $\liminf_{x \to p} f(x) \geq M$. Since $M > 0$ is arbitrary it follows that $\liminf_{x \to p} f(x) = \infty$. Also since in general $\liminf_{x \to p} f(x) \leq \limsup_{x \to p} f(x)$ we can conclude that $\limsup_{x \to p} f(x) = \infty$.

c) $A = -\infty$. Let $M < 0$ be given. There exists $\nu > 0$ such that $f(x) < M$ provided $x \in S_{p,\nu}$. So for all $0 < \delta \leq \nu$ we have $\sup_{x \in S_{p,\delta}} f(x) \leq M$. Passing to the limit (as $\delta \downarrow 0$) we get $\limsup_{x \to p} f(x) \leq M$. Since $M < 0$ is arbitrary it follows that $\limsup_{x \to p} f(x) = -\infty$. Also since in general $\liminf_{x \to p} f(x) \leq \limsup_{x \to p} f(x)$ we can conclude that $\liminf_{x \to p} f(x) = -\infty$.

2. Let $\delta > 0$
\[ \forall x \in S_{p,\delta} \quad f(x) + g(x) \leq \sup_{y \in S_{p,\delta}} f(y) + \sup_{y \in S_{p,\delta}} g(y). \]
So
\[ \sup_{x \in S_{p,\delta}} (f(x) + g(x)) \leq \sup_{y \in S_{p,\delta}} f(y) + \sup_{y \in S_{p,\delta}} g(y). \]
Passing to the limit as $\delta \downarrow 0$ in this equation then implies
\[ \limsup_{x \to p} (f(x) + g(x)) \leq \limsup_{x \to p} f(x) + \limsup_{x \to p} g(x). \]

3. We have,
\[ \liminf_{x \to p} f(x) + \liminf_{x \to p} g(x) = -(\limsup_{x \to p} (-f(x)) + \limsup_{x \to p} (-g(x))) \leq -(\limsup_{x \to p} (-f(x) + g(x))) \]
\[ = \liminf_{x \to p} (f(x) + g(x)). \]

4. We may consider 3 cases:

a) $\liminf_{x \to p} f(x) \in \mathbb{R}$. We have
\[ \limsup_{x \to p} g(x) = \limsup_{x \to p} (g(x) + f(x) - f(x)) \]
\[ \leq \limsup_{x \to p} (f(x) + g(x)) + \limsup_{x \to p} (-f(x)) \]
\[ = \limsup_{x \to p} (f(x) + g(x)) - \liminf_{x \to p} (f(x)). \]

Therefore
\[ \liminf_{x \to p} f(x) + \limsup_{x \to p} g(x) \leq \limsup_{x \to p} (f(x) + g(x)). \]

b) $\liminf_{x \to p} f(x) = -\infty$. By assumption $\liminf_{x \to p} f(x) + \limsup_{x \to p} g(x)$ cannot be $\infty - \infty$, so either $\limsup_{x \to p} g(x) \in \mathbb{R}$ or $\limsup_{x \to p} g(x) = -\infty$. In both cases $\liminf_{x \to p} f(x) + \limsup_{x \to p} g(x) = -\infty$. So obviously the claimed inequality holds true.

c) $\liminf_{x \to p} f(x) = \infty$. By assumption $\liminf_{x \to p} f(x) + \limsup_{x \to p} g(x)$ cannot be $\infty - \infty$, so either $\limsup_{x \to p} g(x) \in \mathbb{R}$ or $\limsup_{x \to p} g(x) = \infty$. Also note that in general $\liminf_{x \to p} f(x) \leq \limsup_{x \to p} f(x)$, so we can conclude that $\limsup_{x \to p} f(x) = \infty$ and hence $\liminf_{x \to p} f(x) = \infty$. In what follows we will show that $\limsup_{x \to p} (f(x) + g(x)) = \infty$ and hence the desired inequality obviously holds true. First let’s consider the case where $A := \limsup_{x \to p} g(x) \in \mathbb{R}$. Let $M > 0$ be given.

\[ \limsup_{x \to p} f(x) = \infty \implies \exists \delta_1 > 0 \quad \forall x \in S_{p,\delta_1} \quad f(x) \geq M + |A|, \]
\[ A = \limsup_{x \to p} g(x) \implies \exists \delta_2 > 0 \quad \forall 0 < \nu \leq \delta_2 \quad \sup_{x \in S_{p,\nu}} g(x) > A/2. \]

Therefore if we let $\delta := \min\{\delta_1, \delta_2\}$ then
\[ \forall 0 < \nu \leq \delta \quad \exists x \in S_{p,\nu} \quad f(x) + g(x) \geq M + |A| + A/2 \geq M. \]

Consequently
\[ \forall 0 < \nu \leq \delta \quad \sup_{x \in S_{p,\nu}} (f(x) + g(x)) \geq M. \]
Passing to the limit (as $\nu \downarrow 0$) we get $\limsup_{x \to p}(f(x) + g(x)) \geq M$. Since $M$ is arbitrary we can conclude that $\limsup_{x \to p}(f(x) + g(x)) = \infty$.

Now let’s consider the case where $\limsup_{x \to p} g(x) = \infty$. Let $M > 0$ be given.

$$\lim_{x \to p} f(x) = \infty \implies \exists \delta_1 > 0 \implies \forall x \in S_{p,\delta_1} \quad f(x) > M/2,$$

$$\limsup_{x \to p} g(x) = \infty \implies \exists \delta_2 > 0 \implies \forall 0 < \nu \leq \delta_2 \quad \sup_{x \in S_{p,\nu}} g(x) > M/2.$$

Therefore if we let $\delta := \min\{\delta_1, \delta_2\}$ then

$$\forall 0 < \nu \leq \delta \quad \exists x \in S_{p,\nu} \quad f(x) + g(x) \geq M/2 + M/2 = M.$$

Consequently

$$\forall 0 < \nu \leq \delta \quad \sup_{x \in S_{p,\nu}} (f(x) + g(x)) \geq M.$$

Passing to the limit as $\nu \downarrow 0$ we get $\limsup_{x \to p}(f(x) + g(x)) \geq M$. Since $M$ is arbitrary we can conclude that $\limsup_{x \to p}(f(x) + g(x)) = \infty$.

By using item 4. we can write

$$\liminf_{x \to p}(g(x) + f(x)) = -\limsup_{x \to p}(-g(x) - f(x))$$

$$\leq -(\limsup_{x \to p}(-f(x)) + \liminf_{x \to p}(-g(x)))$$

$$= \liminf_{x \to p} f(x) + \limsup_{x \to p} g(x).$$

Similarly by item 3. and item 5. we have

$$\liminf_{x \to p} f(x) + \liminf_{x \to p} g(x) \leq \liminf_{x \to p}(f(x) + g(x)) \leq \limsup_{x \to p} f(x) + \liminf_{x \to p} g(x)$$

and so

$$\liminf_{x \to p} f(x) + \liminf_{x \to p} g(x) \leq \liminf_{x \to p}(f(x) + g(x)) \leq \limsup_{x \to p} f(x) + \liminf_{x \to p} g(x)$$

Therefore

$$\liminf_{x \to p}(f(x) + g(x)) = \liminf_{x \to p} f(x) + \liminf_{x \to p} g(x).$$

By assumption there exists $\nu > 0$ such that $f(x), g(x) \geq 0$ for all $x \in S_{p,\nu}$. For all $0 < \delta \leq \nu$ we have

$$\forall x \in S_{p,\delta} \quad f(x).g(x) \leq (\sup_{y \in S_{p,\delta}} f(y))(\sup_{y \in S_{p,\delta}} g(y)),$$

and therefore,

$$\sup_{x \in S_{p,\delta}} (f(x).g(x)) \leq (\sup_{y \in S_{p,\delta}} f(y))(\sup_{y \in S_{p,\delta}} g(y)).$$

Letting $\delta \downarrow 0$ in this inequality implies,

$$\limsup_{x \to p} (f(x).g(x)) \leq \lim\sup_{x \to p} \left(\sup_{y \in S_{p,\delta}} f(y)\right)\left(\sup_{y \in S_{p,\delta}} g(y)\right)$$

$$= \lim_{\delta \downarrow 0} \sup_{y \in S_{p,\delta}} f(y) \lim_{\delta \downarrow 0} \sup_{y \in S_{p,\delta}} g(y)$$

$$= \left(\lim_{x \to p} f(x)\right)\left(\lim_{x \to p} g(x)\right).$$

provided $(\limsup_{x \to p} f(x))(\limsup_{x \to p} g(x))$ is not of the form $0 \cdot \infty$ or $\infty \cdot 0$.

Note that by item 1. $\liminf_{x \to p} f(x) = \limsup_{x \to p} f(x) = \lim_{x \to p} f(x)$; so considering the previous item we just need to show that

$$\limsup_{x \to p}(f(x).g(x)) \geq A.\limsup_{x \to p} g(x).$$

Let $\alpha \in (0, A]$. By assumption there exists $\nu_1 > 0$ such that $f(x), g(x) \geq 0$ for all $x \in S_{p,\nu_1}$. Since $\lim_{x \to p} f(x) = A$ there exists $\nu_2 > 0$ such that $f(x) \geq \alpha$ for all $x \in S_{p,\nu_2}$ and hence $f(x).g(x) \geq \alpha g(x)$ for all $x \in S_{p,\nu}$ where $\nu = \min\{\nu_1, \nu_2\}$. Therefore by item 3. of Proposition D.8 we can conclude that

$$\limsup_{x \to p}(f(x).g(x)) \geq \limsup_{x \to p}(\alpha g(x)) = \alpha.\limsup_{x \to p} g(x).$$

Now we may consider two cases:
a) If \( \limsup_{x \to p} g(x) < \infty \) we may let \( \alpha \uparrow A \) in order to see that
\[
\limsup_{x \to p} (f(x)g(x)) \geq A \limsup_{x \to p} g(x).
\]
b) If \( \limsup_{x \to p} g(x) = \infty \) it follows that
\[
\limsup_{x \to p} (f(x)g(x)) = A \limsup_{x \to p} g(x) = \alpha \limsup_{x \to p} g(x) = \infty.
\]

Remark D.10. By item 1. of the previous proposition if limit exists, then limsup and liminf are both equal to the limit. So whenever the limit exists we may replace the limit by limsup and liminf. The important point is that even if the limit does not exist, limsup and liminf always exist. Therefore if we want to take limit but we do not know whether the limit exists we should take limsup or liminf instead.

Exercise D.2. Prove that if \( f(x) \geq 0 \) then \( \lim sup_{x \to p} f(x) = 0 \) if and only if \( \lim_{x \to p} f(x) = 0 \). Conclude that, even if \( f \) is not necessarily nonnegative, \( \lim_{x \to p} f(x) = c \) if \( \lim sup_{x \to p} |f(x) - c| = 0 \).

Exercise D.3. Prove the followings:

1. If \( f(x), g(x) \geq 0 \) on the intersection of an open ball about \( p \) with \( E \) (or on \( S_{p, \nu} \) for some \( \nu > 0 \)) then
\[
\liminf_{x \to p} (f(x)g(x)) \geq (\liminf_{x \to p} f(x))(\liminf_{x \to p} g(x)),
\]
provided the right hand side of this equation is not of the form \( 0 \cdot \infty \) or \( \infty \cdot 0 \).

2. If \( f(x), g(x) \geq 0 \) on the intersection of an open ball about \( p \) with \( E \) (or on \( S_{p, \nu} \) for some \( \nu > 0 \)) and \( A = \lim_{x \to p} f(x) \) exists in \((0, \infty)\) then
\[
\liminf_{x \to p} (f(x)g(x)) = (\liminf_{x \to p} f(x))(\liminf_{x \to p} g(x)).
\]
The claim of the following theorem is similar to that of Exercise 3.14.

Theorem D.11. Let \( (X, d) \) be a metric space. Suppose \( E \subseteq X \) and \( p \) is a limit point of \( E \). Let \( f \) be a real (or extended real) valued function on \( E \) and suppose \( A \in \mathbb{R} \).

1. \( \lim_{x \to p} f(x) \leq A \) if and only if for every \( \varepsilon > 0 \) \( \exists \delta > 0 \) such that \( f(x) \leq A + \varepsilon \) provided \( x \in S_{p, \delta} \).

2. \( \limsup_{x \to p} f(x) \geq A \) if and only if for every \( \varepsilon > 0 \) \( \forall \delta > 0 \exists x_{\delta, \varepsilon} \in S_{p, \delta} \) such that \( f(x_{\delta, \varepsilon}) \geq A - \varepsilon \).

3. \( \limsup_{x \to p} f(x) = A \) if and only if for every \( \varepsilon > 0 \) \( \exists \delta > 0 \) such that \( f(x) \leq A + \varepsilon \) provided \( x \in S_{p, \delta} \).

Proof.

1. \( \Rightarrow \) Let \( \varepsilon > 0 \) be given, \( \lim_{x \to p} f(x) \leq A \) so there exists \( \delta > 0 \) such that for all \( 0 < \nu \leq \delta \), \( \sup_{x \in S_{p, \nu}} f(x) \leq A + \varepsilon \) so in particular \( f(x) \leq A + \varepsilon \) provided \( x \in S_{p, \delta} \).

2. \( \Leftarrow \) Conversely, if \( \varepsilon > 0 \) then by assumption there exists \( \delta > 0 \) such that \( f(x) \leq A + \varepsilon \) on \( S_{p, \delta} \). Therefore, by item 3. of Proposition D.8\( \limsup_{x \to p} f(x) \leq \limsup_{x \to p} f(A + \varepsilon) = \lim_{x \to p} f(A + \varepsilon) = A + \varepsilon \).

3. \( \Rightarrow \) Let \( \varepsilon > 0 \) be given. Since \( \inf_{x \in S_{p, \delta}} f(x) \geq A \) we can conclude that for all \( \delta > 0 \), \( \sup_{x \in S_{p, \delta}} f(x) \geq A \). Therefore for all \( \delta > 0 \) there exists \( x_{\delta, \varepsilon} \in S_{p, \delta} \) such that \( f(x_{\delta, \varepsilon}) \geq A - \varepsilon \) (otherwise \( \sup_{x \in S_{p, \delta}} f(x) < A - \varepsilon \)).

4. \( \Leftarrow \) Conversely, if \( \varepsilon > 0 \) then the assumption implies that for all \( \delta > 0 \), \( \sup_{x \in S_{p, \delta}} f(x) \geq A - \varepsilon \). Therefore \( \limsup_{x \to p} f(x) = \inf_{\delta > 0} \sup_{x \in S_{p, \delta}} f(x) \geq A - \varepsilon \).

5. This one is a direct consequence of the previous items.

Exercise D.4. Let \( (X, d) \) be a metric space. Suppose \( E \subseteq X \) and \( p \) is a limit point of \( E \). Let \( f \) be a real (or extended real) valued function on \( E \) and suppose \( A \in \mathbb{R} \).

1. \( \lim_{x \to p} f(x) \geq A \) if and only if for every \( \varepsilon > 0 \) \( \exists \delta > 0 \) such that \( f(x) \geq A - \varepsilon \) provided \( x \in S_{p, \delta} \).

2. \( \liminf_{x \to p} f(x) \leq A \) if and only if for every \( \varepsilon > 0 \) \( \forall \delta > 0 \exists x_{\delta, \varepsilon} \in S_{p, \delta} \) such that \( f(x_{\delta, \varepsilon}) \leq A + \varepsilon \).

3. \( \liminf_{x \to p} f(x) = A \) if and only if for every \( \varepsilon > 0 \) \( \exists \delta > 0 \) such that \( f(x) \leq A - \varepsilon \) provided \( x \in S_{p, \delta} \).

Theorem D.12. Let \( (X, d) \) be a metric space. Suppose \( E \subseteq X \) and \( p \) is a limit point of \( E \). Let \( f \) be a real (or extended real) valued function on \( E \). Suppose \( \limsup_{x \to p} f(x) = A \in \mathbb{R} \).

1. If \( \{x_n\}_{n=1}^{\infty} \) is a sequence in \( E - \{p\} \) such that \( p = \lim_{n \to \infty} x_n \), then \( \limsup_{n \to \infty} f(x_n) \leq A \).

2. There exists a sequence \( \{x_n\}_{n=1}^{\infty} \in E - \{p\} \) such that \( p = \lim_{n \to \infty} x_n \) and \( A = \lim_{n \to \infty} f(x_n) \).

Proof.
1. We may consider 3 cases:
   Case 2; $A = -\infty$. In general $\lim\inf_{x\to p} f(x) \leq \lim\sup_{x\to p} f(x)$ so we may conclude that $\lim\inf_{x\to p} f(x) = \lim\sup_{x\to p} f(x) = -\infty$ and hence $\lim_{x\to p} f(x) = -\infty$. Since $\{x_n\} \subseteq E - \{p\}$ and $x_n \to p$, it follows that $\lim_{x\to \infty} f(x_n) = -\infty$. Therefore $\lim\sup_{n\to\infty} f(x_n) = -\infty \leq A = -\infty$.

   Case 3. $A \in \mathbb{R}$. Let $\varepsilon > 0$ be given. By Theorem D.11 there exists $\delta > 0$ such that $f(x) \leq A + \varepsilon$ for all $x \in S_{p,\delta}$. Since $\{x_n\} \subseteq E - \{p\}$ and $x_n \to p$, we have $x_n \in S_{p,\delta}$ a.a.n. Therefore $f(x_n) \leq A + \varepsilon$ a.a.n. Consequently $\lim\sup_{n\to\infty} f(x_n) = \lim\sup_{n\to\infty}(A + \varepsilon) = A + \varepsilon$. Since $\varepsilon > 0$ is arbitrary we can conclude that $\lim\sup_{n\to\infty} f(x_n) \leq A$.

2. By item ii. of Theorem D.11 for all $n \geq 1$ (let $\delta = \varepsilon = 1/n$), there exists $x_n \in S_{p,1/n}$ such that $f(x_n) \geq A - 1/n$. Therefore

   $$x_n \in S_{p,1/n} \implies d(x_n,p) < 1/n, \{x_n\} \subseteq E - \{p\} \implies x_n \to p, \{x_n\} \subseteq E - \{p\}$$

Moreover,

$$f(x_n) \geq A - 1/n \implies \lim\inf_{n\to\infty} f(x_n) \geq \lim\inf(A - 1/n) = \lim_{n\to\infty} (A - 1/n) = A$$

But by item i. $\lim\sup_{n\to\infty} f(x_n) \leq A$ and so

$$A \leq \lim\inf_{n\to\infty} f(x_n) \leq \lim\sup_{n\to\infty} f(x_n) \leq A$$

Consequently $\lim\inf_{n\to\infty} f(x_n) = \lim\sup_{n\to\infty} f(x_n) = A$ which implies $\lim_{x\to \infty} f(x_n) = A$.

\[\square\]

Exercise D.5. Let $(X, d)$ be a metric space. Suppose $E \subseteq X$ and $p$ is a limit point of $E$. Let $f$ be a real (or extended real) valued function on $E$. Suppose $\lim\inf_{x\to p} f(x) = A \in \mathbb{R}$.

1. If $\{x_n\}_{n=1}^\infty$ is a sequence in $E - \{p\}$ such that $p = \lim_{n\to\infty} x_n$, then $\lim\inf_{n\to\infty} f(x_n) \geq A$.

2. There exists a sequence $\{x_n\}_{n=1}^\infty$ in $E - \{p\}$ such that $p = \lim_{n\to\infty} x_n$ and $A = \lim_{n\to\infty} f(x_n)$.

Remark D.13. Considering the above theorem and exercise, one can give the following interpretation of $\lim\inf_{x\to p} f(x)$ and $\lim\sup_{x\to p} f(x)$:

- $\lim\sup_{x\to p} f(x)$ is the largest limiting value of $\{f(x_n)\}_{n=1}^\infty$ over all sequences $\{x_n\}_{n=1}^\infty \subseteq E - \{p\}$ that converge to $\{p\}$.
- $\lim\inf_{x\to p} f(x)$ is the smallest limiting value of $\{f(x_n)\}_{n=1}^\infty$ over all sequences $\{x_n\}_{n=1}^\infty \subseteq E - \{p\}$ that converge to $\{p\}$.

Definition D.14. Let $X = \mathbb{R}$ and $E = (a, b)$ be an open interval in $\mathbb{R}$ containing $p$ (or let $E = (a, b) - \{p\}$ where $p \in (a, b)$). Let $f$ be a real (or extended real) valued function on $E$. We define

$$\lim\sup_{x\to p} f(x) = \lim_{\delta \to 0} \\sup_{0<\delta<\delta} f(x),$$

$$\lim\inf_{x\to p} f(x) = \lim_{\delta \to 0} \\inf_{0<\delta<\delta} f(x),$$

$$\lim\sup_{x\to p} f(x) = \lim_{\delta \to 0} \\sup_{0<\delta<\delta} f(x),$$

$$\lim\inf_{x\to p} f(x) = \lim_{\delta \to 0} \\inf_{0<\delta<\delta} f(x).$$

(Note that the set $E$ can be unbounded.)

Remark D.15. Note that clearly if we let $f_1 := f|_{(a,p)}$ and $f_2 := f|_{(p,b)}$ then

$$\lim\sup_{x\to p} f(x) = \lim\sup_{x\to p} f_2(x), \quad \lim\inf_{x\to p} f(x) = \lim\inf_{x\to p} f_1(x).$$

So in fact the properties that were previously mentioned for limsup and liminf of functions (e.g. Proposition 10) will hold true for one sided limsup and liminf as well (with obvious modifications).

Also it will be helpful to note that $\lim_{x\to p} f(x) = A \in \mathbb{R}$ iff $\lim_{x\to p} f_1(x) = A \in \mathbb{R}$. Similarly $\lim_{x\to p} f(x) = A \in \mathbb{R}$ iff $\lim_{x\to p} f_2(x) = A \in \mathbb{R}$.

Corollary D.16. An immediate consequence of the above remark, item 1. of Proposition D.9 and Theorem 6.41 is that if $f$ is a real valued function on the open interval $E = (a, b)$ which contains $p$ (or on $E = (a, b) - \{p\}$ where $p \in (a, b)$) then

1. $\lim_{x\to p} f(x) = A \in \mathbb{R}$ if $\lim\sup_{x\to p} f(x) = \lim\inf_{x\to p} f(x) = A \in \mathbb{R}$.
2. $\lim_{x\to p} f(x) = A \in \mathbb{R}$ if $\lim\sup_{x\to p} f(x) = \lim\inf_{x\to p} f(x) = A \in \mathbb{R}$.
3. $\lim_{x\to p} f(x) = A \in \mathbb{R}$ if $\lim\sup_{x\to p} f(x) = \lim\inf_{x\to p} f(x) = A \in \mathbb{R}$.

Theorem D.17. Let $X = \mathbb{R}$ and $E = (a, b)$ be an open interval in $\mathbb{R}$ containing $p$ (or let $E = (a, b) - \{p\}$ where $p \in (a, b)$). If $f$ is a real valued function on $E$ then

$$\lim\sup_{x\to p} f(x) = \max\{\lim\sup_{x\to p} f(x), \lim\sup_{x\to p} f(x)\}.$$
Proof. We have
\[\forall \delta > 0 \exists x_\delta \in \mathbb{R} \text{ such that } |x - p| < \delta \Rightarrow f(x) = 1 \implies \sup_{0 < x - p < \delta} f(x) = 1 \implies \limsup_{x \to p^-} f(x) = 1,\]
\[\forall \delta > 0 \exists x_\delta \in \mathbb{R} \text{ such that } |x - p| < \delta \Rightarrow f(x) = 1 \implies \sup_{0 < x - p < \delta} f(x) = 1 \implies \limsup_{x \to p^+} f(x) = 1,\]
\[\forall \delta > 0 \exists x_\delta \in \mathbb{R} \text{ such that } |x - p| < \delta \Rightarrow \inf_{0 < x - p < \delta} f(x) = 0 \implies \liminf_{x \to p^-} f(x) = 0,\]
\[\forall \delta > 0 \exists x_\delta \in \mathbb{R} \text{ such that } |x - p| < \delta \Rightarrow \inf_{0 < x - p < \delta} f(x) = 0 \implies \liminf_{x \to p^+} f(x) = 0.\]

Also by Theorem D.17 and Exercise D.6, \(\limsup_{x \to p} f(x) = \max\{1, 1\} = 1\) and \(\liminf_{x \to p} f(x) = \min\{0, 0\} = 0\). In particular, considering Corollary D.16, we can conclude that \(\lim_{x \to p} f(x)\) does not exist.

Example D.19. Let
Exercise D.8. Prove the followings:

1. If \( f(x) \geq 0 \) on \([M, \infty)\) for some \( M > 0 \) then
   \[
   \lim_{x \to \infty} f(x) = c \text{ if and only if } \lim_{x \to \infty} f(x) = c.
   \]

2. \( \lim_{x \to \infty} f(x) = \lim_{x \to \infty} g(x) \) provided the right side of this equation is not of the form \( \infty - \infty \) or \(-\infty + \infty\).

3. \( \lim_{x \to \infty} f(x) + \lim_{x \to \infty} g(x) \leq \lim_{x \to \infty} (f(x) + g(x)) \) provided the left side of this equation is not of the form \( \infty - \infty \) or \(-\infty + \infty\).

4. \( \lim_{x \to \infty} f(x) + \lim_{x \to \infty} g(x) \leq \lim_{x \to \infty} (f(x) + g(x)) \) provided the left side of this equation is not of the form \( \infty - \infty \) or \(-\infty + \infty\).

5. \( \lim_{x \to \infty} (f(x) + g(x)) \leq \lim_{x \to \infty} f(x) + \lim_{x \to \infty} g(x) \) provided the right side of this equation is not of the form \( \infty - \infty \) or \(-\infty + \infty\).

6. \( \lim_{x \to \infty} f(x) \) exists in \( \mathbb{R} \) then
   \[
   \limsup_{x \to \infty} (f(x) + g(x)) = \lim_{x \to \infty} f(x) + \limsup_{x \to \infty} g(x),
   \]
   \[
   \liminf_{x \to \infty} (f(x) + g(x)) = \lim_{x \to \infty} f(x) + \liminf_{x \to \infty} g(x).
   \]

7. If \( f(x), g(x) \geq 0 \) on \([M, \infty)\) for some \( M > 0 \), then
   \[
   \limsup_{x \to \infty} f(x)g(x) \leq (\limsup_{x \to \infty} f(x))(\limsup_{x \to \infty} g(x)),
   \]
   provided the right hand side of this equation is not of the form \( 0 \cdot \infty \) or \( \infty \cdot 0 \).

8. If \( f(x), g(x) \geq 0 \) on \([M, \infty)\) for some \( M > 0 \) and \( \lim_{x \to \infty} f(x) \) exists in \((0, \infty)\) then
   \[
   \limsup_{x \to \infty} f(x)g(x) = (\lim_{x \to \infty} f(x))(\limsup_{x \to \infty} g(x)).
   \]

(Proof is completely analogous to the proofs of the corresponding items of Proposition D.9)

Exercise D.9. Prove that if \( f(x) \geq 0 \), then \( \limsup_{x \to \infty} f(x) = 0 \) if and only if \( \lim_{x \to \infty} f(x) = 0 \). Conclude that, even if \( f \) is not necessarily nonnegative, \( \lim_{x \to \infty} f(x) = c \) if and only if \( \limsup_{x \to \infty} f(x) - c = 0 \).

Exercise D.8. Prove the followings:

1. If \( f(x), g(x) \geq 0 \) on \([M, \infty)\) for some \( M > 0 \) then
   \[
   \liminf_{x \to \infty} (f(x), g(x)) \geq (\liminf_{x \to \infty} f(x))(\liminf_{x \to \infty} g(x)),
   \]
   provided the right hand side of this equation is not of the form \( 0 \cdot \infty \) or \( \infty \cdot 0 \).

2. If \( f(x), g(x) \geq 0 \) on \([M, \infty)\) for some \( M > 0 \) and \( \limsup_{x \to \infty} f(x) \) exists in \((0, \infty)\) then
   \[
   \liminf_{x \to \infty} (f(x), g(x)) = (\liminf_{x \to \infty} f(x))(\liminf_{x \to \infty} g(x)).
   \]

Theorem D.23. Let \( f \) be a real (or extended real) valued function on \( \mathbb{R} \) and suppose \( A \in \mathbb{R} \).

1. \( \limsup_{x \to \infty} f(x) \leq A \) if and only if for every \( \epsilon > 0 \) \( \exists M > 0 \) such that \( f(x) \leq A + \epsilon \) provided \( x \geq M \).

2. \( \limsup_{x \to \infty} f(x) \geq A \) if and only if for every \( \epsilon > 0 \) \( \forall M > 0 \exists x_{M,\epsilon} \geq M \) such that \( f(x_{M,\epsilon}) \geq A - \epsilon \).

3. \( \limsup_{x \to \infty} f(x) = A \) if and only if for every \( \epsilon > 0 \) \( \exists M > 0 \) such that \( f(x) \geq A + \epsilon \) provided \( x \geq M \), \( \forall M > 0 \exists x_{M,\epsilon} \geq M \) such that \( f(x_{M,\epsilon}) \leq A - \epsilon \).

(Proof is completely analogous to the proof of Theorem D.11)

Exercise D.9. Let \( f \) be a real (or extended real) valued function on \( \mathbb{R} \) and suppose \( A \in \mathbb{R} \).

1. \( \liminf_{x \to \infty} f(x) \geq A \) if and only if for every \( \epsilon > 0 \) \( \exists M > 0 \) such that \( f(x) \geq A - \epsilon \) provided \( x \geq M \).

2. \( \liminf_{x \to \infty} f(x) \leq A \) if and only if for every \( \epsilon > 0 \) \( \forall M > 0 \exists x_{M,\epsilon} \geq M \) such that \( f(x_{M,\epsilon}) \leq A + \epsilon \).

3. \( \liminf_{x \to \infty} f(x) = A \) if and only if for every \( \epsilon > 0 \) \( \exists M > 0 \) such that \( f(x) \leq A - \epsilon \) provided \( x \geq M \), \( \forall M > 0 \exists x_{M,\epsilon} \geq M \) such that \( f(x_{M,\epsilon}) \geq A + \epsilon \).

Theorem D.24. Let \( f \) be a real (or extended real) valued function on \( \mathbb{R} \). Suppose \( \limsup_{x \to \infty} f(x) = A \in \mathbb{R} \).

1. If \( \{x_n\} \to \infty \) is a sequence of real numbers such that \( \lim_{n \to \infty} x_n = \infty \), then \( \limsup_{n \to \infty} f(x_n) \leq A \).

2. There exists a sequence \( \{x_n\} \to \infty \) of real numbers such that \( \lim_{n \to \infty} x_n = \infty \) and \( A = \lim_{n \to \infty} f(x_n) \).

(Proof is completely analogous to the proof of Theorem D.12)

Exercise D.10. Let \( f \) be a real (or extended real) valued function on \( \mathbb{R} \). Suppose \( \liminf_{x \to \infty} f(x) = A \in \mathbb{R} \).

1. If \( \{x_n\} \to \infty \) is a sequence of real numbers such that \( \lim_{n \to \infty} x_n = \infty \), then \( \liminf_{n \to \infty} f(x_n) \geq A \).

2. There exists a sequence \( \{x_n\} \to \infty \) of real numbers such that \( \lim_{n \to \infty} x_n = \infty \) and \( A = \lim_{n \to \infty} f(x_n) \).

Example D.25. Let \( f(x) = \sin x \). We have

\[
\forall t \sup_{x \geq t} f(x) = 1 \iff \lim_{t \to \infty} \sup_{x \geq t} f(x) = 1 \iff \lim_{x \to \infty} f(x) = 1,
\]
\[
\forall t \inf_{x \geq t} f(x) = -1 \iff \lim_{t \to \infty} \inf_{x \geq t} f(x) = -1 \iff \lim_{x \to \infty} f(x) = -1.
\]
Example D.26. Let \( f(x) = x \sin(x^2) \).

For \( \{x_n = \sqrt{2n\pi + \pi/2}\}_{n=1}^{\infty} \) we have \( \lim_{n \to \infty} x_n = \infty \) and

\[
\lim_{n \to \infty} f(x_n) = \lim_{n \to \infty} \sqrt{2n\pi + \pi/2} \sin(2n\pi + \pi/2) = \lim_{n \to \infty} \sqrt{2n\pi + \pi/2} = \infty
\]

So by Theorem D.24 we can conclude that

\[
\infty = \lim_{n \to \infty} f(x_n) = \lim_{n \to \infty} \sup f(x_n) \leq \lim_{x \to \infty} f(x)
\]

and therefore \( \limsup_{x \to \infty} f(x) = \infty \).

Similarly, for \( \{x_n = \sqrt{2n\pi + 3\pi/2}\}_{n=1}^{\infty} \) we have \( \lim_{n \to \infty} x_n = \infty \) and

\[
\lim_{n \to \infty} f(x_n) = \lim_{n \to \infty} \sqrt{2n\pi + \frac{3\pi}{2}} \sin \left(2n\pi + \frac{3\pi}{2}\right) = \lim_{n \to \infty} -\sqrt{2n\pi + 3\pi/2} = -\infty
\]

So by Exercise D.10 we can conclude that

\[
-\infty = \lim_{n \to \infty} f(x_n) = \lim_{n \to \infty} \inf f(x_n) \geq \lim_{x \to \infty} f(x)
\]

and therefore \( \liminf_{x \to \infty} f(x) = -\infty \).
Appendix: Countably Additive Measures

Let \( A \subset 2^\Omega \) be an algebra and \( \mu : A \to [0, \infty) \) be a finitely additive measure. Recall that \( \mu \) is a premeasure on \( A \) if \( \mu \) is \( \sigma \)– additive on \( A \). If \( \mu \) is a premeasure on \( A \) and \( A \) is a \( \sigma \)– algebra (Definition ??), we say that \( \mu \) is a measure on \((\Omega, A)\) and that \((\Omega, A)\) is a measurable space.

Definition E.1. Let \((\Omega, B)\) be a measurable space. We say that \( P : B \to [0, 1] \) is a probability measure on \((\Omega, B)\) if \( P \) is a measure on \( B \) such that \( P(\Omega) = 1 \). In this case we say that \((\Omega, B, P)\) a probability space.

E.1 Overview

The goal of this chapter is to develop methods for proving the existence of probability measures with desirable properties. The main results of this chapter may be summarized in the following theorem.

Theorem E.2. A finitely additive probability measure \( P \) on an algebra, \( A \subset 2^\Omega \), extends to \( \sigma \)– additive measure on \( \sigma(A) \) iff \( P \) is a premeasure on \( A \). If the extension exists it is unique.

Proof. The uniqueness assertion is proved Proposition E.15 below. The existence assertion of the theorem is a corollary of Theorem E.27.

In order to use this theorem it is necessary to determine when a finitely additive probability measure in is in fact a premeasure. The following Proposition is sometimes useful in this regard.

Proposition E.3 (Equivalent premeasure conditions). Suppose that \( P \) is a finitely additive probability measure on an algebra, \( A \subset 2^\Omega \). Then the following are equivalent:

1. \( P \) is a premeasure on \( A \), i.e. \( P \) is \( \sigma \)– additive on \( A \).
2. For all \( A_n \in A \) such that \( A_n \uparrow A \in A \), \( P(A_n) \uparrow P(A) \).
3. For all \( A_n \in A \) such that \( A_n \downarrow A \in A \), \( P(A_n) \downarrow P(A) \).
4. For all \( A_n \in A \) such that \( A_n \uparrow \Omega \), \( P(A_n) \uparrow 1 \).
5. For all \( A_n \in A \) such that \( A_n \downarrow \emptyset \), \( P(A_n) \downarrow 0 \).

Proof. We will start by showing 1 \( \iff \) 2 \( \iff \) 3.

1. \( \implies \) 2. Suppose \( A_n \in A \) such that \( A_n \uparrow A \in A \). Let \( A'_n := A_n \setminus A_{n-1} \) with \( A_0 := \emptyset \). Then \( \{A'_n\}_{n=1}^\infty \) are disjoint, \( A_n = \cup_{k=1}^n A'_k \) and \( A = \cup_{k=1}^\infty A'_k \). Therefore,

\[
P(A) = \sum_{k=1}^\infty P(A'_k) = \lim_{n \to \infty} \sum_{k=1}^n P(A'_k) = \lim_{n \to \infty} P(\cup_{k=1}^n A'_k) = \lim_{n \to \infty} P(A_n).
\]

2. \( \implies \) 1. If \( \{A_n\}_{n=1}^\infty \subset A \) are disjoint and \( A := \cup_{n=1}^\infty A_n \in A \), then \( \cup_{n=1}^\infty A_n \uparrow A \). Therefore,

\[
P(A) = \lim_{N \to \infty} P(\bigcup_{n=1}^N A_n) = \lim_{N \to \infty} \sum_{n=1}^N P(A_n) = \sum_{n=1}^\infty P(A_n).
\]

3. \( \implies \) 2. If \( A_n \in A \) such that \( A_n \uparrow A \in A \), then \( A'_n \uparrow A \) and we again have

\[
\lim_{n \to \infty} (1 - P(A_n)) = \lim_{n \to \infty} P(A'_n) = P(A) = 1 - P(A).
\]

The same proof used for 2. \( \iff \) 3. shows 4. \( \iff \) 5. and it is clear that

3. \( \implies \) 5. To finish the proof we will show 5. \( \implies \) 2.

5. \( \implies \) 2. If \( A_n \in A \) such that \( A_n \uparrow A \in A \), then \( A \setminus A_n \downarrow \emptyset \) and therefore

\[
\lim_{n \to \infty} [P(A) - P(A_n)] = \lim_{n \to \infty} P(A \setminus A_n) = 0.
\]

Remark E.4. Observe that the equivalence of items 1. and 2. in the above proposition hold without the restriction that \( P(\Omega) = 1 \) and in fact \( P(\Omega) = \infty \) may be allowed for this equivalence.

Lemma E.5. If \( \mu : A \to [0, \infty] \) is a premeasure, then \( \mu \) is countably subadditive on \( A \).
Lemma E.8. Suppose that \( A_n \in \mathcal{A} \) with \( \bigcup_{n=1}^{\infty} A_n \in \mathcal{A} \). Let \( A'_n := A_n \setminus (A_1 \cup \ldots \cup A_{n-1}) \) in \( \mathcal{A} \). Then \( \bigcup_{n=1}^{\infty} A'_n = \sum_{n=1}^{\infty} A'_n \) and therefore by the countable additivity and monotonicity of \( \mu \) we have,

\[
\mu \left( \bigcup_{n=1}^{\infty} A_n \right) = \mu \left( \sum_{n=1}^{\infty} A'_n \right) = \sum_{n=1}^{\infty} \mu (A'_n) \leq \sum_{n=1}^{\infty} \mu (A_n).
\]

Proof. Suppose that \( A_n \in \mathcal{A} \) with \( \bigcup_{n=1}^{\infty} A_n \in \mathcal{A} \). Let \( A'_n := A_n \setminus (A_1 \cup \ldots \cup A_{n-1}) \) in \( \mathcal{A} \). Then \( \bigcup_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} A'_n \) and therefore by the countable additivity and monotonicity of \( \mu \) we have,

\[
\mu \left( \bigcup_{n=1}^{\infty} A_n \right) = \mu \left( \sum_{n=1}^{\infty} A'_n \right) = \sum_{n=1}^{\infty} \mu (A'_n) \leq \sum_{n=1}^{\infty} \mu (A_n).
\]

Let us now specialize to the case where \( \Omega = \mathbb{R} \) and \( \mathcal{A} = \mathcal{A} (\{ (a, b] \cap \mathbb{R} : -\infty \leq a \leq b \leq \infty \}) \). In this case we will describe probability measures, \( P \), on \( \mathcal{B}_\mathbb{R} \) by their “cumulative distribution functions.”

Definition E.6. Given a probability measure, \( P \) on \( \mathcal{B}_\mathbb{R} \), the cumulative distribution function (CDF) of \( P \) is defined as the function, \( F = F_P : \mathbb{R} \to [0, 1] \) given as

\[
F(x) := P((\infty, x]).
\]

Example E.7. Suppose that

\[
P = p\delta_{-1} + q\delta_1 + r\delta_\pi
\]

with \( p, q, r > 0 \) and \( p + q + r = 1 \). In this case,

\[
F(x) = \begin{cases} 
0 & \text{for} \quad x < -1 \\
p & \text{for} \quad -1 \leq x < 1 \\
p + q & \text{for} \quad 1 \leq x < \pi \\
p + q + r & \text{for} \quad \pi \leq x < \infty
\end{cases}
\]

A plot of \( F(x) \) with \( p = .2 \), \( q = .3 \), and \( r = .5 \).

Lemma E.8. If \( F = F_P : \mathbb{R} \to [0, 1] \) is a distribution function for a probability measure, \( P \), on \( \mathcal{B}_\mathbb{R} \), then:

1. \( F \) is non-decreasing,
2. \( F \) is right continuous,
3. \( F(-\infty) := \lim_{x \to -\infty} F(x) = 0 \), and \( F(\infty) := \lim_{x \to \infty} F(x) = 1 \).

Proof. The monotonicity of \( F \) follows from the definition of the CDF. The continuity of \( F \) follows from the definition of the CDF. Since \( \{ b_n \}_{n=1}^{\infty} \) is an arbitrary sequence such that \( b_n \downarrow b \), we have shown \( F(b) := \lim_{y \uparrow b} F(y) = F(b) \). This shows that \( F \) is right continuous. Similar arguments show that \( F(\infty) = 1 \) and \( F(-\infty) = 0 \). It turns out that Lemma E.8 has the following important converse.

Theorem E.9. To each function \( F : \mathbb{R} \to [0, 1] \) satisfying properties 1. – 3. in Exercise E.2 and ?? below. The existence portion of the theorem is a special case of Theorem E.33 below. The uniqueness assertion is proved in Corollary E.17 below or see Exercise E.3 below.

Example E.10 (Uniform Distribution). The function,

\[
F(x) := \begin{cases} 
0 & \text{for} \quad x \leq 0 \\
x & \text{for} \quad 0 \leq x < 1 \\
1 & \text{for} \quad 1 \leq x < \infty
\end{cases}
\]

is the distribution function for a measure, \( m \) on \( \mathcal{B}_\mathbb{R} \) which is concentrated on \( (0, 1] \). The measure, \( m \) is called the uniform distribution or Lebesgue measure on \( (0, 1] \).

With this summary in hand, let us now start the formal development. We begin with uniqueness statement in Theorem E.2.

E.2 \( \pi - \lambda \) Theorem

Recall that a collection, \( \mathcal{P} \subset 2^\Omega \), is a \( \pi \)-class or \( \pi \)-system if it is closed under finite intersections. We also need the notion of a \( \lambda \)-system.

Definition E.11 (\( \lambda \)-system). A collection of sets, \( \mathcal{L} \subset 2^\Omega \), is a \( \lambda \)-class or \( \lambda \)-system if

a. \( \Omega \in \mathcal{L} \)
Lemma E.13 (Alternate Axioms for a \( \lambda \) – System*). Suppose that \( \mathcal{L} \subset 2^\Omega \) is a collection of subsets of \( \Omega \). Then \( \mathcal{L} \) is a \( \lambda \) – class iff \( \lambda \) satisfies the following postulates:

1. \( \Omega \in \mathcal{L} \)
2. \( A \in \mathcal{L} \) implies \( A^c \in \mathcal{L} \). (Closed under complementation.)
3. If \( \{A_n\}_{n=1}^\infty \subset \mathcal{L} \) are disjoint, then \( \sum_{n=1}^\infty A_n \in \mathcal{L} \). (Closed under countable increasing unions.)

Proof. Suppose that \( \mathcal{L} \) satisfies a. – c. above. Clearly then postulates 1. and 2. hold. Suppose that \( A, B \in \mathcal{L} \) such that \( A \cap B = \emptyset \), then \( A \subset B^c \) and

\[
A^c \cap B^c = B^c \setminus A \in \mathcal{L}.
\]

Taking complements of this result shows \( A \cup B \in \mathcal{L} \) as well. So by induction, \( B_m := \sum_{n=1}^m A_n \in \mathcal{L} \). Since \( B_m \uparrow \sum_{n=1}^\infty A_n \) it follows from postulate c. that \( \sum_{n=1}^\infty A_n \in \mathcal{L} \).

Now suppose that \( \mathcal{L} \) satisfies postulates 1. – 3. above. Notice that \( \emptyset \in \mathcal{L} \) and by postulate 3., \( \mathcal{L} \) is closed under finite disjoint unions. Therefore if \( A, B \in \mathcal{L} \) with \( A \subset B \), then \( B^c \in \mathcal{L} \) and \( A \cap B^c = \emptyset \) allows us to conclude that \( A \cup B^c \in \mathcal{L} \). Taking complements of this result shows \( B \setminus A = A^c \cap B \in \mathcal{L} \) as well, i.e. postulate b. holds. If \( A_n \in \mathcal{L} \) with \( A_n \uparrow A \), then \( B_n := A_n \setminus A_{n-1} \in \mathcal{L} \) for all \( n \), where by convention \( A_0 = \emptyset \). Hence it follows by postulate 3 that \( \bigcup_{n=1}^\infty A_n = \sum_{n=1}^\infty B_n \in \mathcal{L} \).

**Theorem E.14 (Dynkin’s \( \pi \) – \( \lambda \) Theorem).** If \( \mathcal{L} \) is a \( \lambda \) class which contains a \( \pi \) – class, \( \mathcal{P} \), then \( \sigma(\mathcal{P}) \subset \mathcal{L} \).

**Proof.** We start by proving the following assertion; for any element \( C \in \mathcal{L} \), the collection of sets, \( \mathcal{L}^C := \{ D \in \mathcal{L} : C \cap D \in \mathcal{L} \} \), is a \( \lambda \) – system. To prove this claim, observe that:

1. \( \Omega \in \mathcal{L} \) and \( \mathcal{L} \subset 2^\Omega \), \( \mathcal{L} = 1 \).
2. If \( A, B \in \mathcal{L} \) with \( A, B \in \mathcal{L} \), then \( A \cap C, B \cap C \in \mathcal{L} \) with \( A \cap C \subset B \cap C \) and therefore,

\[
(B \setminus A) \cap C = [B \cap C] \setminus A = [B \cap C] \setminus [A \cap C] \in \mathcal{L}.
\]

This shows that \( \mathcal{L}^C \) is closed under proper differences. c. If \( A_n \in \mathcal{L}^C \) with \( A_n \uparrow A \), then \( A_n \cap C \in \mathcal{L} \) and \( A_n \cap C \uparrow A \cap C \in \mathcal{L} \), i.e. \( A \in \mathcal{L}^C \). Hence we have verified \( \mathcal{L}^C \) is still a \( \lambda \) – system.

For the rest of the proof, we may assume without loss of generality that \( \mathcal{L} \) is the smallest \( \lambda \) – class containing \( \mathcal{P} \) – if not just replace \( \mathcal{L} \) by the intersection of all \( \lambda \) – classes containing \( \mathcal{P} \). Then for \( C \in \mathcal{P} \) we know that \( \mathcal{L}^C \subset \mathcal{L} \) is a \( \lambda \) – class containing \( \mathcal{P} \) and hence \( \mathcal{L}^C = \mathcal{L} \). Since \( C \in \mathcal{P} \) was arbitrary, we have shown, \( C \cap D \in \mathcal{L} \) for all \( C \in \mathcal{P} \) and \( D \in \mathcal{L} \). We may now conclude that if \( C \in \mathcal{L} \), then \( \mathcal{P} \subset \mathcal{L}^C \subset \mathcal{L} \) and hence again \( \mathcal{L}^C = \mathcal{L} \). Since \( C \in \mathcal{L} \) is arbitrary, we have shown \( C \cap D \in \mathcal{L} \) for all \( C \in \mathcal{L} \), i.e. \( \mathcal{L} \) is a \( \pi \) – system. So by Remark E.12 \( \mathcal{L} \) is a \( \sigma \) algebra. Since \( \sigma(\mathcal{P}) \) is the smallest \( \sigma \) – algebra containing \( \mathcal{P} \) it follows that \( \sigma(\mathcal{P}) \subset \mathcal{L} \).

As an immediate corollary, we have the following uniqueness result.

**Proposition E.15.** Suppose that \( \mathcal{P} \subset 2^\Omega \) is a \( \pi \) – system. If \( P \) and \( Q \) are two probability measures on \( \sigma(\mathcal{P}) \) such that \( P = Q \) on \( \mathcal{P} \), then \( P = Q \) on \( \sigma(\mathcal{P}) \).

**Proof.** Let \( \mathcal{L} := \{ A \in \sigma(\mathcal{P}) : P(A) = Q(A) \} \). One easily shows \( \mathcal{L} \) is a \( \lambda \) – class which contains \( \mathcal{P} \) by assumption. Indeed, \( \Omega \in \mathcal{P} \subset \mathcal{L} \), if \( A, B \in \mathcal{L} \) with \( A \subset B \), then

\[
P(B \setminus A) = P(B) - P(A) = Q(B) - Q(A) = Q(B \setminus A)
\]

More generally, \( P \) and \( Q \) could be two measures such that \( P(\Omega) = Q(\Omega) < \infty \).
so that $B \setminus A \in \mathcal{L}$, and if $A_n \in \mathcal{L}$ with $A_n \uparrow A$, then $P(A) = \lim_{n \to \infty} P(A_n) = \lim_{n \to \infty} Q(A_n) = Q(A)$ which shows $A \in \mathcal{L}$. Therefore $\sigma(\mathcal{P}) \subseteq \mathcal{L} = \sigma(\mathcal{P})$ and the proof is complete. \hfill \blacksquare

Example E.16. Let $\Omega := \{a, b, c, d\}$ and let $\mu$ and $\nu$ be the probability measure on $2^\Omega$ determined by, $\mu \{\{x\}\} = \frac{1}{4}$ for all $x \in \Omega$ and $\nu \{\{a\}\} = \nu \{\{d\}\} = \frac{1}{8}$ and $\nu \{\{b\}\} = \nu \{\{c\}\} = 3/8$. In this example,

$$\mathcal{L} := \{A \in 2^\Omega : P(A) = Q(A)\}$$

is $\lambda$ - system which is not an algebra. Indeed, $A = \{a, b\}$ and $B = \{a, c\}$ are in $\mathcal{L}$ but $A \cap B \notin \mathcal{L}$.

Exercise E.1. Suppose that $\mu$ and $\nu$ are two measures (not assumed to be finite) on a measure space, $(\Omega, \mathcal{P})$ such that $\mu = \nu$ on a $\pi$ - system, $\mathcal{P}$. Further assume $\mathcal{B} = \sigma(\mathcal{P})$ and there exists $\Omega_n \in \mathcal{P}$ such that: i) $\mu(\Omega_n) = \nu(\Omega_n) < \infty$ for all $n$ and ii) $\Omega_n \uparrow \Omega$ as $n \uparrow \infty$. Show $\mu = \nu$ on $\mathcal{B}$.

**Hint:** Consider the measures, $\mu_n(A) := \mu(A \cap \Omega_n)$ and $\nu_n(A) = \nu(A \cap \Omega_n)$.

Corollary E.17. A probability measure, $P$, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is uniquely determined by its cumulative distribution function,

$$F(x) := P((-\infty, x]).$$

**Proof.** This follows from Proposition E.15 wherein we use the fact that $\mathcal{P} : = \{(-\infty, x] : x \in \mathbb{R}\}$ is a $\pi$ - system such that $\mathcal{B}_{\mathbb{R}} = \sigma(\mathcal{P})$. \hfill \blacksquare

Remark E.18. Corollary E.17 generalizes to $\mathbb{R}^n$. Namely a probability measure, $P$, on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ is uniquely determined by its CDF,

$$F(x) := P((-\infty, x]^n)$$

where now

$$(-\infty, x]^n := (-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n].$$

E.2.1 A Density Result*

Exercise E.2 (Density of $A$ in $\sigma(A)$). Suppose that $A \subset 2^\Omega$ is an algebra, $\mathcal{B} := \sigma(A)$, and $P$ is a probability measure on $\mathcal{B}$. Let $\rho(A, B) := P(A \Delta B)$. The goal of this exercise is to use the $\pi - \lambda$ theorem to show that $A$ is dense in $\mathcal{B}$ relative to the “metric,” $\rho$. More precisely you are to show using the following outline that for every $B \in \mathcal{B}$ there exists $A \in \mathcal{A}$ such that that $P(A \Delta B) < \varepsilon$.

1. Recall from Exercise ?? that $\rho(a, B) = P(A \Delta B) = E|1_A - 1_B|$.

2. Observe; if $B = \cup_i B_i$ and $A = \cup_i A_i$, then

$$B \setminus A = \cup_i [B_i \setminus A] \subset \cup_i (B_i \setminus A_i) \subset \cup_i A_i \Delta B_i$$

and

$$A \setminus B = \cup_i [A_i \setminus B] \subset \cup_i (A_i \setminus B_i) \subset \cup_i A_i \Delta B_i$$

so that

$$A \Delta B \subset \cup_i (A_i \Delta B_i).$$

3. We also have

$$(B_2 \setminus B_1) \setminus (A_2 \setminus A_1) = B_2 \cap B_1^c \cap (A_2 \setminus A_1)^c$$

$$= B_2 \cap B_1^c \cap (A_2 \cap A_1^c)^c$$

$$= B_2 \cap B_1^c \cap (B_2 \cup A_1)$$

$$= [B_2 \cap B_1^c \cap A_2^c] \cup [B_2 \cap B_1^c \cap A_1]$$

$$\subset (B_2 \setminus A_2) \cup (A_1 \setminus B_1)$$

and similarly,

$$(A_2 \setminus A_1) \setminus (B_2 \setminus B_1) \subset (A_2 \setminus A_2) \cup (B_1 \setminus A_1)$$

so that

$$(A_2 \setminus A_1) \Delta (B_2 \setminus B_1) \subset (A_2 \setminus A_2) \cup (A_1 \setminus B_1) \cup (A_2 \setminus B_2) \cup (B_1 \setminus A_1)$$

$$= (A_1 \Delta B_1) \cup (A_2 \Delta B_2).$$

4. Observe that $A_n \in \mathcal{B}$ and $A_n \uparrow A$, then

$$P(B \Delta A_n) = P(B \setminus A_n) + P(A_n \setminus B)$$

$$\rightarrow P(A \setminus B) + P(B \setminus A) = P(A \Delta B).$$

5. Let $\mathcal{L}$ be the collection of sets $B \in \mathcal{B}$ for which the assertion of the theorem holds. Show $\mathcal{L}$ is a $\lambda$ - system which contains $\mathcal{A}$.

E.3 Construction of Measures

**Definition E.19.** Given a collection of subsets, $\mathcal{E}$, of $\Omega$, let $\mathcal{E}_{\varepsilon}$ denote the collection of subsets of $\Omega$ which are finite or countable unions of sets from $\mathcal{E}$. Similarly let $\mathcal{E}_{\delta}$ denote the collection of subsets of $\Omega$ which are finite or countable intersections of sets from $\mathcal{E}$. We also write $\mathcal{E}_{\varepsilon\delta} = (\mathcal{E}_{\varepsilon})_{\delta}$ and $\mathcal{E}_{\delta\varepsilon} = (\mathcal{E}_{\delta})_{\varepsilon}$, etc.

**Lemma E.20.** Suppose that $A \subset 2^\Omega$ is an algebra. Then:
1. $A_\sigma$ is closed under taking countable unions and finite intersections.
2. $A_\delta$ is closed under taking countable intersections and finite unions.
3. $\{A^c : A \in A_\sigma\} = A_\delta$ and $\{A^c : A \in A_\delta\} = A_\sigma$.

**Proof.** By construction $A_\sigma$ is closed under countable unions. Moreover if $A = \bigcup_{i=1}^\infty A_i$ and $B = \bigcup_{j=1}^\infty B_j$ with $A_i, B_j \in A$, then

$$A \cap B = \bigcup_{i,j=1}^\infty A_i \cap B_j \in A_\sigma,$$

which shows that $A_\sigma$ is also closed under finite intersections. Item 3. is straight forward and item 2. follows from items 1. and 3.

**Remark E.21.** Let us recall from Proposition E.3 and Remark E.4 that a finitely additive measure $\mu : A \to [0, \infty]$ is a premeasure on $A$ if $\mu(\bigcup_{n=1}^\infty A_n)$ for all $\{A_n\}_{n=1}^\infty \subset A$ such that $A_n \uparrow A \in A$. Furthermore if $\mu(\Omega) < \infty$, then $\mu$ is a premeasure on $A$ if $\mu(A_n) \downarrow 0$ for all $\{A_n\}_{n=1}^\infty \subset A$ such that $A_n \downarrow 0$.

**Proposition E.22.** Given a premeasure, $\mu : A \to [0, \infty]$, we extend $\mu$ to $A_\sigma$ by defining

$$\mu(B) := \sup \{\mu(A) : A \ni B \in A\}. \quad (E.2)$$

This function $\mu : A_\sigma \to [0, \infty]$ then satisfies;

1. **(Monotonicity)** If $A, B \in A_\sigma$ with $A \subset B$ then $\mu(A) \leq \mu(B)$.
2. **(Continuity)** If $A_n \in A$ and $A_n \uparrow B \in A_\sigma$, then $\mu(A_n) \uparrow \mu(B)$ as $n \to \infty$.
3. **(Strong Additivity)** If $A, B \in A_\sigma$, then

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B). \quad (E.3)$$

4. **(Sub-Additivity on $A_\sigma$)** The function $\mu$ is sub-additive on $A_\sigma$, i.e. if $\{A_n\}_{n=1}^\infty \subset A_\sigma$, then

$$\mu(\bigcup_{n=1}^\infty A_n) \leq \sum_{n=1}^\infty \mu(A_n). \quad (E.4)$$

5. **(\(\sigma\)-Additivity on $A_\sigma$)** The function $\mu$ is countably additive on $A_\sigma$.

**Proof.** 1. and 2. Monotonicity follows directly from Eq. (E.2) which then implies $\mu(A_n) \leq \mu(B)$ for all $n$. Therefore $M := \lim_{n \to \infty} \mu(A_n) \leq \mu(B)$. To prove the reverse inequality, let $A \ni A \subset B$. Then by the continuity of $\mu$ on $A$ and the fact that $A_n \cap A \uparrow B \cap A = A$, we have $\mu(A_n \cap A) \uparrow \mu(A)$. As $\mu(A_n) \geq \mu(A_n \cap A)$ for all $n$ it then follows that $M := \lim_{n \to \infty} \mu(A_n) \geq \mu(A)$. As $A \in A$ with $A \subset B$ was arbitrary we may conclude,

$$\mu(B) = \sup \{\mu(A) : A \ni A \subset B\} \leq M.$$ 

3. Suppose that $A, B \in A_\sigma$ and $\{A_n\}_{n=1}^\infty$ and $\{B_n\}_{n=1}^\infty$ are sequences in $A$ such that $A_n \uparrow A$ and $B_n \uparrow B$ as $n \to \infty$. Then passing to the limit as $n \to \infty$ in the identity,

$$\mu(A_n \cup B_n) + \mu(A_n \cap B_n) = \mu(A_n) + \mu(B_n)$$

proves Eq. (E.3). In particular, it follows that $\mu$ is finitely additive on $A_\sigma$.

4 and 5. Let $\{A_n\}_{n=1}^\infty$ be any sequence in $A_\sigma$ and choose $\{A_{n,i}\}_{i=1}^\infty \subset A$ such that $A_{n,i} \uparrow A_n$ as $i \to \infty$. Then we have,

$$\mu(\bigcup_{n=1}^N A_{n,N}) \leq \sum_{n=1}^N \mu(A_{n,N}) \leq \sum_{n=1}^N \mu(A_n) \leq \sum_{n=1}^\infty \mu(A_n). \quad (E.5)$$

Since $A \ni \bigcup_{n=1}^N A_{n,N} \uparrow \bigcup_{n=1}^\infty A_n \in A_\sigma$, we may let $N \to \infty$ in Eq. (E.5) to conclude Eq. (E.4) holds. If we further assume that $\{A_n\}_{n=1}^\infty \subset A_\sigma$ are pairwise disjoint, by the finite additivity and monotonicity of $\mu$ on $A_\sigma$, we have

$$\sum_{n=1}^\infty \mu(A_n) = \lim_{N \to \infty} \sum_{n=1}^N \mu(A_n) = \lim_{N \to \infty} \mu(\bigcup_{n=1}^N A_{n,N}) \leq \mu(\bigcup_{n=1}^\infty A_n).$$

This inequality along with Eq. (E.4) shows that $\mu$ is $\sigma$ – additive on $A_\sigma$.

**Lemma E.23.** Suppose $\mu$ is a finite premeasure on an algebra, $A \subset 2^\Omega$, and $A \in A_3 \cap A_\sigma$. Since $A, A^c \in A_\sigma$ and $\Omega = A \cup A^c$, it follows that $\mu(\Omega) = \mu(A) + \mu(A^c)$. From this observation we may extend $\mu$ to a function on $A_3 \cup A_\sigma$ by defining

$$\mu(A) := \mu(\Omega) - \mu(A^c) \text{ for all } A \in A_3. \quad (E.6)$$

**Proof.** 1. If $A \in A_3$ then $\mu(A) = \inf \{\mu(B) : A \subset B \in A\}$.
2. If $A \in A_3$ and $A_n \in A$ such that $A_n \downarrow A$, then $\mu(A) = \lim_{n \to \infty} \mu(A_n)$.
3. $\mu$ is strongly additive when restricted to $A_3$.
4. If $A \in A_3$ and $C \in A_\sigma$ such that $A \subset C$, then $\mu(C \setminus A) = \mu(C) - \mu(A)$.

**Proof.**

1. Since $\mu(B) = \mu(\Omega) - \mu(B^c)$ and $A \subset B$ iff $B^c \subset A^c$, it follows that

$$\inf \{\mu(B) : A \subset B \in A\} = \inf \{\mu(\Omega) - \mu(B^c) : A \ni B^c \subset A^c\} = \mu(\Omega) - \sup \{\mu(B) : A \ni B \subset A^c\} = \mu(\Omega) - \mu(A^c) = \mu(A).$$
2. Similarly, since \( A^c_n \uparrow A^c \in \mathcal{A}_\sigma \), by the definition of \( \mu(A) \) and Proposition \[E.22\] it follows that
\[
\mu(A) = \mu(\Omega) - \mu(A^c) = \mu(\Omega) - \lim_{n \to \infty} \mu(A^c_n) = \lim_{n \to \infty} [\mu(\Omega) - \mu(A^c_n)] = \lim_{n \to \infty} \mu(A_n) .
\]

3. Suppose \( A, B \in \mathcal{A}_\delta \) and \( A_n, B_n \in \mathcal{A} \) such that \( A_n \downarrow A \) and \( B_n \downarrow B \), then \( A_n \cup B_n \uparrow A \cup B \) and \( A_n \cap B_n \downarrow A \cap B \) and therefore,
\[
\mu(A \cup B) + \mu(A \cap B) = \lim_{n \to \infty} [\mu(A_n \cup B_n) + \mu(A_n \cap B_n)] = \lim_{n \to \infty} [\mu(A_n) + \mu(B_n)] = \mu(A) + \mu(B).
\]

All we really need is the finite additivity of \( \mu \) which can be proved as follows. Suppose that \( A, B \in \mathcal{A}_\delta \) are disjoint, then \( A \cap B = \emptyset \) implies \( A^c \cup B^c = \Omega \). So by the strong additivity of \( \mu \) on \( \mathcal{A}_\delta \) it follows that
\[
\mu(\Omega) + \mu(A^c \cap B^c) = \mu(A^c) + \mu(B^c)
\]
from which it follows that
\[
\mu(A \cup B) = \mu(\Omega) - \mu(A^c \cap B^c) = \mu(\Omega) - (\mu(A^c) + \mu(B^c) - \mu(\Omega)) = \mu(A) + \mu(B).
\]

4. Since \( A^c, C \in \mathcal{A}_\sigma \) we may use the strong additivity of \( \mu \) on \( \mathcal{A}_\sigma \) to conclude,
\[
\mu(A^c \cup C) + \mu(A^c \cap C) = \mu(A^c) + \mu(C).
\]
Because \( \Omega = A^c \cup C \), and \( \mu(A^c) = \mu(\Omega) - \mu(A) \), the above equation may be written as
\[
\mu(\Omega) + \mu(C \setminus A) = \mu(\Omega) - \mu(A) + \mu(C)
\]
which finishes the proof.

\[ \square \]

**Notation E.24 (Inner and outer measures)** Let \( \mu : \mathcal{A} \to [0, \infty) \) be a finite premeasure extended to \( \mathcal{A}_\sigma \cup \mathcal{A}_\delta \) as above. The for any \( B \subset \Omega \) let
\[
\mu_+(B) := \sup \{ \mu(A) : \mathcal{A}_\delta \ni A \subset B \} \quad \text{and} \quad \mu^+(B) := \inf \{ \mu(C) : B \subset C \in \mathcal{A}_\sigma \}.
\]
We refer to \( \mu_+(B) \) and \( \mu^+(B) \) as the **inner and outer content** of \( B \) respectively.

If \( B \subset \Omega \) has the same inner and outer content it is reasonable to define the measure of \( B \) as this common value. As we will see in Theorem \[E.27\] below, this extension becomes a \( \sigma \) – additive measure on a \( \sigma \) – algebra of subsets of \( \Omega \).

**Definition E.25 (Measurable Sets)**. Suppose \( \mu \) is a finite premeasure on an algebra \( \mathcal{A} \subset 2^\Omega \). We say that \( B \subset \Omega \) is **measurable** if \( \mu_+(B) = \mu^+(B) \). We will denote the collection of measurable subsets of \( \Omega \) by \( \mathcal{B} = \mathcal{B}(\mu) \) and define \( \bar{\mu} : \mathcal{B} \to [0, \mu(\Omega)] \) by
\[
\bar{\mu}(B) := \mu_+(B) = \mu^+(B) \quad \text{for all} \quad B \in \mathcal{B} .
\]

**Remark E.26**. Observe that \( \mu_+(B) = \mu^+(B) \) iff for all \( \varepsilon > 0 \) there exists \( A \in \mathcal{A}_\delta \) and \( C \in \mathcal{A}_\sigma \) such that \( A \subset B \subset C \) and
\[
\mu(C \setminus A) = \mu(C) - \mu(A) < \varepsilon ,
\]
wherein we have used Lemma \[E.23\] for the first equality. Moreover we will use below that if \( B \in \mathcal{B} \) and \( \mathcal{A}_\delta \ni A \subset B \subset C \in \mathcal{A}_\sigma \), then
\[
\mu(A) \leq \mu_+(B) = \bar{\mu}(B) = \mu^+(B) \leq \mu(C) .
\]

**Theorem E.27 (Finite Premeasure Extension Theorem)**. Suppose \( \mu \) is a finite premeasure on an algebra \( \mathcal{A} \subset 2^\Omega \) and \( \bar{\mu} : \mathcal{B} := \mathcal{B}(\mu) \to [0, \mu(\Omega)] \) be as in Definition \[E.25\]. Then \( \mathcal{B} \) is a \( \sigma \) – algebra on \( \Omega \) which contains \( \mathcal{A} \) and \( \bar{\mu} \) is a \( \sigma \) – additive measure on \( \mathcal{B} \). Moreover, \( \bar{\mu} \) is the unique measure on \( \mathcal{B} \) such that \( \bar{\mu}|_{\mathcal{A}} = \mu \).

**Proof.** 1. \( \mathcal{B} \) is an algebra. It is clear that \( \mathcal{A} \subset \mathcal{B} \) and that \( \mathcal{B} \) is closed under complementation – see Eq. \[E.8\] and use the fact that \( A^c \cap C^c = C \setminus A \). Now suppose that \( B_i \in \mathcal{B} \) for \( i = 1, 2 \) and \( \varepsilon > 0 \) is given. We may then choose \( A_i \subset B_i \subset C_i \) such that \( A_i \in \mathcal{A}_\delta \), \( C_i \in \mathcal{A}_\sigma \), and \( \mu(C_i \setminus A_i) < \varepsilon \) for \( i = 1, 2 \). Then with \( A = A_1 \cup A_2 \), \( B = B_1 \cup B_2 \) and \( C = C_1 \cup C_2 \), we have \( \mathcal{A}_\delta \ni A \subset B \subset C \in \mathcal{A}_\sigma \). Since
\[
C \setminus A = (C_1 \setminus A) \cup (C_2 \setminus A) \subset (C_1 \setminus A_1) \cup (C_2 \setminus A_2) ,
\]
it follows from the sub-additivity of \( \mu \) that with
\[
\mu(C \setminus A) \leq \mu(C_1 \setminus A_1) + \mu(C_2 \setminus A_2) < 2 \varepsilon .
\]

Since \( \varepsilon > 0 \) was arbitrary, we have shown that \( B \in \mathcal{B} \) which completes the proof that \( \mathcal{B} \) is an algebra.

2. \( \mathcal{B} \) is a \( \sigma \) – algebra. As we know \( \mathcal{B} \) is an algebra, to show \( \mathcal{B} \) is a \( \sigma \) – algebra it suffices to show that \( B = \sum_{n=1}^{\infty} B_n \in \mathcal{B} \) whenever \( \{B_n\}_{n=1}^{\infty} \) is a disjoint sequence in \( \mathcal{B} \). To this end, let \( \varepsilon > 0 \) be given and choose \( A_i \subset B_i \subset C_i \) such
that \(A_i \in \mathcal{A}_3, C_i \in \mathcal{A}_{\sigma}\), and \(\mu(C_i \setminus A_i) < \varepsilon 2^{-i}\) for all \(i\). Let \(C := \bigcup_{i=1}^{\infty} C_i \in \mathcal{A}_{\sigma}\) and for \(n \in \mathbb{N}\) let \(A^n := \bigcup_{i=1}^{n} A_i \in \mathcal{A}_3\). Since the \(\{A_i\}_{i=1}^{\infty}\) are pairwise disjoint we may use Lemma E.23 to show,

\[
\sum_{i=1}^{n} \mu(C_i) = \sum_{i=1}^{n} (\mu(A_i) + \mu(C_i \setminus A_i)) = \mu(A^n) + \sum_{i=1}^{n} \mu(C_i \setminus A_i) \leq \mu(\Omega) + \sum_{i=1}^{n} \varepsilon 2^{-i}
\]

which on letting \(n \to \infty\) implies

\[
\sum_{i=1}^{\infty} \mu(C_i) \leq \mu(\Omega) + \varepsilon < \infty. \quad (E.10)
\]

Using

\[
C \setminus A^n = \bigcup_{i=1}^{\infty} (C_i \setminus A^n) \subset [\bigcup_{i=1}^{n} (C_i \setminus A_i)] \cup [\bigcup_{i=n+1}^{\infty} C_i] \in \mathcal{A}_{\sigma},
\]

and the sub-additivity of \(\mu\) on \(\mathcal{A}_{\sigma}\) it follows that

\[
\mu(C \setminus A^n) \leq \sum_{i=1}^{n} \mu(C_i \setminus A_i) + \sum_{i=n+1}^{\infty} \mu(C_i) \leq \varepsilon \sum_{i=1}^{n} 2^{-i} + \sum_{i=n+1}^{\infty} \mu(C_i) \leq \varepsilon + \sum_{i=n+1}^{\infty} \mu(C_i) \to \varepsilon \text{ as } n \to \infty,
\]

wherein we have used Eq. (E.10) in computing the limit. In summary, \(B = \bigcup_{i=1}^{\infty} B_i, \mathcal{A}_3 \supseteq A^n \subset B \subset C \in \mathcal{A}_{\sigma}\), \(C \setminus A^n \in \mathcal{A}_{\sigma}\) with \(\mu(C \setminus A^n) \leq 2\varepsilon\) for all \(n\) sufficiently large. Since \(\varepsilon > 0\) is arbitrary, it follows that \(B \in \mathcal{B}\).

3. \(\bar{\mu}\) is a measure. Continuing the notation in step 2, we have

\[
\sum_{i=1}^{\infty} \mu(A_i) = \sup_{n \to \infty} \sum_{i=1}^{n} \mu(A_i) = \mu(A^n) \leq \bar{\mu}(B) \leq \mu(C) \leq \sum_{i=1}^{\infty} \mu(C_i). \quad (E.11)
\]

On the other hand, since \(A_i \subset B_i \subset C_i\), it follows (see Eq. (E.9)) that \(\mu(A_i) \leq \bar{\mu}(B_i) \leq \mu(C_i)\) and therefore that

\[
\sum_{i=1}^{\infty} \mu(A_i) \leq \sum_{i=1}^{\infty} \bar{\mu}(B_i) \leq \sum_{i=1}^{\infty} \mu(C_i). \quad (E.12)
\]

Equations (E.11) and (E.12) show that \(\bar{\mu}(B)\) and \(\sum_{i=1}^{\infty} \bar{\mu}(B_i)\) are both between \(\sum_{i=1}^{\infty} \mu(A_i)\) and \(\sum_{i=1}^{\infty} \mu(C_i)\) and so

\[
\left| \bar{\mu}(B) - \sum_{i=1}^{\infty} \bar{\mu}(B_i) \right| \leq \sum_{i=1}^{\infty} \mu(C_i) - \sum_{i=1}^{\infty} \mu(A_i) = \sum_{i=1}^{\infty} \mu(C_i \setminus A_i) \leq \sum_{i=1}^{\infty} \varepsilon 2^{-i} = \varepsilon.
\]

Since \(\varepsilon > 0\) is arbitrary, we have shown \(\bar{\mu}(B) = \sum_{i=1}^{\infty} \bar{\mu}(B_i)\), i.e. \(\bar{\mu}\) is a measure on \(\mathcal{B}\).

Since we really had no choice as to how to extend \(\mu\), it is to be expected that the extension is unique. You are asked to supply the details in Exercise E.3 below.

**Exercise E.3.** Let \(\mu, \bar{\mu}, \mathcal{A}, \text{ and } B := B(\mu)\) be as in Theorem E.27. Further suppose that \(B_0 \subset 2^\Omega\) is a \(\sigma\) – algebra such that \(\mathcal{A} \subset B_0 \subset B\) and \(\nu : B_0 \to [0, \mu(\Omega))]\) is a \(\sigma\) – additive measure on \(B_0\) such that \(\nu = \mu|_A\). Show that \(\nu = \bar{\mu}\) on \(B_0\) as well. (When \(B_0 = \sigma(\mathcal{A})\) this exercise is of course a consequence of Proposition E.15. It is not necessary to use this information to complete the exercise.)

**Corollary E.28.** Suppose that \(A \subset 2^\Omega\) is an algebra and \(\mu : B_0 := \sigma(\mathcal{A}) \to [0, \mu(\Omega))\) is a \(\sigma\) – additive finite measure. Then for every \(B \in \sigma(\mathcal{A})\) and \(\varepsilon > 0;\)

1. there exists \(\mathcal{A}_3 \supseteq A \subset B \subset C \in \mathcal{A}_\sigma\) and \(\varepsilon > 0\) such that \(\mu(C \setminus A) < \varepsilon\) and
2. there exists \(A \in \mathcal{A}\) such that \(\mu(A \Delta B) < \varepsilon\).

**Exercise E.4.** Prove corollary E.28 by considering \(\bar{\nu}\) where \(\nu := \mu|_A\). **Hint:** you may find Exercise ?? useful here.

**Theorem E.29 (\(\sigma\) – Finite Premeasure Extension Theorem).** Suppose that \(\mu\) is a \(\sigma\) – finite premeasure on an algebra \(\mathcal{A}\). Then

\[
\bar{\mu}(B) := \inf \{\mu(C) : B \subset C \in \mathcal{A}_\sigma\} \quad \forall B \in \sigma(\mathcal{A}) \quad (E.13)
\]

defines a measure on \(\sigma(\mathcal{A})\) and this measure is the unique extension of \(\mu\) on \(\mathcal{A}\) to a measure on \(\sigma(\mathcal{A})\). Recall that

\[
\mu(C) = \sup \{\mu(A) : A \supseteq C \subset \mathcal{A}_\sigma\} \quad \forall C \in \mathcal{A}_\sigma.
\]

**Proof.** Let \(\{\Omega_n\}_{n=1}^{\infty} \subset \mathcal{A}\) be chosen so that \(\mu(\Omega_n) < \infty\) for all \(n\) and \(\Omega_n \uparrow \Omega\) as \(n \to \infty\) and let

\[
\mu_n(A) := \mu_n(A \cap \Omega_n) \quad \forall A \in \mathcal{A}.
\]

Each \(\mu_n\) is a premeasure (as is easily verified) on \(\mathcal{A}\) and hence by Theorem E.27 each \(\mu_n\) has an extension, \(\bar{\mu}_n\), to a measure on \(\sigma(\mathcal{A})\). Since the measure \(\bar{\mu}_n\) are increasing, \(\bar{\mu} := \lim_{n \to \infty} \bar{\mu}_n\) is a measure which extends \(\mu\).

The proof will be completed by verifying that Eq. (E.13) holds. Let \(B \in \sigma(\mathcal{A})\), \(B_m = \Omega_m \cap B\) and \(\varepsilon > 0\) be given. By Theorem E.27 there exists
Proposition E.31. Suppose that $S \subset 2^\Omega$ is a semi-algebra, $A = A(S)$ and $\mu : A \to [0, \infty]$ is a finitely additive measure. Then $\mu$ is a premeasure on $A$ iff $\mu$ is countably sub-additive on $S$.

Proof. Clearly if $\mu$ is a premeasure on $A$ then $\mu$ is $\sigma$-additive and hence sub-additive on $S$. Because of Proposition 7, to prove the converse it suffices to show that the sub-additivity of $\mu$ on $S$ implies the sub-additivity of $\mu$ on $A$.

So suppose $A = \sum_{n=1}^{\infty} A_n \in A$ with each $A_n \in A$. By Proposition 7 we may write $A = \sum_{j=1}^{k} E_j$ and $A_n = \sum_{i=1}^{N_n} E_{n,i}$ with $E_j, E_{n,i} \in S$. Intersecting the identity, $A = \sum_{n=1}^{\infty} A_n$, with $E_j$ implies

$$E_j = A \cap E_j = \sum_{n=1}^{\infty} A_n \cap E_j = \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} E_{n,i} \cap E_j.$$

By the assumed sub-additivity of $\mu$ on $S$,

$$\mu(E_j) \leq \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \mu(E_{n,i} \cap E_j).$$

Summing this equation on $j$ and using the finite additivity of $\mu$ shows

$$\mu(A) \leq \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \mu(E_{n,i} \cap E_j) = \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \mu(E_{n,i}) = \sum_{n=1}^{\infty} \mu(A_n).$$

Suppose now that $\mu$ is a Radon measure on $(\mathbb{R}, B_{\mathbb{R}})$ and $F : \mathbb{R} \to \mathbb{R}$ is chosen so that

$$\mu((a, b]) = F(b) - F(a) \quad \text{for all } -\infty < a < b < \infty. \quad (E.14)$$

For example if $\mu(\mathbb{R}) < \infty$ we can take $F(x) = \mu(\{-\infty, x]\}$ while if $\mu(\mathbb{R}) = \infty$ we might take

$$F(x) = \begin{cases} \mu((0, x]) & \text{if } x \geq 0 \\ -\mu((x, 0]) & \text{if } x \leq 0 \end{cases}.$$

The function $F$ is uniquely determined modulo translation by a constant.

Lemma E.32. If $\mu$ is a Radon measure on $(\mathbb{R}, B_{\mathbb{R}})$ and $F : \mathbb{R} \to \mathbb{R}$ is chosen so that $\mu((a, b]) = F(b) - F(a)$, then $F$ is increasing and right continuous.

Proof. The function $F$ is increasing by the monotonicity of $\mu$. To see that $F$ is right continuous, let $b \in \mathbb{R}$ and choose $a \in (\infty, b)$ and any sequence $\{b_n\}_{n=1}^{\infty} \subset (b, \infty)$ such that $b_n \downarrow b$ as $n \to \infty$. Since $\mu((a, b_n]) < \infty$ and $\{a, b_n\} \downarrow (a, b)$ as $n \to \infty$, it follows that

$$F(b_n) - F(a) = \mu((a, b_n]) \downarrow \mu((a, b]) = F(b) - F(a).$$
Since \( \{b_n\}_{n=1}^\infty \) was an arbitrary sequence such that \( b_n \downarrow b \), we have shown 

\[
\lim_{y \downarrow b} F(y) = F(b).
\]

The key result of this section is the converse to this lemma.

**Theorem E.33.** Suppose \( F : \mathbb{R} \to \mathbb{R} \) is a right continuous increasing function. Then there exists a unique Radon measure, \( \mu = \mu_F \), on \( (\mathbb{R}, \mathcal{B}_{\mathbb{R}}) \) such that Eq. (E.14) holds.

**Proof.** Let \( S := \{[a, b] \cap \mathbb{R} : -\infty \leq a \leq b \leq \infty\} \), and \( A = A(S) \) consists of those sets, \( A \subset \mathbb{R} \) which may be written as finite disjoint unions of sets from \( S \) as in Example 11.1. Recall that \( \mathcal{B}_{\mathbb{R}} = \sigma(A) = \sigma(S) \). Further define \( F(\pm \infty) := \lim_{x \to \pm \infty} F(x) \) and let \( \mu = \mu_F \) be the finitely additive measure on \( (\mathbb{R}, A) \) described in Proposition ?? and Remark ???. To finish the proof it suffices by Theorem [E.29] to show that \( \mu \) is a premeasure on \( A = A(S) \) where

\[
S := \{[a, b] \cap \mathbb{R} : -\infty \leq a < b \leq \infty\}.
\]

So in light of Proposition E.31, to finish the proof it suffices to show \( \mu \) is subadditive on \( S \), i.e. we must show

\[
\mu(J) \leq \sum_{n=1}^\infty \mu(J_n).
\]

(E.15)

where \( J = \sum_{n=1}^\infty J_n \) with \( J = [a, b] \cap \mathbb{R} \) and \( J_n = (a_n, b_n] \cap \mathbb{R} \). Recall from Proposition ?? that the finite additivity of \( \mu \) implies

\[
\sum_{n=1}^\infty \mu(J_n) \leq \mu(J).
\]

(E.16)

We begin with the special case where \( -\infty < a < b < \infty \). Our proof will be by “continuous induction.” The strategy is to show \( a \in A \) where

\[
A := \left\{ \alpha \in [a, b] : \mu(J \cap (\alpha, b]) \leq \sum_{n=1}^\infty \mu(J_n \cap (\alpha, b]) \right\}.
\]

(E.17)

As \( b \in J \), there exists an \( k \) such that \( b \in J_k \) and hence \( (a_k, b_k] = (a, b] \) for this \( k \). It now easily follows that \( J_k \subset A \) so that \( A \) is not empty. To finish the proof we are going to show \( \bar{a} := \inf A \in A \) and that \( \bar{a} = a \).

- Let \( \alpha_m \in A \) such that \( \alpha_m \downarrow \bar{a} \), i.e.

\[
\mu(J \cap (\alpha_m, b]) \leq \sum_{n=1}^\infty \mu(J_n \cap (\alpha_m, b]).
\]

(E.18)

The right continuity of \( F \) implies \( \alpha \to \mu(J \cap (\alpha, b]) \) is right continuous. So by the dominated convergence theorem for sums,\(^2\) DCT applies as \( \mu(J \cap (\alpha_m, b]) \leq \mu(J_n) \) and \( \sum_{n=1}^\infty \mu(J_n) \leq \mu(J) < \infty \) by Eq. (E.16).

\[
\mu(J \cap (\bar{a}, b]) = \lim_{m \to \infty} \mu(J \cap (\alpha_m, b])
\]

\[
\leq \lim_{m \to \infty} \sum_{n=1}^\infty \mu(J_n \cap (\alpha_m, b])
\]

\[
= \sum_{n=1}^\infty \lim_{m \to \infty} \mu(J_n \cap (\alpha_m, b]) = \sum_{n=1}^\infty \mu(J_n \cap (\bar{a}, b]),
\]

i.e. \( \bar{a} \in A \).

- If \( \bar{a} > a \), then \( \bar{a} \in J_l = (a_l, b_l] \) for some \( l \). Letting \( \alpha = a_l < \bar{a} \), we have,

\[
\mu(J \cap (\alpha, b]) = \mu(J \cap (\alpha, \bar{a}]) + \mu(J \cap (\bar{a}, b])
\]

\[
\leq \mu(J_l \cap (\alpha, \bar{a}]) + \mu(J_l \cap (\bar{a}, b])
\]

\[
= \mu(J_l \cap (\alpha, \bar{a}]) + \mu(J_l \cap (\bar{a}, b]) + \mu(J_n \cap (\bar{a}, b])
\]

\[
= \mu(J_l \cap (\alpha, b]) + \sum_{n \neq l} \mu(J_n \cap (\bar{a}, b])
\]

\[
\leq \sum_{n=1}^\infty \mu(J_n \cap (\alpha, b])
\]

This shows \( \alpha \in A \) and \( \alpha < \bar{a} \) which violates the definition of \( \bar{a} \). Thus we must conclude that \( \bar{a} = a \).

The hard work is now done but we still have to check the cases where \( a = -\infty \) or \( b = \infty \). For example, suppose that \( b = \infty \) so that

\[
J = (a, \infty) = \sum_{n=1}^\infty J_n
\]

with \( J_n = (a_n, b_n] \cap \mathbb{R} \). Then

\[
I_M := (a, \infty] = J \cap I_M = \sum_{n=1}^\infty J_n \cap I_M
\]

and so by what we have already proved,

\[
F(M) - F(a) = \mu(I_M) \leq \sum_{n=1}^\infty \mu(J_n \cap I_M) \leq \sum_{n=1}^\infty \mu(J_n).
\]

Now let \( M \to \infty \) in this last inequality to find that
The other cases where \( a = -\infty \) and \( b \in \mathbb{R} \) and \( a = -\infty \) and \( b = \infty \) are handled similarly.

**Proof.** Old Proof of Theorem [E.33] Case 1. First suppose that \( -\infty < a < b < \infty \). Choose numbers \( \tilde{a} > a \), \( \tilde{b} > b \) in which case \( I := (\tilde{a}, \tilde{b}) \subset J \)
\[
\tilde{J}_n := (a_n, \tilde{b}) \supset \tilde{J}_n^\ast := (a_n, \tilde{b}) \supset J_n.
\]
Since \( \tilde{I} = [\tilde{a}, \tilde{b}] \) is compact and \( \tilde{I} \subset J \subset \bigcup_{n=1}^\infty \tilde{J}_n \) there exists \( N < \infty \) such that
\[
\tilde{I} \subset \tilde{I} \subset \bigcup_{n=1}^N \tilde{J}_n \subset \bigcup_{n=1}^N \tilde{J}_n.
\]
Hence by finite sub-additivity of \( \mu \),
\[
F(\tilde{b}) - F(\tilde{a}) = \mu(\tilde{I}) \leq \sum_{n=1}^N \mu(\tilde{J}_n) \leq \sum_{n=1}^\infty \mu(\tilde{J}_n).
\]
Using the right continuity of \( F \) and letting \( \tilde{a} \downarrow a \) in the above inequality,
\[
\mu(J) = \mu((a, b]) = F(b) - F(a) \leq \sum_{n=1}^\infty \mu(\tilde{J}_n) = \sum_{n=1}^\infty \mu(J_n) + \sum_{n=1}^\infty \mu(J_n \setminus J_{n+1}).
\]
Given \( \varepsilon > 0 \), we may use the right continuity of \( F \) to choose \( \tilde{b}_n \) so that
\[
\mu(J_n \setminus J_{n+1}) = F(\tilde{b}_n) - F(b_n) \leq \varepsilon 2^{-n} \quad \forall \ n \in \mathbb{N}.
\]
Using this in Eq. (E.19) shows
\[
\mu(J) = \mu((a, b]) \leq \sum_{n=1}^\infty \mu(J_n) + \varepsilon
\]
which verifies Eq. (E.15) since \( \varepsilon > 0 \) was arbitrary. \( \blacksquare \)

---

Example E.34 (The construction of all probability measures on \((\mathbb{R}, B_\mathbb{R})\)). Suppose that \( \Omega = \mathbb{R} \), \( \mathcal{S} := \{(a, b] \cap \mathbb{R} : -\infty \leq a \leq b \leq \infty \} \), and \( \mathcal{A} = \mathcal{A}(\mathcal{S}) \) consists of those sets, \( A \subset \mathbb{R} \) which may be written as finite disjoint unions of sets from \( \mathcal{S} \) as in Example [11.13] Recall that \( B_\mathbb{R} = \sigma(\mathcal{A}) = \sigma(\mathcal{S}) \).

1. By Proposition 11.10, to every increasing function, \( F : \mathbb{R} \to [0, 1] \) such that
\[
F(-\infty) := \lim_{x \to -\infty} F(x) = 0 \quad \text{and} \quad F(+\infty) := \lim_{x \to +\infty} F(x) = 1
\]
there exists a finitely additive probability measure, \( P = P_F \) on \( \mathcal{A} \) such that
\[
P((a, b] \cap \mathbb{R}) = F(b) - F(a) \quad \text{for all} \quad -\infty < a \leq b \leq \infty.
\]
2. By Lemma E.8 and Theorem E.33 the measure \( P \) is \( \sigma \) – additive on \( \mathcal{A} \) iff \( F \) is right continuous.
3. Now apply Theorem E.27 to learn that \( P \) extends to a probability measure on \( B_\mathbb{R} \) iff \( F \) is right continuous.

---

E.4.1 Lebesgue Measure

If \( F(x) = x \) for all \( x \in \mathbb{R} \), we denote \( \mu_F \) by \( m \) and call \( m \) Lebesgue measure on \((\mathbb{R}, B_\mathbb{R})\).

**Theorem E.35.** Lebesgue measure \( m \) is invariant under translations, i.e. for \( B \in B_\mathbb{R} \) and \( x \in \mathbb{R} \),
\[
m(x + B) = m(B).
\]
Lebesgue measure, \( m \), is the unique measure on \( B_\mathbb{R} \) such that \( m((0, 1]) = 1 \) and Eq. (E.20) holds for \( B \in B_\mathbb{R} \) and \( x \in \mathbb{R} \). Moreover, \( m \) has the scaling property
\[
m(\lambda B) = |\lambda| m(B)
\]
where \( \lambda \in \mathbb{R} \), \( B \in B_\mathbb{R} \) 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 \( B_\mathbb{R} \) such that \( m_x((a, b]) = b - a \) for all \( a < b \). Therefore, \( m_x = m \) by the uniqueness assertion in Exercise 11.10. For the converse, suppose that \( m \) is translation invariant and \( m((0, 1]) = 1 \). Given \( n \in \mathbb{N} \), we have
\[
(0, 1] = \bigcup_{k=1}^n \left( \frac{k-1}{n}, \frac{k}{n} \right] = \bigcup_{k=1}^n \left( \frac{k-1}{n} + (0, \frac{1}{n}) \right).
\]
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
\[ m((0, \frac{1}{n}]) = \frac{1}{n}. \]

Similarly, \( m((0, \frac{1}{n}]) = \frac{l}{n} \) 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. (E.21) 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 \( m_\lambda \) is again a measure on \( B_\mathbb{R} \) which satisfies
\[ m_\lambda((a,b]) = \lambda^{-1} m((\lambda a, \lambda b]) = \lambda^{-1} (\lambda b - \lambda a) = b - a \]
if \( \lambda > 0 \) and
\[ m_\lambda((a,b]) = |\lambda|^{-1} m((\lambda b, \lambda a]) = -|\lambda|^{-1} (\lambda b - \lambda a) = b - a \]
if \( \lambda < 0 \). Hence \( m_\lambda = m \).

**Exercise E.5 (Riesz Markov Theorem for an Interval).** Suppose that \( X = [0,1] \) and \( \lambda \in C(X)^* \) and \( \lambda \geq 0 \). Define \( F(1) := \lambda(1) \) and for \( b \in [0,1) \), let \( F(b) = \inf_{\varepsilon > 0} \lambda(\varphi_{b,\varepsilon}) = \lim_{\varepsilon \to 0} \lambda(\varphi_{b,\varepsilon}) \) where
\[ \varphi_{b,\varepsilon}(x) = \begin{cases} 
1 & \text{if } x \leq b \\
1 - (x - b)/\varepsilon & \text{if } b \leq x \leq b + \varepsilon \\
0 & \text{if } x \geq b + \varepsilon.
\end{cases} \] (E.22)

Show \( F \) is an increasing right continuous function on \( X \). Let \( \mu \) be the unique measure on \( B_X \) such that \( \mu([0,b]) = F(b) \) for all \( b \in X \). Show also that
\[ \lambda(f) = \int_{0}^{1} f \, d\mu \text{ for all } f \in C(X). \] (E.23)

This is a simple version of Theorem ??.
Appendix: Math 140A Topics

F.1 Summary of Key Facts about Real Numbers

1. The real numbers, \( \mathbb{R} \), is the unique (up to order preserving field isomorphism) ordered field with the least upper bound property or equivalently which is Cauchy complete.

2. Informally the real numbers are the rational numbers with the (irrational) hole filled in.

3. Monotone bounded sequence always converge in \( \mathbb{R} \).

4. A sequence converges in \( \mathbb{R} \) iff it is Cauchy.

5. Cauchy sequences are bounded.

6. \( \mathbb{N} \) is unbounded from above in \( \mathbb{R} \).

7. For all \( \epsilon > 0 \) in \( \mathbb{R} \) there exists \( n \in \mathbb{N} \) such that \( \frac{1}{n} < \epsilon \).

8. \( \mathbb{Q} \) and \( \mathbb{R} \setminus \mathbb{Q} \) is dense in \( \mathbb{R} \). In particular, between any two real numbers \( a < b \), there are infinitely many rational and irrational numbers.

9. Decimal numbers map (almost 1-1) into the real numbers by taking the hole filled in.

10. If \( a, b, \epsilon \in \mathbb{R} \), then
   a) \( a \leq b \) by showing that \( a \leq b + \epsilon \) for all \( \epsilon > 0 \).
   b) \( a = b \) by proving \( a \leq b \) and \( b \leq a \) or
   c) \( a = b \) by showing \( |b - a| \leq \epsilon \) for all \( \epsilon > 0 \).

11. A number of standard limit theorems hold, see Theorem 3.13.

12. Unlike limits, \( \limsup \) and \( \liminf \) always exist. Moreover we have;
   \[ \liminf_{n \to \infty} a_n \leq \limsup_{n \to \infty} a_n \]
   with equality iff \( \lim_{n \to \infty} a_n \) exists in which case
   \[ \liminf_{n \to \infty} a_n = \lim_{n \to \infty} a_n = \limsup_{n \to \infty} a_n. \]

   We may allow the values of \( \pm \infty \) in these statements.

13. If \( b_k := \{a_{n_k}\}_{k=1}^\infty \) is a convergent subsequence of \( \{a_n\} \), then
   \[ \liminf_{n \to \infty} a_n \leq \lim_{k \to \infty} b_k \leq \limsup_{n \to \infty} a_n \]
   and we may choose \( \{b_k\} \) so that \( \lim_{k \to \infty} b_k = \limsup_{n \to \infty} a_n \) or
   \( \lim_{k \to \infty} b_k = \lim \inf_{n \to \infty} a_n \).

14. Bounded sequences of real numbers always have convergence subsequences.

15. If \( S \subset \mathbb{R} \) and \( A := \sup(S) \), then there exists \( \{a_n\}_{n=1}^\infty \subset S \) such that \( a_n \leq a_{n+1} \) for all \( n \) and \( \lim_{n \to \infty} a_n = \sup(S) \).

16. If \( S \subset \mathbb{R} \) and \( A := \inf(S) \), then there exists \( \{a_n\}_{n=1}^\infty \subset S \) such that \( a_{n+1} \leq a_n \) for all \( n \) and \( \lim_{n \to \infty} a_n = \inf(S) \).

F.2 Test 2 Review Topics:

1. Understand the basic properties of complex numbers.

2. Countability. Key facts are that countable union of countable sets is countable and the finite product of countable sets is countable.

3. Definitions of metric and normed spaces and their basic properties which in the end of the day typically follow from the triangle inequality.

4. You should know that metrics and norms are continuous functions that satisfy,
   \[ ||x|| - ||y|| \leq ||x - y|| \]

5. Be aware of different norms, \( ||\cdot||_1, ||\cdot||_1, \) and \( ||\cdot||_2 \).

6. Understand the notion of limits of sequences, Cauchy sequences, completeness, limits and continuity of functions.

7. Know what is meant by pointwise and uniform convergence. You should be able to compute pointwise limits and know how to test if the limit is uniform or not. A key theorem is the uniform limit of continuous functions is still continuous.

F.3 After Test 2 Review Topics:

Let (\( X, ||\cdot|| \)) be a Banach space.

1. Know \( \sum_{n=1}^\infty x_n = \lim_{N \to \infty} \sum_{n=1}^N x_n \) if the limit exists.

2. Know \( \sum_{n=1}^\infty x_n \) converges absolutely iff \( \sum_{n=1}^\infty ||x_n|| < \infty \) and absolute convergence implies convergence in a Banach space.

3. Telescoping series and geometric series.

5. Absolute convergence tests: 1) integral test, 2) root test, 3) ratio test often combined with the 4) comparison test.
7. $n^{th}$–term test for divergence.
8. Cauchy criteria for convergence and the fact that tails of convergent series tend to zero. i.e. tails of convergent series tend to zero.
10. Uniform convergence of sums and the Weierstrass $M$–test
11. Power series including radius of convergence notion.
12. The exponential function and its relatives, $\sin$, $\sinh$, $\cos$, $\cosh$. 