Analysis

May 8, 2021

Contents

Part I
Topology

1 Basic Definitions

Definition 1.1. A **topological space** is an ordered pair (X, τ), where X is a nonempty set and τ is a collection of subsets of X, satisfying the following axioms:

1. The empty set $\emptyset$ and the enitre set X belongs to τ.
2. τ is closed under arbitrary unions: if $U_i \in \tau$ for all i in some arbitrary index set I, then $\bigcup_{i \in I} U_i \in \tau$.
3. τ is closed under finite intersections: if $U_1, \dots, U_n \in \tau$, where $n \in \mathbb{N}$, then $\bigcap_{m=1}^{n} U_m \in \tau$.

The elements of τ are called **open sets** and the collection τ is called a **topology** on X.

Remark 1.

1. We often just write X instead of (X, τ) to denote a topological space. We also say τ gives X a topology.
2. Typically, one describes a topological space by specifying its open sets.

1.0.1 Comparison of Topologies

Definition 1.2. Let τ and τ' be two topologies on a set X. If $\tau \subseteq \tau'$, then we say τ is a **coarser** (**weaker** or **smaller**) topology than τ. Similarly, if $\tau \subseteq \tau'$, then we say τ' is a **finer** (**stronger** or **larger**) **topology** than τ.

Proposition 1.1. *Let τ and τ' be two topologies on a set X. Suppose that for every $x \in X$ and for every τ-open neighborhood U_x of x, there exists a τ'-open neighborhood U'_x of x such that $U'_x \subseteq U_x$. Then τ' is finer than τ.*

Proof. Let $U \in \tau$. For each $x \in U$, choose a τ'-open neighborhood U'_x of x such that $U'_x \subseteq U$. Then

$$\begin{aligned} U &= \bigcup_{x \in U} U'_x \\ &\in \tau'. \end{aligned}$$

It follows that τ' is finer than τ. □

1.0.2 Subspace Topology

Let (X, τ) be a topological space and let Z be a subset of X. We can give Z a topology by declaring open subsets of Z to be all sets of the form $U \cap Z$, where U is an open subset of X. One easily verifies that the collection of these open sets satisfy the axioms of forming a topology. We call this the **subspace topology induced by** τ.

1.0.3 Generating a Topology from a Collection of Subsets

Proposition 1.2. *Let X be a set and let $\mathscr{C}$ be a nonempty collection of subsets of X. Then there exists a smallest topology on X which contains $\mathscr{C}$. It is called the **topology generated by** $\mathscr{C}$ and is denoted $\tau(\mathscr{C})$. We also call $\mathscr{C}$ a **subbase** for $\tau(\mathscr{C})$.*

Proof. We define $\tau(\mathscr{C})$ to be the collection of all subsets of X obtained by adjoining to $\mathscr{C}$ the set X itself, empty set, and all arbitary unions of finite intersections of members of $\mathscr{C}$. To see that $\tau(\mathscr{C})$ is a topology, note that arbitrary unions of arbitrary unions of finite intersections is an arbitrary union of finite intersections, so it suffices to show that $\tau(\mathscr{C})$ is closed under finite intersections. Let $A, A' \in \tau(\mathscr{C})$. Then A has the form

$$A = \bigcup_{i \in I} (C_{i,1} \cap C_{i,2} \cap \cdots \cap C_{i,n_i})$$

where $C_{i,j} \in \mathscr{C}$ and $n_i \in \mathbb{N}$ for all $i \in I$ and $1 \leq j \leq n_i$. Similarly, A' has the form

$$A' = \bigcup_{i' \in I'} (C'_{i',1} \cap C'_{i',2} \cap \cdots \cap C'_{i',n'_{i'}})$$

where $C'_{i',j'} \in \mathscr{C}$ and $n'_{i'} \in \mathbb{N}$ for all $i' \in I'$ and $1 \leq j' \leq n'_{i'}$. Thus

$$\begin{aligned}
A \cap A' &= A \cap \left(\bigcup_{i' \in I'} (C'_{i',1} \cap C'_{i',2} \cap \cdots \cap C'_{i',n'_{i'}}) \right) \\
&= \bigcup_{i' \in I'} \left(A \cap C'_{i',1} \cap C'_{i',2} \cap \cdots \cap C'_{i',n'_{i'}} \right) \\
&= \bigcup_{i' \in I'} \left(\left(\bigcup_{i \in I} (C_{i,1} \cap C_{i,2} \cap \cdots \cap C_{i,n_i}) \right) \cap C'_{i',1} \cap C'_{i',2} \cap \cdots \cap C'_{i',n'_{i'}} \right) \\
&= \bigcup_{i' \in I'} \left(\bigcup_{i \in I} (C_{i,1} \cap C_{i,2} \cap \cdots \cap C_{i,n_i} \cap C'_{i',1} \cap C'_{i',2} \cap \cdots \cap C'_{i',n'_{i'}}) \right) \\
&= \bigcup_{(i,i') \in I \times I'} (C_{i,1} \cap C_{i,2} \cap \cdots \cap C_{i,n_i} \cap C'_{i',1} \cap C'_{i',2} \cap \cdots \cap C'_{i',n'_{i'}}) \\
&\in \tau(\mathscr{C}).
\end{aligned}$$

Thus $\tau(\mathscr{C})$ is closed under finite intersections. □

Definition 1.3. Let (X, τ) be a topological space. A collection $\mathscr{B}$ of open subsets of X is called a **base** (or **basis**) for τ if

1. $\mathscr{B}$ covers X,
2. For all $U, V \in \mathscr{B}$ and for all points $x \in U \cap V$, there exists $W \in \mathscr{B}$ such that $x \in W \subseteq U \cap V$.

1.0.4 Neighborhoods

Definition 1.4. Let X be a topological space and let $x \in X$. An open subset U of X is called an **open neighborhood of** x if $x \in U$. If U is a basis element in the topology, then we say U is a **basic open neighborhood of** x.

1.1 Continuous Functions

Definition 1.5. Let X and Y be topological spaces. A function $f \colon X \to Y$ is called **continuous** if $f^{-1}(V)$ is an open subset of X whenever V is an open subset of Y. We say f is **continous at a point** $x \in X$ if for any open neighborhood V of $f(x)$, there exists an open neighborhood U of x such that $f(U) \subseteq V$.

Remark 2. To check continuity of f at $x \in X$, it is enough to shows that for any *basic* open neighborhood V_0 of $f(x)$, there exists an open neighborhood U of x such that $f(U) \subseteq V_0$. Indeed, if this property holds, then for any open neighborhood V of $f(x)$, we choose a basic open neighborhood V_0 of $f(x)$ such that $V_0 \subseteq V$ and an open neighborhood U of x such that $f(U) \subseteq V_0 \subseteq V$.

Proposition 1.3. *Let X and Y be topological spaces. A function $f \colon X \to Y$ is continous if and only if it is continuous at every point $x \in X$.*

Proof. First assume that f is continous. Let $x \in X$ and V be an open neighborhood of $f(x)$. Then $f^{-1}(V)$ is an open subset of X since f is continous and, moreover, it contains x. Thus $f^{-1}(V)$ is an open neighborhood of x. Since x was arbitary, f is continous at every point in X.

Conversely, assume that f is continous at every point in X. Let V be an open subset of Y. We need to show that $f^{-1}(V)$ is an open subset of X. For all $x \in f^{-1}(V)$, we can find an open neighborhood U_x of x such that $U_x \subset f^{-1}(V)$ (i.e. $f(U_x) \subseteq V$), since f is continous at every point in X. Then

$$f^{-1}(V) = \bigcup_{x \in f^{-1}(V)} U_x$$

implies that $f^{-1}(V)$ is open. □

Remark 3. Suppose $f\colon X \to Y$ is continuous at a point $x_0 \in X$. One may suspect that f is continuous in some open neighborhood of x_0, but this is not the case. For a counterexample, consider the function $f : \mathbb{R} \to \mathbb{R}$, given by

$$f(x) = \begin{cases} x^2 & \text{if } x \text{ is rational,} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

We claim that f is continuous at 0 but is not continuous anywhere else. To show that f is continuous at 0, let $B_\varepsilon(0)$ be an ε-ball centered at $f(0) = 0$. Then $f\left(B_{\sqrt{\varepsilon}}(0)\right) \subset B_\varepsilon(0)$. Indeed if $x \in B_{\sqrt{\varepsilon}}(0)$ is rational, then $f(x) = 0 \in B_\varepsilon(0)$ and if $x \in B_{\sqrt{\varepsilon}}(0)$ is irrational, then $f(x) = x^2 \in B_\varepsilon(0)$ (since $|x| < \sqrt{\varepsilon}$ and hence $x^2 < \varepsilon$). It is an easy exercise to show that f is not continuous anywhere else.

Proposition 1.4. *Let X, Y be topological spaces, $f\colon X \to Y$ be a continuous function, and let $A \subset X$ be given the subspace topology. Then $f|_A\colon A \to Y$ is continuous.*

Proof. Let V be an open subset of Y. Then $f^{-1}(V)$ is an open subset of X since f is continuous. Therefore $(f|_A)^{-1}(V) = f^{-1}(V) \cap A$ is an open subset of A. □

1.2 Continuity in Metric Spaces

Definition 1.6. A **metric** on a set X is a function $d : X \times X \to \mathbb{R}$ which satisfies the following three properties:

1. (Identity of indiscernibles) $d(x, y) = 0$ if and only if $x = y$ for all $x, y \in X$,
2. (Symmetric) $d(x, y) = d(y, x)$ for all $x, y \in X$,
3. (Triangle Inequality) $d(x, z) \le d(x, y) + d(y, z)$ for all $x, y, z \in X$.

A set X together with a choice of a metric d is called a **metric space** and is denoted (X, d), or just denoted X if the metric is understood from context.

Remark 4. Given the three axioms above, we also have $d(x, y) \ge 0$ (positive-definitene) for all $x, y \in X$. Indeed,

$$\begin{aligned} 0 &= d(x, x) \\ &\le d(x, y) + d(y, x) \\ &= d(x, y) + d(x, y) \\ &= 2d(x, y). \end{aligned}$$

This implies $d(x, y) \ge 0$.

1.2.1 Open Balls

Definition 1.7. Let (X, d) be a metric space. For $x \in X$ and $\varepsilon > 0$, we define the **open ball centered at x of radius** ε, denoted $B_\varepsilon(x)$, to be the set

$$B_\varepsilon(x) := \{y \in X \mid d(x, y) < \varepsilon\}.$$

Proposition 1.5. *Let (X, d) be a metric space. If $B_\varepsilon(x)$ and $B_{\varepsilon'}(x')$ are two open balls centered at $x \in X$ (resp. $x' \in X$) of radius $\varepsilon > 0$ (resp. $\varepsilon' > 0$) such that $B_\varepsilon(x) \cap B_{\varepsilon'}(x') \neq \emptyset$, then there exists $x'' \in X$ and $\varepsilon'' > 0$ such that*

$$B_{\varepsilon''}(x'') \subseteq B_\varepsilon(x) \cap B_{\varepsilon'}(x).$$

Proof. Pick any $x'' \in B_\varepsilon(x) \cap B_{\varepsilon'}(x')$. Set $\delta = d(x, x'')$ and $\delta' = d(x', x'')$. Without loss of generality, say $\varepsilon - \delta \le \varepsilon' - \delta'$. Then we set $\varepsilon'' = \varepsilon - \delta$. If $y \in B_{\varepsilon''}(x)$, then

$$\begin{aligned} d(x, y) &\le d(x, x'') + d(x'', y) \\ &= \delta + d(x'', y) \\ &< \varepsilon \end{aligned}$$

implies $y \in B_\varepsilon(x)$ and

$$\begin{aligned} d(x', y) &\leq d(x', x'') + d(x'', y) \\ &= \delta' + d(x'', y) \\ &< \delta' + \varepsilon - \delta \\ &\leq \varepsilon' \end{aligned}$$

implies $y \in B_{\varepsilon'}(x)$. □

Let (X, d) be a metric space. The proposition above implies that the open balls form the base for a topology of X, making it a topological space. A topological space which can arise in this way from a metric space is called a **metrizable** space.

1.2.2 Epsilon-Delta and Metric Spaces

Let U be an open subset of $\mathbb{R}^n$. Then a function $f\colon U \to \mathbb{R}$ is continuous at a point $p \in U$ if for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\|q - p\| < \delta \text{ implies } \|f(q) - f(p)\| < \varepsilon.$$

In terms of open sets, this says for all basic open neighborhoods $B_\varepsilon(f(p))$ of $f(p)$, there is a basic open neighborhood $B_\delta(p)$ of p such that $f(B_\delta(p)) \subseteq B_\varepsilon(f(p))$.

Proposition 1.6. *Let (X, d) and (Y, d) be metric spaces. Then $f\colon X \to Y$ is continous at at a point $x \in X$ if and only if for all sequences (x_n) in X such that $x_n \to x$, we have $f(x_n) \to f(x)$.*

Proof. First suppose f is continous at $x \in X$. Let (x_n) be a sequence in X such that $x_n \to x$. We need to show that $f(x_n) \to f(x)$. Let V be an open neighborhood of $f(x)$. Since f is continuous at x, there exists an open neighborhood U of x such that $f(U) \subseteq V$. Since $x_n \to x$, there exists $N \in \mathbb{N}$ such that $x_n \in U$ for all $n \geq N$. Then $f(x_n) \in V$ for all $n \geq N$. This shows that $f(x_n) \to f(x)$.

Conversely, suppose that for all sequences (x_n) in X such that $x_n \to x$, we have $f(x_n) \to f(x)$. We need to show that f is continuous. We will prove that f is continuous at x by contradiction: assume that f is not continuous at x. Choose an open neighborhood V of $f(x)$ such that there is no neighborhood U of x where $f(U) \subseteq V$. Let

$$U_n := B_{\frac{1}{n}}(x) := \left\{ x' \in X \mid d(x, x') < \frac{1}{n} \right\}.$$

Choose $x'_n \in U_n$ such that $f(x'_n) \notin V$. Then (x'_n) is a sequence in X such that $x'_n \to x$, but $f(x'_n) \not\to f(x)$ (indeed, $f(x'_n)$ is never in V). Contradiction. □

Example 1.1. Consider the step function $f\colon \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0. \end{cases}$$

Then f is not continuous at $x = 0$ since, for example, the sequence $(1/n) \to 0$ but $(f(1/n)) \to 1 \neq f(0)$. On the other hand, the sequence $(-1/n) \to 0$ and $f(-1/n) = 0 = f(0)$. Thus we really do need f to preserve *all* convergent sequences in order for it to be continuous.

1.3 First-Countable Spaces

Definition 1.8. A topological space X is said to be **first-countable** if each point has a countable neighborhood basis. That is, for each $x \in X$, there exists a sequence (U_n) of open neighborhoods of x such that for any open neighborhood U of x there exists an $n \in \mathbb{N}$ such that $U_n \subseteq U$.

Proposition 1.7. *Let $f\colon X \to Y$ be a function and assume that X is first-countable. Then f is continous at a point $x \in X$ if and only if for all sequences (x_n) in X such that $x_n \to x$, we have $f(x_n) \to f(x)$.*

Proof. First suppose f is continous at $x \in X$. Let (x_n) be a sequence in X such that $x_n \to x$. Let V be an open neighborhood of $f(x)$. Since f is continuous at x, we can choose an open neighborhood U of x such that $f(U) \subseteq V$. Since $x_n \to x$, there exists an $N \in \mathbb{N}$ such that $n \geq N$ implies $x_n \in U$. Then $n \geq N$ implies

$$\begin{aligned} f(x_n) &\in f(U) \\ &\subseteq V. \end{aligned}$$

It follows that $f(x_n) \to f(x)$.

Conversely, suppose that for all sequences (x_n) in X such that $x_n \to x$, we have $f(x_n) \to f(x)$. Assume that f is not continuous at x. Choose an open neighborhood V of $f(x)$ such that there does not exist an open neighborhood U of x with $f(U) \subseteq V$. Now we apply first-countability of X. Choose a neighborhood basis of x, say (U_n). For each $n \in \mathbb{N}$ choose $x_n \in U_n$ such that $f(x_n) \notin V$. Then $x_n \to x$ since for any open neighborhood U of x, we can find an $N \in \mathbb{N}$ such that $n \geq N$ implies

$$x_n \in U_N \subseteq U.$$

On the other hand, $f(x_n) \not\to f(x)$ since $f(x_n) \notin V$ for all $n \in \mathbb{N}$. Contradiction. □

1.4 Discrete Topologies

Definition 1.9. Let X be a set. The **discrete topology** on X is defined by letting every subset of X be open.

Proposition 1.8. *Let X and Y be topological spaces.*

1. *If Y has the discrete topology. Then every continous map $f : X \to Y$ is locally constant*[a].
2. *If X has the discrete topology, then every function $f : X \to Y$ is continous.*

[a] This means for every $x \in X$, there exists an open neighborhood U_x of x such that f is constant on U_x: $f(y) = f(x)$ for all $y \in U_x$.

Proof.

1. Let $f : X \to Y$ be a continous function and $x \in X$. Then $\{f(x)\}$ is open in Y since Y has the discrete topology. Denote U_x to be the inverse image of $\{f(x)\}$ under f:

   $$U_x := f^{-1}\{f(x)\}.$$

   Then U_x is an open neighborhood x on which f is constant.

2. Let $f : X \to Y$ be a function and let V be an open subset of Y. Since X is discrete, every subset of X is open. In particular, $f^{-1}(V)$ is open.

□

1.4.1 Weakest Topology on Codomain

Let X be a set, $\{X_i\}_{i \in I}$ be a collection of topological spaces, and let $\{f_i : X_i \to X\}$ be a collection of functions. We want to give X a topology such that the maps f_i become continuous. We do this by declaring a subset U of X to be open if and only if $f_i^{-1}(U)$ is open in X_i for all i. That this really is a topology follows from the identities:

$$f_i^{-1}\left(\bigcup_{\lambda \in \Lambda} U_\lambda\right) = \bigcup_{\lambda \in \Lambda} f_i^{-1}(U_\lambda) \text{ and } f_i^{-1}\left(\bigcap_{\lambda \in \Lambda} U_\lambda\right) = \bigcap_{\lambda \in \Lambda} f_i^{-1}(U_\lambda).$$

1.4.2 Weakest Topology on Domain

Let X be a set, $\{X_i\}_{i \in I}$ be a collection of topological spaces, and let $\{f_i : X \to X_i\}$ be a collection of functions. We want to give X a topology such that the maps f_i become continuous. If U_i is an open subset of X_i, then certainly we need $f_i^{-1}(U_i)$ to be an open subset of X. We give X the smallest topology that contains all sets of the form $f_i^{-1}(U_i)$, where $i \in I$ and U_i is an open subset of i.

1.5 Gluing

Definition 1.10. Let X be a topological space. An **open covering** of X is a collection $\{U_i\}_{i\in I}$ of open subsets U_i of X such that

$$\bigcup_{i\in I} U_i = X.$$

Let X be a topological space and let $\{X_i\}$ be an open covering, so each X_i gets an induced topology. Note that a subset $U \subseteq X$ is open if and only if $U \cap X_i$ is open in X_i for each i. Indeed, one direction is clear. For the other direction, suppose $U \cap X_i$ is open in X_i for each i. Then for each i, there exists an open subset U_i of X such that $U_i \cap X_i = U \cap X$. Therefore

$$U = \bigcup_{i\in I} U \cap X_i = \bigcup_{i\in I} U_i \cap X_i,$$

shows that U is a union of open subsets of X.

If $f : X \to Y$ is a continouous map, then by restriction to X_i we get continuous maps $f_i : X_i \to Y$ such that

$$f_i \mid_{X_i \cap X_j} = f_j \mid_{X_i \cap X_j} \text{ for all } i \text{ and } j \tag{1}$$

Conversely, if we are given continuous maps $f_i : X_i \to Y$ such that (1) holds, then there is a unique set-theoretic map $f : X \to Y$ satisfying $f \mid_{X_i} = f_i$ for all i, and moreover it is continuous. Indeed, for any open subset V of Y we have $f^{-1}(V)$ is open in X because $f^{-1}(V) \cap X_i = f_i^{-1}(V)$ is open in X_i for every i. Hence, we can view continuous maps $X \to Y$ as collections of continuous maps $X_i \to Y$ that are compatible on the overlaps $X_i \cap X_j$. We want to run this procedure in reverse.

Theorem 1.1. *Let X be a set, and let $\{X_i\}$ be a collection of subsets whose union is X. Suppose on each X_i there is a given topology τ_i and that the τ_i's are compatible in the following sense: $X_i \cap X_j$ is open in each of X_i and X_j, and the induced topologies on $X_i \cap X_j$ from both X_i and X_j coincide. There is a unique topology on X that induces upon each X_i the topology τ_i.*

Remark 5. We say that the topology in this theorem is obtained by **gluing** the given topologies on the X_i's (We may also say that the topological space (X, τ) is obtained by **gluing** the topological spaces (X_i, τ_i).

Proof. We first prove uniqueness. If τ is a topology on X inducing τ_i for each i and making X_i open in X for each i, then a subset $U \subseteq X$ is open for τ if and only if $U \cap X_i$ is open for the induced topology on X_i for each i (as X_i is τ-open for every i), and hence (by the assumption that the induced topology on X_i is τ_i) if and only if $U \cap X_i$ is τ_i-open in X_i for each i. This final formulation of the openness condition for τ is expressed entirely in terms of the τ_i's and so establishes uniqueness: we have no choice as to what the condition of τ-openness is to be, and it must be the case that the τ-open sets in X are exactly those that meet each X_i in a τ_i-open subset of X_i for each i.

We now run the process in reverse to verify the existence. We *define* τ to be the collection of subsets $U \subseteq X$ such that $U \cap X_i$ is τ_i-open in X_i for each i. This topology is the weakest topology which makes the inclusion maps $X_i \hookrightarrow X$ continuous. Since for each fixed i_0 the overlap $X_{i_0} \cap X_j$ is τ_j-open in X_j for every j, it follows that X_{i_0} is τ-open in X for every i_0. □

2 Compactness

Definition 2.1. Let X be a topological space. We say X is **compact** every open covering of X contains a finite subcovering of X: if $\{U_i\}_{i\in I}$ covers X, then for some $n \in \mathbb{N}$ there exists $U_{i_1}, U_{i_2}, \ldots, U_{i_n} \in \{U_i\}_{i\in I}$ such that $\{U_{i_k}\}_{k=1}^n$ covers X. We say a subset K of X is a **compact subset** of X if K is compact with respect to the subspace topology.

Let $\mathcal{B}$ be a basis for X. To check for compactness for X, it is enough to only consider open coverings $\{U_i\}_{i\in I}$ where the U_i are in $\mathcal{B}$:

Proposition 2.1. *Let X be a topological space and let $\mathcal{B}$ be a basis for X. Then X is compact if and only if every open covering of X consisting of basis elements contains a finite subcovering of X.*

Proof. One direction is clear. For the other direction, assume that every open covering of X consisting of basis elements contains a finite subcovering of X. Let $\{U_i\}_{i\in I}$ be an open covering of X (where the U_i are not necessarily basis elements). For each $i \in I$, let $\{V_{i,j}\}_{j\in J}$ be an open covering of U_i consisting of basis elements (so the $V_{i,j}$

are basis elements). Then $\{V_{i,j}\}_{i\in I, j\in J}$ is an open covering of X consisting of basis elements and so there exists a finite subcovering, say $\{V_{i_\lambda, j_\gamma}\}_{\lambda\in\Lambda,\gamma\in\Gamma}$ where Λ and Γ are finite sets. Then $\{U_{i_\lambda}\}_{\lambda\in\Lambda}$ is a finite subcovering of $\{U_i\}_{i\in I}$. □

Remark 6. The proposition above is still true if we replace $\mathcal{B}$ with a subbase. However to prove this, we would need to use the Ultrafilter principle.

Lemma 2.1. *Let X be Hausdorff and let K be a compact subset of X. Then K is closed in X.*

Proof. We show that $X\backslash K$ is open. Let $x \in X\backslash K$. For each $y \in K$, choose an open neighborhood U_y of y and an open neighborhood V_y of x such that $U_y \cap V_y = \emptyset$. Since K is compact, the open covering $\{U_y \cap K\}_{y\in K}$ of K contains a finite subcovering of K, say $\{U_{y_i} \cap K\}_{i=1}^n$ where $y_i \in K$ for $i = 1, \dots, n$. Then

$$V_x := \bigcap_{i=1}^{n} V_{y_i}$$

is an open neighborhood of x which does not meet K. Therefore

$$X\backslash K = \bigcup_{x\in X\backslash K} V_x,$$

which implies $X\backslash K$ is open, which implies K is closed. □

2.0.1 Image of a Compact Space is Compact

Proposition 2.2. *Let $f\colon X \to Y$ be a continuous function from a compact space X to a topological space Y. Then $f(X)$ is a compact subspace of Y.*

Proof. Let $\{V_j \cap f(X)\}_{j\in J}$ be an open covering of $f(X)$, where the V_j are open subsets of Y. Then $\{f^{-1}(V_j)\}_{j\in J}$ is an open covering of X. Since X is compact, there exists a finite subcover of $\{f^{-1}(V_j)\}_{j\in J}$ wich covers X, say $\{f^{-1}(V_{j_1}), \dots, f^{-1}(V_{j_k})\}$. Then $\{V_{j_1} \cap f(X), \dots, V_{j_k} \cap f(X)\}$ is a finite subcover of $\{V_j \cap f(X)\}_{j\in J}$ which covers $f(X)$. Thus $f(X)$ is compact. □

2.0.2 Finite Intersection Property

There is another way of thinking about compactness.

Definition 2.2. Let X be a topological space. We say that X satisfies the **finite intersection property** (or **FIP**) for closed sets if any collection $\{Z_i\}_{i\in I}$ of closed sets in X with all finite intersections

$$Z_{i_1} \cap \cdots \cap Z_{i_n} \neq \emptyset,$$

the intersection $\bigcap_{i\in I} Z_i$ of all Z_i's is non-empty.

Theorem 2.2. *Let X be a topological space. Then X is compact if and only if it satisfies FIP for closed sets.*

Proof. This is an exercise in linguistics. Suppose first that X is compact. To obtain a contradiction, assume that X does not satisfy FIP for closed sets. Then there exists a collection $\{Z_i\}_{i\in I}$ of closed sets in X with all finite intersections $Z_{i_1} \cap \cdots \cap Z_{i_n} \neq \emptyset$ and with $\bigcap_{i\in I} Z_i = \emptyset$. But this implies $\{X\backslash Z_i\}_{i\in I}$ is an open cover of X with no finite subcover. The converse is proved in exactly the same way. □

2.0.3 When a continuous bijection is a homeomorphism

Lemma 2.3. *Let X be a compact space and let E be a closed subset of X. Then E is also compact.*

Proof. Let $\{U_i \cap E\}_{i\in I}$ be an open cover of E. Then $(X\backslash E) \cup \{U_i \cap E\}_{i\in I}$ is an open cover of X. Since X is compact, there exists a finite subcover in $(X\backslash E) \cup \{U_i \cap E\}_{i\in I}$ of X. In particular, this implies that there exists a finite subcover in $\{U_i \cap E\}_{i\in I}$ of E. □

Lemma 2.4. *Let X be a compact space, Y be any topological space, and let $f : X \to Y$ be continuous surjective map. Then Y is compact.*

Proof. Let $\{V_i\}_{i\in I}$ be an open cover of Y. Since f is continuous, $\{f^{-1}(V_i)\}_{i\in I}$ is an open cover of X. Since X is compact, there exists a finite subcover in $\{f^{-1}(V_i)\}_{i\in I}$ of X, say $\{f^{-1}(V_{i_1}),\dots,f^{-1}(V_{i_n})\}$. But then $\{V_{i_1},\dots,V_{i_n}\}$ is a finite subcover in $\{V_i\}_{i\in I}$ of Y. $\square$

Theorem 2.5. *Let X and Y be topological spaces such that X is compact and Y is Hausdorff, and let $f : X \to Y$ be a continuous bijection. Then f is a homeomorphism.*

Proof. Let $g : Y \to X$ denote the inverse of f. We need to show that g is continuous. We do this by showing that the inverse image of a closed set in X is a closed set in Y: Let E be a closed set in X. Since X is compact, E is compact by Lemma (2.3). Since E is compact, $f(E)$ is compact by Lemma (2.4). Since Y is Hausdorff and $f(E)$ is compact, $f(E)$ is closed by Lemma (2.1). But $f(V) = g^{-1}(V)$, so $g^{-1}(V)$ is closed. $\square$

2.0.4 Closes subspaces of compact spaces are compact

Proposition 2.3. *Let X be a compact space and let A be a closed subspace of X. Then A is compact.*

Proof. Let $\{U_i \cap A\}_{i\in I}$ be an open covering of A. Then $(X\backslash A) \cup \{U_i\}_{i\in I}$ is an open covering of X. Since X is compact, it contains a finite subcovering of X, say $(X\backslash A) \cup \{U_{i_k}\}_{k=1}^n$. But then $\{U_{i_k} \cap A\}_{k=1}^n$ must be a finite subcovering of $\{U_i \cap A\}_{i\in I}$. $\square$

2.1 Heine-Borel Theorem

Definition 2.3. Let S be a subset of a topological space X. We say $x \in X$ is a **limit point** of S if every open neighborhood of x meets S: if U is an open subset of X such that $x \in U$, then $U \cap S \neq \emptyset$.

Theorem 2.6. *Let S be a subset of Euclidean space $\mathbb{R}^n$. Then S is compact if and only if it is closed and bounded.*

Proof. Suppose that S is compact. Since $\mathbb{R}^n$ is Hausdorff, Lemma (2.1) implies S is closed. It remains to show that S is bounded, which we will do by contradiction: assume S is not bounded. For each $x \in S$, let $U_x = B_1(x)$ be the open ball of radius 1 centered at x. Then $\{U_x\}_{x\in S}$ forms an open cover of S. Since S is compact, there eixsts a finite subcover of $\{U_x\}_{x\in S}$, say $\{U_{x_1},\dots,U_{x_n}\}$. Let

$$L_{ij} = \sup\left\{\|a_i - a_j\| \mid a_i \in U_{x_i} \text{ and } a_j \in U_{x_j}\right\}$$

clearly L_{ij} is finite since $L_{ij} \le \|x_i - x_j\| + 2$. Setting $L = \max_{1\le i,j\le n}\{L_{ij}\}$, we see that for all $a, a' \in S$, we must have $\|a - a'\| \le L$. Thus, S is bounded.

Conversely, suppose that S is closed and bounded. Since S is bounded, it is enclosed within an n-box $T_0 = [-a, a]^n$ where $a > 0$. Since $\mathbb{R}^n$ is Hausdorff, a closed subset of a compact set is compact, and so it suffices to show T_0 is compact. Assume, by way of contradiction, that T_0 is not compact. Then there exists an infinite open cover $\{U_i\}_{i\in I}$ of T_0 that does not admit any finite subcover. Through bisection of each of the sides of T_0, the box T_0 can be broken up into 2^n sub n-boxes, each of which has diameter equal to half the diameter of T_0. Then at least one of the 2^n sections of T_0 must require an infinite subcover of $\{U_i\}$, otherwise $\{U_i\}$ itself would have a finite subcover, by uniting together the finite covers of the sections. Call this section T_1.

Likewise, the sides of T_1 can be bisected, yielded 2^n sections of T_1, at least one of which must require an infinite subcover of $\{U_i\}$. Continuing in this manner yields a decreasing sequence of nexted n-boxes:

$$T_0 \supset T_1 \supset T_2 \supset \cdots \supset T_k \supset \cdots$$

where the side length of T_k is $(2a)/2^k$, which tends to 0 as k tends to infinity. Let us define a sequence (x_k) such that each x_k is in T_k. This sequence is Cauchy, so it must converge to some limit x. Since each T_k is closed, and for each k the sequence (x_k) is eventually always inside T_k, we see that $x \in T_k$ for each k.

Since $\{U_i\}$ covers T_0, it has some member $U \in \{U_i\}$ such that $x \in U$. Since U is open, there is an n-ball $B_\varepsilon(x) \subset U$ for some $\varepsilon > 0$. For large enough k (for example such that $(2a)/2^k < \varepsilon$), one has

$$T_k \subset B_\varepsilon(x) \subset U,$$

but then the infinite number of members of $\{U_i\}$ needed to cover T_k can be replaced by just one: U, a contradiction. Thus, T_0 is compact. $\square$

Remark 7. The Heine-Borel theorem does not hold as stated for general metric and topological vector spaces. For instance, at one point in our proof we used completeness, which doesn't hold in a general metric space. A metric space (X, d) is said to have the **Heine-Borel property** if each closed bounded set in X is compact.

2.1.1 Sequential Compactness

A topological space X is said to be **sequentially compact** if every sequence of points in X has a convergent subsequence converging to a point in X. In general, there are compact spaces which are not sequentially compact and there are sequentially compact spaces that are not compact. However, when it comes to metric spaces, these notions are equivalent. We will prove this in the case of the Euclidean space $\mathbb{R}^n$.

Theorem 2.7. *Let S be a subset of Euclidean space $\mathbb{R}^n$. Then S is sequentially compact if and only if it is closed and bounded.*

Proof. We first assume that S is sequentially compact. We will first show that S is closed.

Let (x_n) be a convergent sequence in S, and suppose that $x_n \to x$ as $n \to \infty$. Since S is sequentially compact, we can choose a convergent subsequence (x_{n_k}) of (x_n) which converges to a point in X. Since every convergent subsequence of a convergent sequence converges to the same limit, we have $x_{n_k} \to x$ as $k \to \infty$. This establishes that S is closed.

Now we will show that S is bounded. Assume (for a contradiction) that S is unbounded. Since S is unbounded, there exists a sequence (x_n) in S such that

$$x_n \notin \bigcup_{m=1}^{n-1} B_1(x_m).$$

for all $n \in \mathbb{N}$. Such a sequence has no convergent subsequence since for each $n \in \mathbb{N}$, the neighborhood $B_1(x_n)$ contains only one member in the sequence (namely x_n).

To complete the proof of the theorem, we now assume that S is closed and bounded. We will show that S is sequentially compact. Let (x_n) be a sequence in S. Since S is closed and bounded, it lies in a closed box, say $B_0 = [-a, a]^n$ where $a > 0$. Through bisection of each of the sides of B_0, the box B_0 can be broken up into 2^n sub n-boxes, each of which has diameter equal to half the diameter of B_0. Then at least one of the 2^n sections of B_0 contains infinitely elements in the sequence (x_n). Call this section B_1.

Likewise, the sides of B_1 can be bisected, yielded 2^n sections of B_1, at least one of which must contain infinitely many elements in the sequence (x_n). Continuing in this manner yields a decreasing sequence of nexted n-boxes:

$$B_0 \supset B_1 \supset B_2 \supset \cdots \supset B_k \supset \cdots$$

where the side length of B_k is $(2a)/2^k$, which tends to 0 as k tends to infinity. Now we define a convergent subsequence of (x_n) as follows: For each $k \in \mathbb{N}$, we choose x_{n_k} inductively on k to be a member of the sequence (x_n) which lies in the box B_k and such that $x_{n_k} \neq x_{n_{k-1}}$ for all $k \in \mathbb{N}$. The sequence (x_{n_k}) is is Cauchy, so it must converge to some limit x. Since each T_k is closed, and for each k the sequence (x_k) is eventually always inside T_k, we see that $x \in T_k$ for each k. Finally, since S is closed, we must have $x \in S$. This establishes that S is sequentially compact. □

2.1.2 Extreme Value Theorem

Proposition 2.4. *Let X be a compact space and let $f\colon X \to \mathbb{R}$ be continuous. Then f obtains a maximum value, i.e. there exists $x_0 \in X$ such that $f(x_0) \geq f(x)$ for all $x \in X$.*

Proof. Assume X is nonempty, otherwise it is trivial. Since X is compact, $f(X)$ is a compact subset of $\mathbb{R}$. By the Heine-Borel theorem, $f(X)$ is a closed and bounded subset of $\mathbb{R}$. Since $f(X)$ is nonempty and bounded above, the limit $\sup(f(X))$ exists. Moreover, since $f(X)$ is closed and $\sup f(X)$ is a limit point of $f(X)$, we have $\sup(f(X)) \in f(X)$. Thus $\sup(f(X)) = f(x_0)$ for some $x_0 \in X$, and this is clearly the maximum value. □

3 Closure and Interior

3.1 Closure

Definition 3.1. Let X be a topological space and let A be a subset of X. The **closure** of A in X, denoted $\overline{A}$, is the smallest closed set in X which contains A. It is characterized by the universal property that if E is a closed set in X such that $E \subseteq A$, then $E \subseteq \overline{A}$. Indeed,

$$\overline{A} = \bigcap_{\substack{E \text{ closed} \\ A \subseteq E}} E.$$

3.1.1 Uniqueness of Continuous Extensions of Functions from a Set to its Closure

Lemma 3.1. *Let X be a topological space, A be a subset of X, and let $x \in \overline{A}$. If U is an open neighborhood of x, then U meets A (i.e. $U \cap A \neq \emptyset$).*

Proof. To obtain a contradiction, assume $U \cap A = \emptyset$. Then A is contained in the closed set X/U, which implies $\overline{A}$ is contained in the closed set $X \backslash U$. But this is a contradiction since $x \in \overline{A} \cap U$, whence $x \in \overline{A}$ and $x \notin X \backslash U$. □

Proposition 3.1. *Let X and Y be topological spaces such that Y is Hausdorff. Let A be a subset of X and let $f \colon A \to Y$ be a continuous map. Suppose there exists a continuous extension $\widetilde{f} \colon \overline{A} \to Y$ of f (i.e. $\widetilde{f}$ is continuous and $\widetilde{f}(a) = f(a)$ for all $a \in A$), then $\widetilde{f}$ is unique.*

Proof. To prove uniqueness, suppose that $\widetilde{f}_1 \colon \overline{A} \to Y$ and $\widetilde{f}_2 \colon \overline{A} \to Y$ are two continuous extensions of f. Then there exists $x \in \overline{A}$ such that $\widetilde{f}_1(x) \neq \widetilde{f}_2(x)$. Choose open neighborhoods V_1 and V_2 of $\widetilde{f}_1(x)$ and $\widetilde{f}_2(x)$ respectively such that $V_1 \cap V_2 = \emptyset$ (we can do this since Y is Hausdorff). Then $\widetilde{f}_1^{-1}(V_1) \cap \widetilde{f}_2^{-1}(V_2)$ is an open neighborhood of $x \in \overline{A}$, and so it must meet A by Lemma (3.1). This is a contradiction though, since $a \in A \cap \widetilde{f}_1^{-1}(V_1) \cap f_2^{-1}(V_2)$ implies $V_1 \cap V_2 \neq \emptyset$. Indeed,

$$\begin{aligned} V_1 \ni \widetilde{f}_1(a) \\ = f_1(a) \\ = f_2(a) \\ = \widetilde{f}_2(a) \in V_2. \end{aligned}$$

□

Proposition 3.2. *Let X be a topological spaces such that Y is Hausdorff. Suppose that for every subset A of X and let $f \colon A \to Y$ be a continuous map. Suppose there exists a continuous extension $\widetilde{f} \colon \overline{A} \to Y$ of f (i.e. $\widetilde{f}$ is continuous and $\widetilde{f}(a) = f(a)$ for all $a \in A$), then $\widetilde{f}$ is unique.*

Proof. To prove uniqueness, suppose that $\widetilde{f}_1 \colon \overline{A} \to Y$ and $\widetilde{f}_2 \colon \overline{A} \to Y$ are two continuous extensions of f. Then there exists $x \in \overline{A}$ such that $\widetilde{f}_1(x) \neq \widetilde{f}_2(x)$. Choose open neighborhoods V_1 and V_2 of $\widetilde{f}_1(x)$ and $\widetilde{f}_2(x)$ respectively such that $V_1 \cap V_2 = \emptyset$ (we can do this since Y is Hausdorff). Then $\widetilde{f}_1^{-1}(V_1) \cap \widetilde{f}_2^{-1}(V_2)$ is an open neighborhood of $x \in \overline{A}$, and so it must meet A by Lemma (3.1). This is a contradiction though, since $a \in A \cap \widetilde{f}_1^{-1}(V_1) \cap f_2^{-1}(V_2)$ implies $V_1 \cap V_2 \neq \emptyset$. Indeed,

$$\begin{aligned} V_1 \ni \widetilde{f}_1(a) \\ = f_1(a) \\ = f_2(a) \\ = \widetilde{f}_2(a) \in V_2. \end{aligned}$$

□

4 Metric Spaces

Definition 4.1. A **metric** on a set X is a function $d \colon X \times X \to \mathbb{R}$ which satisfies the following three properties:

1. (Identity of Indiscernibles) $d(x, y) = 0$ if and only if $x = y$ for all $x, y \in X$,
2. (Symmetric) $d(x, y) = d(y, x)$ for all $x, y \in X$,
3. (Triangle Inequality) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

A set X together with a choice of a metric d is called a **metric space** and is denoted (X, d), or just denoted X if the metric is understood from context.

Remark 8. Given the three axioms above, we also have $d(x,y) \geq 0$ (Positive-Definiteness) for all $x, y \in X$. Indeed,

$$\begin{aligned} 0 &= d(x,x) \\ &\leq d(x,y) + d(y,x) \\ &= d(x,y) + d(x,y) \\ &= 2d(x,y). \end{aligned}$$

This implies $d(x,y) \geq 0$.

Example 4.1. On $\mathbb{R}^m$ the **Euclidean metric** is

$$d_E(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\| = \sqrt{(x_1-y_1)^2 + \cdots + (x_m-y_m)^2}.$$

This is the usual distance used in $\mathbb{R}^m$, and when we speak about $\mathbb{R}^m$ as a metric space without specifying a metric, it's the Euclidean metric that is intended.

To check that d_E is a metric on $\mathbb{R}^m$, the first two conditions in the definition are obvious. The third condition is a consequence of the inequality $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (replace $\mathbf{x}$ with $\mathbf{x}-\mathbf{z}$ and $\mathbf{y}$ with $\mathbf{z}-\mathbf{y}$), and to show this inequality holds we will write $\|\mathbf{x}\|$ in terms of the dot product: $\|\mathbf{x}\|^2 = x_1^2 + \cdots + x_m^2 = \mathbf{x}\cdot\mathbf{x}$, so

$$\begin{aligned} \|\mathbf{x}+\mathbf{y}\|^2 &= (\mathbf{x}+\mathbf{y})\cdot(\mathbf{x}+\mathbf{y}) \\ &= \mathbf{x}\cdot\mathbf{x} + \mathbf{x}\cdot\mathbf{y} + \mathbf{y}\cdot\mathbf{x} + \mathbf{y}\cdot\mathbf{y} \\ &= \|\mathbf{x}\|^2 + 2\mathbf{x}\cdot\mathbf{y} + \|\mathbf{y}\|^2 \\ &\leq \|\mathbf{x}\|^2 + 2|\mathbf{x}\cdot\mathbf{y}| + \|\mathbf{y}\|^2. \end{aligned}$$

The famous Cauchy-Schwarz inequality says $|\mathbf{x}\cdot\mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|$, so

$$\|\mathbf{x}+\mathbf{y}\|^2 \leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2$$

and now take square roots.

A different metric on $\mathbb{R}^m$ is

$$d_\infty(\mathbf{x},\mathbf{y}) = \max_{1\leq i\leq m} |x_i - y_i|.$$

Again, the first two conditions of being a metric are clear, and to check the triangle inequality we use the fact that it is known for the absolute value. If $\max |x_i - y_i| = |x_k - y_k|$ for a particular k from 1 to m, then $d_\infty(\mathbf{x},\mathbf{y}) = |x_k - y_k|$, so

$$\begin{aligned} d_\infty(\mathbf{x},\mathbf{y}) &\leq |x_k - z_k| + |z_k - y_k| \\ &\leq \max_{1\leq i\leq m} |x_i - z_i| + \max_{1\leq i\leq m} |z_i - y_i| \\ &= d_\infty(\mathbf{x},\mathbf{z}) + d_\infty(\mathbf{z},\mathbf{y}). \end{aligned}$$

While the metrics d_E and d_∞ on $\mathbb{R}^m$ are different, they're not that different from each other since each is bounded by a constant multiple of the other one:

$$d_E(\mathbf{x},\mathbf{y}) \leq \sqrt{m}d_\infty(\mathbf{x},\mathbf{y}) \text{ and } d_\infty(\mathbf{x},\mathbf{y}) \leq d_E(\mathbf{x},\mathbf{y}).$$

Example 4.2. Let $C[0,1]$ be the space of all continuous functions from $[0,1]$ to $\mathbb{R}$. Two metrics used on $C[0,1]$ are

$$d_\infty(f,g) = \max_{0\leq x\leq 1} |f(x) - g(x)| \text{ and } d_1(f,g) = \int_0^1 |f(x) - g(x)|dx.$$

Unlike with the two metrics on $\mathbb{R}^m$ while we have $d_1(f,g) \leq d_\infty(f,g)$ there is no constant $A > 0$ that makes $d_\infty(f,g) \leq Ad_1(f,g)$ for all f and g. These metrics d_1 and d_∞ on $C[0,1]$ are quite different.

4.1 Metric Space Induced by a Norm

Definition 4.2. Let V be a vector space over a subfield F of the complex numbers. A **norm** on V is a nonnegative-valued scalar function $p : V \to [0, \infty)$ such that for all $a \in F$ and $u, v \in V$, we have

1. (Subadditivity) $p(u+v) \leq p(u) + p(v)$,
2. (Absolutely Homogeneous) $p(av) = |a|p(v)$,
3. (Positive-Definite) $p(v) = 0$ implies $v = 0$.

We call the pair (V, p) a **normed vector space**.

Proposition 4.1. *Let (V, p) be a normed vector space. Define $d : V \times V \to \mathbb{R}$ by $d(u, v) = p(u - v)$ for all $(u, v) \in V \times V$. Then (V, d) is a metric space.*

Proof. Let us first check that d satisfies the identity of indiscernibles property. Since p is positive-definite, $d(u, v) = 0$ implies $p(u - v) = 0$ which implies $u = v$. On the other hand, suppose $u = v$. Then since p is absolutely homogeneous, we have $p(0) = |0|p(0) = 0$, and so $d(u, u) = p(0) = 0$.

Next we check that d is symmetric. For all $(u, v) \in V \times V$, we have

$$\begin{aligned} d(u, v) &= p(u - v) \\ &= p(-1(v - u)) \\ &= |-1|p(v - u) \\ &= p(v - u) \\ &= d(v, u). \end{aligned}$$

Finally, triangle inequality for d follows from subadditivity of p. Indeed, for all $u, v, w \in V$, we have

$$\begin{aligned} d(u, v) + d(v, w) &= p(u - v) + p(v - w) \\ &\geq p(u - w) \\ &= d(u, w). \end{aligned}$$

□

Remark 9. The metric d induced by a norm p has additional properties that are not true of general metrics. These are

1. (Translation Invariance) $d(u + w, v + w) = d(u, v)$ for all $u, v, w \in V$
2. (Scaling Property) $d(au, av) = |a|d(u, v)$ for all $a \in F$ and $u, v \in V$.

Convserely, if a metric has these properties, then $d(u, 0)$ is a norm.

4.2 Limit of a Sequence in a Metric Space

Definition 4.3. For a sequence (x_n) in a metric space (X, d), we say (x_n) **converges to** $x \in X$, and write $\lim_{n\to\infty} x_n = 0$ or $x_n \to x$, if for every $\varepsilon > 0$ there is an $N = N_\varepsilon \in \mathbb{N}$ such that

$$n \geq N \text{ implies } d(x_n, x) < \varepsilon.$$

If a sequence in (X, d) has a limit, then we say that the sequence is **convergent**.

Theorem 4.1. *If a sequence (x_n) in a metric space (X, d) converges, then $d(x_n, x_{n+1}) \to 0$.*

Proof. Suppose $x_n \to x$. From the triangle inequality, we have

$$d(x_n, x_{n+1}) \leq d(x_n, x) + d(x, x_{n+1}) = d(x_n, x) + d(x_{n+1}, x).$$

The two terms on the right get small when n is large, so $d(x_n, x_{n+1})$ gets small when n is large. To be precise, for $\varepsilon/2 > 0$ there's an $N \geq 1$ such that for all $m \geq N$ we have $d(x_m, x) < \varepsilon/2$. Therefore

$$n \geq N \text{ implies } n + 1 \geq N \text{ implies } d(x_n, x_{n+1}) \leq d(x_n, x) + d(x_{n+1}, x) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

□

Theorem 4.2. *Every subsequence of a convergent sequence in a metric space is also convergent, with the same limit.*

Proof. Let $x_n \to x$ in (X, d) and let (x_{n_i}) be a subsequence of (x_n). Set $y_i = x_{n_i}$. We want ot show $y_i \to x$.

For $\varepsilon > 0$ there is an N such that $n \geq N$ implies $d(x_n, x) < \varepsilon$. Since the interegers n_i are increasing, we have $n_i \geq N$ if we go out far enough: there's an I such that $i \geq I$ implies $n_i \geq N$ which imlpies $d(x_{n_i}, x) < \varepsilon$, so $d(y_i, x) < \varepsilon$. Thus $y_i \to x$. □

Theorem 4.3. *In a metric space (X, d), if two sequences (x_n) and (x'_n) converge to the same value, then $d(x_n, x'_n) \to 0$.*

Proof. Suppose $x_n \to x$ and $x'_n \to x$ for some $x \in X$ and let $\varepsilon > 0$. Then there exists some integer $N \in \mathbb{N}$ such that

$$n \geq N \text{ implies } d(x_n, x) < \frac{\varepsilon}{2} \text{ and } d(x'_n, x) < \frac{\varepsilon}{2}.$$

In particular,

$$n \geq N \text{ implies } d(x_n, x'_n) \leq d(x_n, x) + d(x, x'_n) < \varepsilon.$$

□

Example 4.3. On $\mathbb{R}^m$, because the metrics d_E and d_∞ are each bounded above by a constant multiple of the other, we have $d_E(x_n, x) \to 0$ if and only if $d_\infty(x_n, x) \to 0$. Indeed, let $\varepsilon/\sqrt{m} > 0$. Then there exists some $N \in \mathbb{N}$ such that

$$n \geq N \text{ implies } d_\infty(x_n, x) < \frac{\varepsilon}{\sqrt{m}},$$

and since $d_E(x_n, x) \leq \sqrt{m} d_\infty(x_n, x)$,

$$n \geq N \text{ implies } d_E(x_n, x) < \varepsilon.$$

The converse is shown the same way. Therefore convergence of sequences in $\mathbb{R}^m$ for both metrics means the same thing (with the same limits).

Example 4.4. In $C[0, 1]$ consider the sequence of functions x^n for $n \geq 1$.This sequence converges to 0 in the metric d_1 but not in the metric d_∞:

$$d_1(x^n, 0) = \int_0^1 |x^n| dx = \frac{1}{n+1} \to 0, \qquad d_\infty(x^n, 0) = \max_{0 \leq x \leq 1} |x^n| = 1.$$

In fact the sequence (x^n) in $C[0, 1]$ has no limit at all relative to the metric d_∞. To prove (x^n) has no limit in $(C[0, 1], d_\infty)$, not just that the constant function 0 is not a limit, we seek a property that all convergent sequences satisfy and the sequence (x^n) in $(C[0, 1], d_\infty)$ does not satisfy. This will be provided to us in the next section.

4.3 Cauchy Sequences and Completeness

Recall from Theorem (4.1) that for a sequence (x_n) in a metric space (X, d) to converge, it is necessary that $d(x_n, x_{n+1}) \to 0$. On the other hand, this condition is not sufficient. Indeed, in $C[0, 1]$, we have

$$d_\infty(x^n, x^{n+1}) = \max_{0 \leq x \leq 1} |x^n - x^{n+1}| = \max_{0 \leq x \leq 1} (x^n - x^{n+1}),$$

To find the maximal value, we first compute the derivative

$$\frac{d}{dx}\left(x^n - x^{n+1}\right) = nx^{n-1} - (n+1)x^n = x^{n-1}(n - (n+1)x).$$

Setting this equal to 0, we have either $x = 0$ or $x = n/(n+1)$. Since $x^n - x^{n+1} = x^n(1 - x)$ is always positive on $[0, 1]$, it must be maximized on $[0, 1]$ at $x = n/(n+1)$, where the value is

$$\left(\frac{n}{n+1}\right)^n \left(1 - \frac{n}{n+1}\right) \sim \frac{1}{e}\left(\frac{1}{n+1}\right) \to 0.$$

Thus, we do have $d_\infty(x^n, x^{n+1}) \to 0$. However we shall see shortly that this sequence does not converge under the d_∞ metric. Indeed, by the exact same reasoning in the proof of Theorem (4.1), we should also have $d_\infty(x^n, x^{2n}) \to 0$. But

$$d_\infty(x^n, x^{2n}) = \max_{0 \leq x \leq 1} |x^n - x^{2n}| = \max_{0 \leq x \leq 1} (x^n(1 - x^n)),$$

has its maximum value at $x = 1/\sqrt[n]{2}$ where $x^n(1-x^n) = 1/4$, which is independent of n. This proves that (x^n) has not limit in $(C[0,1], d_\infty)$.

Theorem 4.4. *If (x_n) is a convergent sequence in a metric space (X,d), then the terms of the sequence become "uniformly close": for every $\varepsilon > 0$ there is an $N \in \mathbb{N}$ such that*

$$m, n \geq N \text{ implies } d(x_n, x_m) < \varepsilon.$$

Proof. Letting $x = \lim_{n\to\infty} x_n$, the triangle inequality tells us for all m and n that

$$d(x_m, x_n) \leq d(x_m, x) + d(x, x_n) = d(x_m, x) + d(x_n, x).$$

We now make an $\varepsilon/2$ argument. For every $\varepsilon > 0$ there's an $N \in \mathbb{N}$ such that for all $n \geq N$ we have $d(x_n, x) < \varepsilon/2$. Therefore $m, n \geq N$ implies

$$\begin{aligned} d(x_m, x_n) &\leq d(x_m, x) + d(x, x_n) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon. \end{aligned}$$

□

Definition 4.4. A sequence (x_n) in a metric space (X,d) is called a **Cauchy sequence** if for every $\varepsilon > 0$ there exists an $N = N_\varepsilon \in \mathbb{N}$ such that

$$m, n \geq N \text{ implies } d(x_m, x_n) < \varepsilon.$$

Corollary 1. *If (X,d) is a metric space and Y is a subset of X given the induced metric $d\,|_Y$, then any sequence in Y that converges in X is a Cauchy sequence in $(Y, d\,|_Y)$.*

Proof. A sequence (y_n) in Y that converges in X is Cauchy in X by Theorem (4.4). Since the metric d on X is the metric we are using on Y, the Cauchy property of (y_n) in X can be viewed as the Cauchy property in Y. □

Example 4.5. Consider the interval $(0,\infty)$ as a metric space using the absolute value metric induced from $\mathbb{R}$. We have $1/n \to 0$ in $\mathbb{R}$, but the sequence $(1/n)$ has no limit in $(0,\infty)$ since $0 \notin (0,\infty)$. The sequence $(1/n)$ is a Cauchy sequence in $(0,\infty)$ by Corollary (1) but it is not a convergent sequence in $(0,\infty)$.

Theorem 4.5. *If (x_n) is a sequence in a metric space (X,d) such that $d(x_n, x_{n+1}) \leq ar^n$ for all n, where $a > 0$ and $0 < r < 1$, then (x_n) is a Cauchy sequence.*

Proof. For $1 \leq m < n$, a massive use of the triangle inequality tells us

$$\begin{aligned} d(x_m, x_n) &\leq d(x_m, x_{m+1}) + d(x_{m+1}, x_{m+2}) + \cdots + d(x_{n-1}, x_n) \\ &\leq ar^m + ar^{m+1} + \cdots + ar^{n-1} \\ &< \sum_{k=m}^{\infty} ar^k \\ &= \frac{ar^m}{1-r}. \end{aligned}$$

Since $0 < r < 1$, the terms ar^n tend to 0 as $n \to \infty$. Now if we pick an $\varepsilon > 0$, choose N large enough that $ar^N < (1-r)\varepsilon$. For $m, n \geq N$, without loss of generality $m \leq n$ so by our prior calculation

$$\begin{aligned} d(x_m, x_n) &< \frac{ar^m}{1-r} \\ &\leq \frac{ar^N}{1-r} \\ &< \varepsilon. \end{aligned}$$

□

4.4 Complete Metric Spaces

Definition 4.5. A metric space (X,d) is called **complete** if every Cauchy sequence in X converges in X: if (x_n) is Cauchy in X then there's an $x \in X$ such that $x_n \to x$.

4.4.1 Completions exist

Our goal in this subsection is to prove the following theorem:

Theorem 4.6. *Let X be a metric space. Then a completion of X exists.*

Let C_X denote the set of Cauchy sequences in the given metric space X. We say two elements $(x_n), (y_n) \in C_X$ are **equivalent**, written $(x_n) \sim (y_n)$, if $\rho(x_n, y_n) \to 0$ as $n \to \infty$. We claim that this is an equivalence relation on C_X. The only nontrivial part to check is transitivity: Suppose $(x_n) \sim (y_n)$ and $(y_n) \sim (z_n)$. Then

$$\rho(x_n, z_n) \leq \rho(x_n, y_n) + \rho(y_n, z_n) \to 0 + 0$$

as $n \to \infty$.

We denote by $C_X/\sim$ to be the set of equivalence classes in C_X under $\sim$. Note that there is a natural map of sets

$$\iota_X : X \to C_X/\sim,$$

which assigns to each $x \in X$, the equivalence class of constant sequences $(x) \in C_X$. The map ι_X is injective. Indeed, if $\iota_X(x) = \iota_X(x')$, then $(x) \sim (x')$, and so $\rho(x, x') \to 0$ forces $x = x'$.

We now must enhance the structure of $C_X/\sim$ by giving it a metric (with respect to which we'll easily see ι_X is an isometry with dense image). We first define a "pseudo-metric" on the set C_X. Let's first define what we mean by "pseudo-metric":

Definition 4.6. A **pseudo-metric** on a set S is a function $d : S \times S \to \mathbb{R}$ which satisfies the following three properties:

1. $d(s, s) = 0$ for all $s \in S$,
2. $d(s, t) = d(t, s)$ for all $s, t \in S$,
3. $d(s, u) \leq d(s, t) + d(t, u)$ for all $s, t, u \in S$.

Remark 10.

1. The same proof as in the metric case shows that a pseudo-metric is positive-definite (i.e. $d(s, t) \geq 0$ for all $s, t \in S$).
2. The only difference between a metric and a pseudo-metric is that a pseudo-metric might satisfy $d(s, t) = 0$ for some pair $s, t \in S$ with $s \neq t$.

Lemma 4.7. *If (s_n) and (t_n) are Cauchy sequences in S, then $(\rho(s_n, t_n))$ is a convergent sequence in $\mathbb{R}$.*

Proof. We show that $(\rho(s_n, t_n))$ is a Cauchy sequence in $\mathbb{R}$. Let $\varepsilon > 0$. Since (s_n) and (t_n) are Cauchy sequences, there exists $N \in \mathbb{N}$ such that $\rho(s_n, s_m) < \varepsilon/2$ and $\rho(t_n, t_m) < \varepsilon/2$ for all $n, m \geq N$.

Note that $\rho(s_n, t_n) \leq \rho(s_n, s_m) + \rho(s_m, t_m) + \rho(t_m, t_n)$ implies

$$\rho(s_n, t_n) - \rho(s_m, t_m) \leq \rho(s_n, s_m) + \rho(t_m, t_n),$$

and implies

$$\rho(s_m, t_m) - \rho(s_n, t_n) \leq \rho(s_m, s_n) + \rho(t_n, t_m).$$

Thus, for all $n, m \geq N$, we have

$$|\rho(s_n, t_n) - \rho(s_m, t_m)| \leq \rho(s_n, s_m) + \rho(t_m, t_n) < \varepsilon.$$

□

With the lemma established, the following definition makes sense:

Definition 4.7. For $(x_n), (y_n) \in C_X$, we define the **pseudo-distance** between them to be

$$d((x_n), (y_n)) = \lim_{n \to \infty} \rho(x_n, y_n).$$

The pseudo-distance defined is a pseudo-metric since it inherits the properties these from ρ.

Corollary 2. *For $c_1, c_2, c_1', c_2' \in C_X$ with $c_j \sim c_j'$ we have $d(c_1, c_2) = d(c_1', c_2')$. In other words, the pseudo-metric d on C_X respects the equivalence relation.*

Proof. Note that $c_j \sim c'_j$ if and only if $d(c_j, c'_j) = 0$. Thus

$$\begin{aligned} d(c_1, c_2) &\le d(c_1, c'_1) + d(c'_1, c'_2) + d(c'_2, c_2) \\ &= d(c'_1, c'_2). \end{aligned}$$

Similarly,

$$\begin{aligned} d(c'_1, c'_2) &\le d(c'_1, c_1) + d(c_1, c_2) + d(c_2, c'_2) \\ &= d(c_1, c_2). \end{aligned}$$

Thus $d(c_1, c_2) = d(c'_1, c'_2)$. $\square$

Definition 4.8. Define $X' = C_X/\sim$ and define the function $\rho' : X' \times X' \to \mathbb{R}$ by

$$\rho'(\xi_1, \xi_2) = d(c_1, c_2)$$

where $c_1, c_2 \in C_X$ are respective representative elements for $\xi_1, \xi_2 \in X' = C_X/\sim$.

This definition is well-defined in view of Corollary (2). In fact, the function $\rho' : X' \times X' \to \mathbb{R}$ is a metric since it inherits the pseudo-metric properties from ρ, and modding out by the equivalence relation ensures that we have identity of indiscernibles. Note that we also have

$$\rho'(\iota_X(x), \iota_X(y)) = \rho(x, y)$$

for all $x, y \in X$. Thus, $\iota_X : X \to X'$ is an isometric embedding. In fact the image of X is dense in X'. Indeed, fix a choice of $\xi \in X'$. Choose a representative Cauchy sequence $(x_n) \in C_X$ for the equivalence class $\xi \in X'$. Then the sequence of elements $\iota_X(x_n) \in \iota_X(X) \subseteq X'$ has limit ξ.

4.5 Open and Closed Subsets

In this section we generalize open and closed intervals in $\mathbb{R}$ to open and closed balls of a metric space. Throughout, let (X, d) be a metric space.

Definition 4.9. For $a \in X$ and $r \geq 0$, the **open ball** with center a and radius r is

$$\mathrm{B}_r(a) = \{x \in X \mid d(a, x) < r\},$$

and the **closed ball** with center a and radius r is

$$\mathrm{B}_r[a] = \{x \in X \mid d(a, x) \leq r\}.$$

When $r = 0$, we have $B(a, 0) = \emptyset$ and $\overline{B}(a, 0) = \{a\}$.

Definition 4.10. A subset of X is called **bounded** if it is contained in some ball $B(a, r)$. A subset that is not bounded is called **unbounded**.

Definition 4.11. A subset $U \subset X$ is called **open** if for each $x \in U$ there's an $r > 0$ such that $B(x, r) \subset U$. We also consider the empty subset of X to be an open subset.

5 Pseudometric Spaces

Definition 5.1. A **pseudometric** on a set X is a function $\mathrm{d}\colon X \times X \to \mathbb{R}$ which satisfies the following three properties:

1. (Reflexivity) $\mathrm{d}(x, x) = 0$ for all $x \in X$;
2. (Symmetry) $\mathrm{d}(x, y) = \mathrm{d}(y, x)$ for all $x, y \in X$,
3. (Triangle Inequality) $\mathrm{d}(x, z) \leq \mathrm{d}(x, y) + \mathrm{d}(y, z)$ for all $x, y, z \in X$.

If d is a pseudometric on a set X, then we call the pair (X, d) a **pseudometric space**. If the pseudometric is understood from context, then we often denote a pseudometric space by X instead of (X, d).

Remark 11. Given the three axioms above, we also have $d(x,y) \geq 0$ for all $x, y \in X$. Indeed,

$$\begin{aligned} 0 &= d(x,x) \\ &\leq d(x,y) + d(y,x) \\ &= d(x,y) + d(x,y) \\ &= 2d(x,y). \end{aligned}$$

This implies $d(x,y) \geq 0$.

5.1 Topology Induced by Pseudometric Space

Proposition 5.1. *Let (X, d) be a pseudometric space. For each $x \in X$ and $r > 0$, define*

$$\mathrm{B}_r^{\mathrm{d}}(x) := \{y \in X \mid \mathrm{d}(x,y) < r\},$$

and let

$$\mathcal{B}^{\mathrm{d}} = \{\mathrm{B}_r(x) \mid x \in X \text{ and } r > 0\}.$$

Finally, let $\tau(\mathcal{B}^{\mathrm{d}})$ be the smallest topology on X which contains $\mathcal{B}^{\mathrm{d}}$. Then $\mathcal{B}^{\mathrm{d}}$ is a basis for $\tau(\mathcal{B}^{\mathrm{d}})$.

Remark 12. We often remove the d in the superscript in $\mathrm{B}_r^{\mathrm{d}}(x)$ and $\mathcal{B}^{\mathrm{d}}$ whenever context is clear.

Proof. First note that $\mathcal{B}$ covers X. Indeed, for any $r > 0$, we have

$$X \subseteq \bigcup_{x \in X} \mathrm{B}_r(x).$$

Next, let $\mathrm{B}_r(x)$ and $\mathrm{B}_{r'}(x')$ be two members of $\mathcal{B}$ which have nontrivial intersection and let $x'' \in \mathrm{B}_r(x) \cap \mathrm{B}_{r'}(x')$. Set

$$r'' = \min\{r' - \mathrm{d}(x', x''), r - \mathrm{d}(x, x'')\}.$$

We claim that $\mathrm{B}_{r''}(x'') \subseteq \mathrm{B}_r(x) \cap \mathrm{B}_{r'}(x')$. Indeed, assume without loss of generality that $r'' = r - \mathrm{d}(x, x'')$. Let $y \in \mathrm{B}_{r''}(x'')$. Then

$$\begin{aligned} \mathrm{d}(y,x) &\leq \mathrm{d}(y,x'') + \mathrm{d}(x'',x) \\ &< r - \mathrm{d}(x,x'') + \mathrm{d}(x'',x) \\ &= r - \mathrm{d}(x'',x) + \mathrm{d}(x'',x) \\ &= r \end{aligned}$$

implies $y \in \mathrm{B}_r(x)$. Similarly,

$$\begin{aligned} \mathrm{d}(y,x') &\leq \mathrm{d}(y,x'') + \mathrm{d}(x'',x') \\ &< r' - \mathrm{d}(x',x'') + \mathrm{d}(x'',x') \\ &= r' - \mathrm{d}(x'',x') + \mathrm{d}(x'',x') \\ &= r' \end{aligned}$$

implies $y \in \mathrm{B}_{r'}(x')$. Thus $\mathrm{B}_{r''}(x'') \subseteq \mathrm{B}_r(x) \cap \mathrm{B}_{r'}(x')$, and so $\mathcal{B}$ is a basis for $\tau(\mathcal{B})$. □

Definition 5.2. The topology $\tau(\mathcal{B})$ in Proposition (41.1) is called the **topology induced by the pseudometric** d. We also denote this topology by τ_{d}.

5.1.1 Subspace topology agrees with topology induced by pseudometric

Let (X, d) be a pseudometric space and let $A \subseteq X$. Then the pseudometric on X restricts to a pseudometric on A. We denote this restriction by $\mathrm{d}|_A$. Thus there are two natural topologies on A. One is the subspace is the subspace topology given by

$$\tau \cap A := \{U \cap A \mid U \in \tau\}.$$

The other is the topology induced by the pseudometric $\mathrm{d}|_A$ given by

$$\tau_{\mathrm{d}|_A} := \tau(\mathcal{B}^{\mathrm{d}}).$$

The next proposition tells us that these are actually the same.

Proposition 5.2. *Let (X, d) be a pseudometric space and let $A \subseteq X$. Then*

$$\tau_{\mathrm{d}} \cap A = \tau_{\mathrm{d}|_A}.$$

Proof. Let $a \in A$ and $r > 0$. Then

$$\begin{aligned} \mathrm{B}_r^{\mathrm{d}|_A}(a) &= \{b \in A \mid \mathrm{d}|_A(a, b) < r\} \\ &= \{b \in A \mid \mathrm{d}(a, b) < r\} \\ &= A \cap \{x \in X \mid \mathrm{d}(a, x) < r\} \\ &= A \cap \mathrm{B}_r^{\mathrm{d}}(a). \end{aligned}$$

It follows that $\tau_{\mathrm{d}|_A}$ and $\tau_{\mathrm{d}} \cap A$ have the same basis, and hence $\tau_{\mathrm{d}} \cap A = \tau_{\mathrm{d}|_A}$. □

5.1.2 Convergence in (X, d)

Concepts like convergence and completion still make sense in pseudometric spaces. This is because these are purely topological concepts.

Definition 5.3. Let (X, d) be a pseudometric space and let (x_n) be a sequence in X.

1. We say the sequence (x_n) converges to $x \in X$ if for all $\varepsilon > 0$ there exists an $N_\varepsilon \in \mathbb{N}$[a] such that

$$n \geq N_\varepsilon \text{ implies } \mathrm{d}(x_n, x) < \varepsilon.$$

   In this case, we say (x_n) is a **convergent** and that it **converges** to x. We denote this by $x_n \to x$ as $n \to \infty$, or $\lim\limits_{n\to\infty} x_n = x$, or even just $x_n \to x$.

2. We say the sequence (x_n) is **Cauchy** if for all $\varepsilon > 0$ there exists an $N_\varepsilon \in \mathbb{N}$ such that

$$n, m \geq N_\varepsilon \text{ implies } \mathrm{d}(x_n, x_m) < \varepsilon.$$

[a] We write ε in the subscript to remind the reader that N_ε depends on ε. Usually we omit ε in the subscript and just write N.

5.1.3 Completeness in (X, d)

In a metric space, every Cauchy sequence is convergence but the converse may not hold. The same thing is true for pseudometric spaces since the proof is purely topological. Let's go over the proof again:

Proposition 5.3. *Let (x_n) be a sequence in X, let $x \in X$, and suppose $x_n \to x$. Then (x_n) is Cauchy.*

Proof. Let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies

$$\mathrm{d}(x_n, x) < \varepsilon/2.$$

Then $n, m \geq N$ implies

$$\begin{aligned} \mathrm{d}(x_n, x_m) &\leq \mathrm{d}(x_n, x) + \mathrm{d}(x, x_m) \\ &< \varepsilon/2 + \varepsilon/2 \\ &= \varepsilon. \end{aligned}$$

This implies (x_n) is Cauchy. □

Thus, the concept of completeness makes sense in a pseudometric space.

Definition 5.4. Let (X, d) be a pseudometric space. We say (X, d) is **complete** if every Cauchy sequence in (X, d) is a convergent.

5.2 Metric Obtained by Pseudometric

Unless otherwise specified, we let (X, d) be a pseudometric space throughout the remainder of this section. There is a natural way to obtain a metric space from (X, d) which we now describe as follows: define a relation $\sim$ on X by

$$x \sim y \text{ if and only if } \mathrm{d}(x,y) = 0.$$

Then $\sim$ is an equivalence relation. Indeed, we have reflexivity of $\sim$ since $\mathrm{d}(x,x) = 0$ for all $x \in X$, we have symmetry of $\sim$ since $\mathrm{d}(x,y) = \mathrm{d}(y,x)$ for all $x, y \in X$, and we have transitivity of $\sim$ since d satisfies the triangle inequality: if $x \sim y$ and $y \sim z$, then

$$\begin{aligned} \mathrm{d}(x,z) &\leq \mathrm{d}(x,y) + \mathrm{d}(y,z) \\ &= 0 + 0 \\ &= 0. \end{aligned}$$

Thus $\mathrm{d}(x,z) = 0$ which implies $x \sim z$.

Therefore we may consider the quotient space of X with respect to the equivalence relation above. We shall denote this quotient space by $[X] := X/\sim$. A coset in $[X]$ which is represented by $x \in X$ will be written as $[x]$. There is a natural **projection map** $\pi \colon X \to [X]$ that sends $x \in X$ to its equivalence class $[x]$. Since π is surjective, any subset of $[X]$ has the form

$$[A] = \{[a] \in [X] \mid a \in A\}.$$

We are ready to define the metric on $[X]$.

Theorem 5.1. *Define* $[\mathrm{d}] \colon [X] \times [X] \to \mathbb{R}$ *by*

$$[\mathrm{d}]([x],[y]) = \mathrm{d}(x,y) \tag{2}$$

for all $[x], [y] \in [X]$*. Then* $[\mathrm{d}]$ *is a metric on* $[X]$*. It is called the metric* ***induced*** *by the pseudometric.*

Proof. We first show that (104) is well-defined. Indeed, choose different coset representatives of $[x]$ and $[y]$, say x' and y' respectively (so $\mathrm{d}(x,x') = 0$ and $\mathrm{d}(y,y') = 0$). Then

$$\begin{aligned} [\mathrm{d}]([x'],[y']) &= \mathrm{d}(x',y') \\ &\leq \mathrm{d}(x',x) + \mathrm{d}(x,y) + \mathrm{d}(y,y') \\ &= \mathrm{d}(x,y) \\ &= [\mathrm{d}]([x],[y]). \end{aligned}$$

Thus $[\mathrm{d}]$ is well-defined.

Next we show that $[\mathrm{d}]$ is in fact a metric on $[X]$. First we check $[\mathrm{d}]$ is symmetric. Let $[x], [y] \in [X]$. Then

$$\begin{aligned} [\mathrm{d}]([x],[y]) &= \mathrm{d}(x,y) \\ &= \mathrm{d}(y,x) \\ &= [\mathrm{d}]([y],[x]). \end{aligned}$$

Thus $[\mathrm{d}]$ is symmetric. Next we check $[\mathrm{d}]$ satisfies triangle inequality. Let $[x], [y], [z] \in [X]$. Then

$$\begin{aligned} [\mathrm{d}]([x],[z]) &= \mathrm{d}(x,z) \\ &\leq \mathrm{d}(x,y) + \mathrm{d}(y,z) \\ &= [\mathrm{d}]([x],[y]) + [\mathrm{d}]([y],[z]). \end{aligned}$$

Thus $[\mathrm{d}]$ satisfies triangle inequality. Finally we check $[\mathrm{d}]$ satisfies identify of indiscernables. Let $[x], [y] \in [X]$ and suppose $[\mathrm{d}]([x],[y]) = 0$. Then

$$\begin{aligned} 0 &= [\mathrm{d}]([x],[y]) \\ &= \mathrm{d}(x,y) \end{aligned}$$

implies $x \sim y$ by definition. Therefore $[x] = [y]$. Thus $[\mathrm{d}]$ satisfies identify of indiscernables. □

5.2.1 Completeness in (X, d) is equivalent to completeness in $([X], [\mathrm{d}])$

As in the case of the pseudometric d, the metric $[\mathrm{d}]$ induces a topology on $[X]$. We denote this topology by $\tau_{[\mathrm{d}]}$.

Proposition 5.4. *(X, d) is complete if and only if $([X], [\mathrm{d}])$ is complete.*

Proof. Suppose that (X, d) is complete. Let $([x_n])$ be a Cauchy sequence in $([X], [\mathrm{d}])$. We claim (x_n) is a Cauchy sequence in (X, d). Indeed, let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $m, n \geq N$ implies

$$[\mathrm{d}]([x_n], [x_m]) < \varepsilon.$$

Then $m, n \geq N$ implies

$$\begin{aligned} \mathrm{d}(x_n, x_m) &= [\mathrm{d}]([x_n], [x_m]) \\ &< \varepsilon. \end{aligned}$$

This implies (x_n) is a Cauchy sequence in (X, d). Since (X, d) is complete, the sequence converges to a (not necessarily unique) $x \in X$. Then we claim that $[x_n] \to [x]$. Indeed, let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies

$$\mathrm{d}(x_n, x) < \varepsilon.$$

Then $n \geq N$ implies

$$\begin{aligned} [\mathrm{d}]([x_n], [x]) &= \mathrm{d}(x_n, x) \\ &< \varepsilon. \end{aligned}$$

This implies $[x_n] \to [x]$. Thus $([X], [\mathrm{d}])$ is complete.

Conversely, suppose $([X], [\mathrm{d}])$ is complete. Let (x_n) be a Cauchy sequence in (X, d). We claim $([x_n])$ is a Cauchy sequence in $([X], [\mathrm{d}])$. Indeed, let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $m, n \geq N$ implies

$$\mathrm{d}(x_n, x_m) < \varepsilon.$$

Then $m, n \geq N$ implies

$$\begin{aligned} [\mathrm{d}]([x_n], [x_m]) &= \mathrm{d}(x_n, x_m) \\ &< \varepsilon. \end{aligned}$$

This implies (x_n) is a Cauchy sequence in $([X], [\mathrm{d}])$. Since $([X], [\mathrm{d}])$ is complete, the sequence converges to a unique $[x] \in [X]$. We claim that $x_n \to x$ (in fact it converges to any $y \in X$ such that $y \sim x$). Indeed, let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies

$$[\mathrm{d}]([x_n], [x]) < \varepsilon.$$

Then $n \geq N$ implies

$$\begin{aligned} \mathrm{d}(x_n, x) &= [\mathrm{d}]([x_n], [x]) \\ &< \varepsilon. \end{aligned}$$

This implies $x_n \to x$. Thus (X, d) is complete. □

5.3 Quotient Topology

Recall that we view X as a topological space with toplogy τ_{d}; the topology induced by the pseudometric d. It turns out that there are two natural topologies on $[X]$. One such topology is $\tau_{[\mathrm{d}]}$; the topology induced by the metric $[\mathrm{d}]$. The other topology is called the **quotient topology with respect to** $\sim$, and is denoted by $[\tau_{\mathrm{d}}]$, where $[\tau_{\mathrm{d}}]$ is defined by

$$[\tau_{\mathrm{d}}] = \{[A] \subseteq [X] \mid \pi^{-1}([A]) \in \tau_{\mathrm{d}}\}.$$

In other words, we declare a subset $[A]$ of $[X]$ to be open in $[X]$ if and only if

$$\begin{aligned} \pi^{-1}([A]) &= \{x \in X \mid x \sim a \text{ for some } a \in A\} \\ &= \{x \in X \mid \mathrm{d}(x, a) = 0 \text{ for some } a \in A\} \end{aligned}$$

is open in X. Since $\pi^{-1}(\emptyset) = \emptyset$ and $\pi^{-1}([X]) = X$, we see that both $\emptyset$ and $[X]$ are open in $[X]$. Furthermore, since

$$\pi^{-1}\left(\bigcup_{i \in I}[A_i]\right) = \bigcup_{i \in I} \pi^{-1}([A_i]) \text{ and } \pi^{-1}\left(\bigcap_{i \in I}[A_i]\right) = \bigcap_{i \in I} \pi^{-1}([A_i]),$$

we see that the collection of open sets in $[X]$ is closed under arbitrary unions and finite intersections. Therefore $[\tau_{\mathrm{d}}]$ is indeed a topology on $[X]$. Note that $[\tau_{\mathrm{d}}]$ was defined in such a way that it makes the projection map $\pi\colon X \to [X]$ continous.

5.3.1 Universal Mapping Property For Quotient Space

Quotient spaces satisfy the following universal mapping property.

Proposition 5.5. *Let $f\colon X \to Y$ be any continuous function which is constant on each equivalence class. Then there exists a unique continuous function $[f]\colon [X] \to Y$ such that $f = [f] \circ \pi$.*

Proof. We define $[f]\colon [X] \to Y$ by

$$[f]([x]) = f(x) \tag{3}$$

for all $x \in X$. We first show that (**??**) is well-defined. Suppose x and x' are two different representatives of the same coset (so $x \sim x'$). Then $f(x) = f(x')$ as f was assumed to be constant on equivalence classes, and so

$$\begin{aligned} [f]([x']) &= f(x') \\ &= f(x) \\ &= [f]([x]). \end{aligned}$$

Thus (**??**) is well-defined.

Next we want to show that $[f]$ is continuous. Let V be an open set in Y. Then

$$f^{-1}(V) = \pi^{-1}\left([f]^{-1}(V)\right)$$

is open in X. By the definition of quotient topology, this implies $[f]^{-1}(V)$ is open in $[X]$. This implies $[f]$ is continuous.

Finally, we want to show that $f = [f] \circ \pi$ holds. Let $x \in X$. Then we have

$$\begin{aligned} ([f] \circ \pi)(x) &= [f](\pi(x)) \\ &= [f]([x]) \\ &= f(x). \end{aligned}$$

It follows that $[f] \circ \pi = f$. This establishes existence of f.

For uniqueness, assume for a contradiction that $\overline{f}\colon [X] \to Y$ is a continuous function such that $f = \overline{f} \circ \pi$ and such that $\overline{f} \neq [f]$. Choose $[x] \in [X]$ such that $\overline{f}[x] \neq [f][x]$. Then

$$\begin{aligned} f(x) &= (\overline{f} \circ \pi)(x) \\ &= \overline{f}(\pi(x)) \\ &= \overline{f}([x]) \\ &\neq [f]([x]) \\ &= f(x), \end{aligned}$$

which gives us a contradiction. □

It follows from Proposition (41.5) that we have the following bijection of sets

$$\{f\colon X \to Y \mid f \text{ is continuous and constant on equivalence classes}\} \cong \{\text{continuous functions from } [X] \text{ to } Y\}.$$

In particular, if we want to study continuous functions out of $[X]$, then we just need to study the continuous functions out of X which are constant on equivalence classes.

Proposition 5.6. *Suppose (Y, d_Y) is a metric space and $f\colon (X, \mathrm{d}) \to (Y, \mathrm{d}_Y)$ is continuous. The f is constant on equivalence classes.*

Proof. Let $x, x' \in X$ such that $x \sim x'$. Thus $\mathrm{d}(x, x') = 0$. Let $\varepsilon > 0$. Choose $\delta > 0$ such that

$$\mathrm{d}(x, y) < \delta \text{ implies } \mathrm{d}_Y(f(x), f(y)) < \varepsilon.$$

We want to show that $f(x) = f(x')$. □

5.3.2 Open Equivalence Relation

An equivalence relation $\sim$ on a topological space X is said to be **open** if the projection map $\pi\colon X \to [X]$ is open. In other words, the equivalence relation $\sim$ on X is open if and only if for every open set U in X, the set

$$\pi^{-1}(\pi(U)) = \bigcup_{x \in U} [x]$$

of all points equivalent to some point of U is open. The imporance of open equivalence relations is that if $\mathscr{B}$ is a basis for X, then $[\mathscr{B}]$ is a basis for $[X]$.

Lemma 5.2. *Let $x \in X$ and $r > 0$. Then*

$$\mathrm{B}_r(x) = \pi^{-1}([\mathrm{B}_r(x)]).$$

In particular, π is an open mapping.

Proof. We have

$$\begin{aligned} \mathrm{B}_r(x) &\subseteq \pi^{-1}\left(\pi\left(\mathrm{B}_r(x)\right)\right) \\ &= \pi^{-1}([\mathrm{B}_r(x)]). \end{aligned}$$

For the reverse inclusion, let $y \in \pi^{-1}([\mathrm{B}_r(x)])$. Then $\mathrm{d}(y,z) = 0$ for some $z \in \mathrm{B}_r(x)$. Choose such a $z \in \mathrm{B}_r(x)$. Then

$$\begin{aligned} \mathrm{d}(y,x) &\leq \mathrm{d}(y,z) + \mathrm{d}(z,x) \\ &= \mathrm{d}(z,x) \\ &< r \end{aligned}$$

implies implies $y \in \mathrm{B}_r(x)$. Therefore

$$\pi^{-1}([\mathrm{B}_r(x)]) \subseteq \mathrm{B}_r(x).$$

Thus each subset in $[X]$ of the form $[\mathrm{B}_r(x)]$ is open in $[X]$.

To see that π is an open mapping, let U be an open set in X. Since the set of all open balls is a basis for τ_{d}, we can cover U by open balls, say

$$U = \bigcup_{i \in I} \mathrm{B}_{r_i}(x_i).$$

Then

$$\begin{aligned} \pi(U) &= \pi\left(\bigcup_{i \in I} \mathrm{B}_{r_i}(x_i)\right) \\ &= \bigcup_{i \in I} \pi\left(\mathrm{B}_{r_i}(x_i)\right) \\ &= \bigcup_{i \in I} [\mathrm{B}_{r_i}(x_i)] \\ &\in [\tau_{\mathrm{d}}]. \end{aligned}$$

Thus π is an open mapping. □

5.3.3 Quotient Topology Agrees With Metric Topology

Theorem 5.3. *With the notation as above, we have*

$$[\tau_{\mathrm{d}}] = \tau_{[\mathrm{d}]}.$$

Proof. We first note that for each $x \in X$ and $r > 0$, we have

$$\begin{aligned} [\mathrm{B}_r(x)] &= \{[y] \in [X] \mid y \in \mathrm{B}_r(x)\} \\ &= \{[y] \in [X] \mid \mathrm{d}(y,x) < r\} \\ &= \{[y] \in [X] \mid [\mathrm{d}]([y],[x]) < r\} \\ &= \mathrm{B}_r([x]). \end{aligned}$$

In particular, $\tau_{[\mathrm{d}]}$ and $[\tau_{\mathrm{d}}]$ share a common basis. Therefore $\tau_{[\mathrm{d}]} = [\tau_{\mathrm{d}}]$. □

6 Quotient Topology

Let (X, τ) be a topological space and let $\sim$ be an equivalence relation on X. We denote $[X] := X/\sim$. A coset in $[X]$ which is represented by $x \in X$ will be written as $[x]$. There is a natural **projection map** $\pi : X \to [X]$ that sends $x \in X$ to its equivalence class $[x]$. Since π is surjective, any subset of $[X]$ has the form

$$[A] = \{[a] \in [X] \mid a \in A\}.$$

With these considerations in mind, we define a topology on $[X]$ called the **quotient topology with respect to the equivalence relation** $\sim$, which we denote by $[\tau]$, by

$$[\tau] = \{[A] \subseteq [X] \mid \pi^{-1}([A]) \in \tau\}.$$

In other words, we declare a subset $[A]$ of $[X]$ to be open in $[X]$ if and only if

$$\pi^{-1}([A]) = \{x \in X \mid x \sim a \text{ for some } a \in A\}$$

is open in X. Since $\pi^{-1}(\varnothing) = \varnothing$ and $\pi^{-1}([X]) = X$, we see that both $\varnothing$ and $[X]$ are open in $[X]$. Furthermore, since

$$\pi^{-1}\left(\bigcup_{i \in I}[A_i]\right) = \bigcup_{i \in I}\pi^{-1}([A_i]) \text{ and } \pi^{-1}\left(\bigcap_{i \in I}[A_i]\right) = \bigcap_{i \in I}\pi^{-1}([A_i]),$$

we see that the collection of open sets in $[X]$ is closed under arbitrary unions and finite intersections. Therefore $[\tau]$ is indeed a topology on $[X]$.

Note that $[\tau]$ was defined in such a way that it makes the projection map $\pi\colon X \to [X]$ continuous. Also, any open set in $[X]$ can be represented by an open set in X. Indeed, suppose $[A]$ is open in $[X]$. Denote $U = \pi^{-1}([A])$. Then $[U] = [A]$.

Proposition 6.1. *Let $\mathscr{B}$ be a basis for X. Then $[\mathscr{B}]$ is a basis for $[X]$.*

Proof. It is clear that $\mathscr{B}$ covers $[X]$. Let $[U]$ and $[V]$ be two elements in $\mathscr{B}$ and assume that U and V are open in X. Then □

6.0.1 Continuity of a Map on a Quotient

Let $f\colon X \to Y$ be a continuous function. If f is constant on each equivalence class, then it induces a map $[f]\colon [X] \to Y$ defined by

$$[f][x] = f(x) \tag{4}$$

for all $x \in X$. To see that (**??**) is well-defined, suppose x and x' are two different representatives of the same coset (so $x \sim x'$). Then $f(x) = f(x')$ as f was assumed to be constant on equivalence classes, and so

$$\begin{aligned}[f][x'] &= f(x') \\ &= f(x) \\ &= [f][x].\end{aligned}$$

Thus (**??**) is well-defined. We also have continuity:

Proposition 6.2. *The induced map $[f]\colon [X] \to Y$ is continuous if and only if the map $f\colon X \to Y$ is continuous.*

Proof. Suppose $[f]$ is continuous. Then f is continuous since $f = [f] \circ \pi$ is a composition of two continuous functions. Conversely, suppose f is continuous. Let V be an open set in Y. Then

$$f^{-1}(V) = \pi^{-1}\left([f]^{-1}(V)\right)$$

is open in X. By the definition of quotient topology, this implies $[f]^{-1}(V)$ is open in $[X]$. This implies $[f]$ is continuous. □

6.0.2 Identification of a Subset to a Point

If A is a subspace of a topological space X, we can define a relation $\sim$ on X by declaring

$$x \sim x \text{ for all } x \in X \text{ and } x \sim y \text{ for all } x, y \in A.$$

This is an equivalence relation on X. We say that the quotient space $X/\sim$ is obtained from X by **identifying A to a point**.

Example 6.1. Let I be the unit interval $[0,1]$ and $I/\sim$ be the quotient space obtained from I by identifying the two points $\{0,1\}$ to a point. Denote by S^1 the unit circle in the complex plane. The function $f : I \to S^1$, given by $f(x) = e^{2\pi i x}$, assumes the same value at 0 and 1, and so induces a function $\overline{f} : I/\sim \to S^1$. Since f is continuous, $\overline{f}$ is continuous. As the continuous image of a compact set I, the quotient $I/\sim$ is compact. Thus $\overline{f}$ is a continuous bijection from the compact space $I/\sim$ to the Hausdorff space S^1. Hence it is a homeomorphism.

6.1 Open Equivalence Relations

An equivalence relation $\sim$ on a topological space X is said to be **open** if the projection map $\pi : X \to X/\sim$ is open. In other words, the equivalence relation $\sim$ on X is open if and only if for every open set U in X, the set

$$\pi^{-1}(\pi(U)) = \bigcup_{x \in U} [x]$$

of all points equivalent to some point of U is open.

Example 6.2. Let $\sim$ be the equivalence relation on the real line $\mathbb{R}$ that identifies the two points 1 and -1 and let $\pi : \mathbb{R} \to \mathbb{R}/\sim$ be the projection map. Then π is not an open map. Indeed, let V be the open interval $(-2,0)$ in $\mathbb{R}$. Then

$$\pi^{-1}(\pi(V)) = (-2,0) \cup \{1\},$$

which is not open in $\mathbb{R}$.

Given an equivalence relation $\sim$ on X, let R be the subset of $X \times X$ that defines the relation

$$R = \{(x,y) \in X \times X \mid x \sim y\}.$$

We call R the **graph** of the equivalence relation $\sim$.

Theorem 6.1. *Suppose $\sim$ is an open equivalence relation on a topological space X. Then the quotient space $X/\sim$ is Hausdorff if and only if the graph R of $\sim$ is closed in $X \times X$.*

Proof. There is a sequence of equivalent statements: R is closed in $X \times X$ iff $(X \times X)\backslash R$ is open in $X \times X$ iff for every $(x,y) \in (X \times X)\backslash R$, there is a basic open set $U \times V$ containing (x,y) such that $(U \times V) \cap R = \emptyset$ iff for every pair $x \not\sim y$ in X, there exist neighborhoods U of x and V of y in X such that no element of U is equivalent to an element of V iff for any two points $[x] \neq [y]$ in $X/\sim$, there exist neighborhoods U of x and V of y in X such that $\pi(U) \cap \pi(V) = \emptyset$ in $X/\sim$.

We now show that this last statement is equivalent to $X/\sim$ being Hausdorff. Since $\sim$ is an open equivalence relation, $\pi(U)$ and $\pi(V)$ are disjoint open sets in $X/\sim$ containing $[x]$ and $[y]$ respectively, so $X/\sim$ is Hausdorff. Conversely, suppose $X/\sim$ is Hausdorff. Let $[x] \neq [y]$ in $X/\sim$. Then there exist disjoint open sets A and B in $X/\sim$ such that $[x] \in A$ and $[y] \in B$. By the surjectivity of π, we have $A = \pi(\pi^{-1}A)$ and $B = \pi(\pi^{-1}B)$. Let $U = \pi^{-1}A$ and $V = \pi^{-1}B$. Then $x \in U$, $y \in V$, and $A = \pi(U)$ and $B = \pi(V)$ are disjoint open sets in $X/\sim$. $\square$

Theorem 6.2. *Let $\sim$ be an open equivalence relation on a topological space X. If $\mathcal{B} = \{B_\alpha\}$ is a basis for X, then its image $\{\pi(B_\alpha)\}$ under π is a basis for $X/\sim$.*

Proof. Since π is an open map, $\{\pi(B(\alpha)\}$ is a collection of open sets in $X/\sim$. Let W be an open set in $X/\sim$ and $[x] \in W$. Then $x \in \pi^{-1}(W)$. Since $\pi^{-1}(W)$ is open, there is a basic open set $B \in \mathcal{B}$ such that $x \in B \subset \pi^{-1}(W)$. Then $[x] = \pi(x) \in \pi(B) \subset W$, which proves that $\{\pi(B_\alpha)\}$ is a basis for $S/\sim$. $\square$

Corollary 3. *If $\sim$ is an open equivalence relation on a second-countable space X, then the quotient space is second-countable.*

6.2 Quotients by Group Actions

Many important manifolds are constructed as quotients by actions of groups on other manifolds, and this often provides a useful way to understand spaces that may have been constructed by other means. As a basic example, the Klein bottle will be defined as a quotient of $S^1 \times S^1$ by the action of a group of order 2. The circle as defined concretely in $\mathbb{R}^2$ is isomorphic to the quotient of $\mathbb{R}$ by additive translation by $\mathbb{Z}$.

Definition 6.1. Let X be a topological space and G a discrete group. A right action of G on X is **continuous** if for each $g \in G$ the action map $X \to X$ defined by $x \mapsto x.g$ is continuous (and hence a homeomorphism, as the action of g^{-1} gives an inverse). The action is **free** if for each $x \in X$ the stabilizer subgroup $\{g \in G \mid x.g = x\}$ is the trivial subgroup (in other words, $x.g = x$ implies $g = 1$). The action is **properly discontinuous** when it is continuous for the discrete topology on G and each $x \in X$ admits an open neighborhood U_x so that the G-translate $U_x.g$ meets U_x for only finitely many $g \in G$.

Proposition 6.3. *A right action of G on X is continuous if $\pi : X \times G \to X$ is continuous.*

Remark 13. Here, G has the discrete topology.

Proof. Suppose we have a right action of G on X which is continuous. Let U be an open set in X. For each $g \in G$, let $U_g := g^{-1}(U)$. Then
$$\pi^{-1}(U) = \bigcup_{g \in G} U_g \times \{g\},$$
which is open. Conversely, suppose π is continuous and let $g \in G$. Let U be open in X and set $U_g := g^{-1}(U)$. Then
$$\pi^{-1}(U) \cap X \times \{g\} = U_g \times \{g\},$$
which shows that g is continuous since $\pi^{-1}(U)$ and $X \times \{g\}$ are open in $X \times G$. $\square$

Example 6.3. Suppose that X is a locally Hausdorff space, and that G acts on X on the right via a properly discontinuous action. For each $x \in X$, we get an open subset U_x such that U_x meets $U_x.g$ for only finitely many $g \in G$. This property is unaffected by replacing U_x with a smaller open subset around x, so by the locally Hausdorff property we can assume that U_x is Hausdorff. The key is that we can do better: there exists an open set $U'_x \subseteq U_x$ such that U'_x meets $U'_x.g$ if and only if $x = x.g$. Thus, if the action is also free then U'_x is disjoint from $U'_x.g$ for all $g \in G$ with $g \neq 1$.

To find U'_x, let $g_1, \ldots, g_n \in G$ be an enumeration of the finite set of elements $g \in G$ such that U_x meets $U_x.g$. For any open subset $U \subseteq U_x$ we can only have $U \cap U.g \neq \emptyset$ for g equal to one of the g_i's, so it suffices to show that for each i with $x.g_i \in U_x \backslash \{x\}$ there is an open subset $U_i \subseteq U_x$ such that $U_i \cap (U_i).g_i = \emptyset$ (and then we may take U'_x to be the intersection of the U_i's over the finitely many i such that $x.g_i \neq x$). By the Hausdorff property of U_x, when $x.g_i \in U_x \backslash \{x\}$ there exist disjoint opens $V_i, V'_i \subseteq U_x$ around x and $x.g_i$ respectively. By continuity of the action on X by $g_i \in G$ there is an open $W_i \subseteq X$ around x such that $(W_i).g_i \subseteq V'_i$. Thus $U_i = W_i \cap V_i$ is disjoint from V'_i yet satisfies $(U_i).g_i \subseteq V'_i$, so $U_i \cap (U_i).g_i = \emptyset$. This completes the construction of U'_x.

The interest in free and properly discontinuous actions is that for such actions in the locally Hausdorff case we may find an open U_x around each $x \in X$ such that U_x is disjoint from $U_x.g$ whenever $g \neq 1$. Thus, for such actions we may say that in X/G we are identifying points in the same G-orbit with this identification process not "crushing" the space X by identifying points in X that are arbitrarily close to each other. An example where things go horribly wrong is the action of $G = \mathbb{Q}$ on $\mathbb{R}$ via additive translations. This is a continuous action, but the quotient $\mathbb{R}/\mathbb{Q}$ is very bad: any two $\mathbb{Q}$-orbits in $\mathbb{R}$ contain arbitrarily close points!

Here are some examples of free and properly discontinuous actions.

Example 6.4. The antipodal map on S^n, given by $(a_1, \ldots, a_{n+1}) \mapsto (-a_1, \ldots, -a_{n+1})$, viewed as an action of the integers mod 2 is free and properly discontinuous: freeness is clear, as is continuity, and for any $x \in S^n$ the points near x all have their antipodes far away!

Example 6.5. Consider the curve $X := \mathbf{V}(x^3 + y^3 + z^3 - 1) \subseteq \mathbb{C}^3$. Then the action $(a_1, a_2, a_3) \mapsto (\zeta_3 a_1, \zeta_3 a_2, \zeta_3)$, viewed as an action of the integers mod 3 is free and properly discontinuous.

Example 6.6. Let $X = S^1 \times S^1$ be a product of two circles, where the circle
$$S^1 = \{z \in \mathbb{C} \mid |z| = 1\}$$

is viewed as a topological group (using multiplication in $\mathbb{C}$, so both the group law and inversion $z \mapsto 1/z = \bar{z}$ on S^1 are continuous). The visibly continuous map $(z, w) \mapsto (1/z, -w) = (\bar{z}, -w)$ reflects through the x-axis in the first circle and rotates 180-degree in the second circle, and it is its own inverse. Thus, this give an action by the order-2 group G of integers mod 2. The associated quotient X/G will be called the (set-theoretic) **Klein bottle**.

Theorem 6.3. *Let X be a locally Hausdorff topological space with a free and properly discontinuous action by a group G. There is a unique topology on X/G such that the quotient map $\pi : X \to X/G$ is a continuous map that is a local homeomorphism (i.e. each $x \in X$ admits a neighborhood mapping homeomorphically onto an open subset of X/G). Moreover, the quotient map is open.*

A subset $S \subseteq X/G$ is open if and only if its preimage in X is open, and if $U \subseteq X$ is an open set that is disjoint from $U.g$ for all nontrivial $g \in G$ then the map $U \to X/G$ is a homeomorphism onto its open image $\overline{U}$ and the natural map $U \times G \to \pi^{-1}(\overline{U})$ over $\overline{U}$ given by $(u, g) \mapsto u.g$ is a homeomorphism when G is given the discrete topology.

Remark 14. The topology in this theorem is called the **quotient topology**, and it is locally Hausdorff since $X \to X/G$ is a local homeomorphism.

Proof. Sketch: we show that π is an open map. Let $x \in X$ and pick U_x such that $U_x.g \cap U_x = \emptyset$ for all $g \in G \setminus \{1\}$. We first show that $\pi(U_x)$ is open. The inverse image of $\pi(U_x)$ under π is a disjoint union of open sets $\bigcup_{g \in G} U_x.g$. Therefore $\pi(U_x)$ is open. Now let U be any open subset of X. For each $x \in X$, choose U_x such that $U_x.g \cap U_x = \emptyset$ for all $g \in G \setminus \{1\}$ and $U_x \subset U$. Then

$$\pi(U) = \pi\left(\bigcup_{x \in U} U_x\right) = \bigcup_{x \in U} \pi(U_x)$$

implies $\pi(U)$ is open. $\square$

Example 6.7. (Möbius Strip) Choose $a > 0$. Let $X = (-a, a) \times S^1$, and let the group of order 2 act on it with the non-trivial element acting by $(t, w) \mapsto (-t, -w)$. This is easily checked to be a continuous action for the discrete topology of the group of order 2, and it is free and properly discontinuous. The quotient M_a is the **Möbius strip** of height $2a$.

To check that the Möbius strip M_a is Hausdorff, we use the quotient criterion: the set of points in $X \times X$ with the form $((t, w), (t', w'))$ with $(t', w') = (t, w)$ or $(t', w') = (-t, -w)$ is checked to be closed by using the sequential criterion in $X \times X$: suppose $(t_n, w_n) \sim (t'_n, w'_n)$ are sequences in $X \times X$ which converge (t, w) and (t', w') respectively. Then we need to show that $(t, w) \sim (t', w')$. Assume that $(t, w) \neq (t', w')$. Choose open neighborhoods U of (t, w) and U' of (t', w') respectively such that $U \cap U' = \emptyset$ and such that eventually $(t_n, w_n) \neq (t'_n, w'_n)$ (We can do this because they converge to different limits and our space $X \times X$ is Hausdorff). Thus, eventually we have $(t'_n, w'_n) = (-t_n, -w_n) \to (-t, -w)$.

7 Product Topology

Let Λ be a set and let $(X_\lambda, \tau_\lambda)$ be a topological space for each $\lambda \in \Lambda$. For each $\lambda \in \Lambda$, we denote by $\pi_\lambda \colon \prod_\lambda X_\lambda \to X_\lambda$ to be the λth **projection map** defined by

$$\pi_\lambda((x_\lambda)) = x_\lambda$$

for all $(x_\lambda) \in \prod_\lambda X_\lambda$. We define the **product topology** on $\prod_\lambda X_\lambda$, denoted $\prod_\lambda \tau_\lambda$, to be the topology generated by sets of the form

$$\{\pi_\lambda^{-1}(U_\lambda) \mid \lambda \in \Lambda \text{ and } U_\lambda \in \tau_\lambda\}.$$

In particular, the product is the *weakest* topology on $\prod_\lambda X_\lambda$ which makes all of the projection maps π_λ continuous. Recall that the topology on X generated by a subcollection $\mathcal{C} \subseteq \mathcal{P}(X)$ is obtained by adjoining X and $\emptyset$ to the entire collection as well as all adjoint all arbitrary unions of finite intersections of members of $\mathcal{C}$ to the entire collection. Note that for each $\mu \in \Lambda$ and $U_\mu \in \tau_\mu$ we have

$$\pi_\mu^{-1}(U_\mu) = U_\mu \times \prod_{\lambda \in \Lambda \setminus \{\mu\}} X_\lambda,$$

and for each distinct $\mu, \kappa \in \Lambda$ and $U_\mu \in \tau_\mu$ and $U_\kappa \in \tau_\kappa$, we have

$$\pi_\mu^{-1}(U_\mu) \cap \pi_\kappa^{-1}(U_\kappa) = U_\mu \times U_\kappa \times \prod_{\lambda \in \Lambda \setminus \{\mu, \kappa\}} X_\lambda.$$

In general, a basis of $\prod_\lambda \tau_\lambda$ consists of sets of the form

$$\prod_{\lambda_0 \in \Lambda_0} U_{\lambda_0} \times \prod_{\lambda \in \Lambda \setminus \Lambda_0} X_\lambda,$$

where Λ_0 is a finite subset of Λ and U_{λ_0} is an open subset of X_{λ_0} for each $\lambda_0 \in \Lambda_0$.

Proposition 7.1. *Let Λ be a set, let $(X_\lambda, \tau_\lambda)$ be a topological space for each $\lambda \in \Lambda$, let Y be a topological space, and let $f\colon Y \to \prod_\lambda X_\lambda$ be a function. Then f is continuous if and only if $\pi_\lambda \circ f\colon Y \to X_\lambda$ is continuous for each $\lambda \in \Lambda$.*

Proof. If f is continous, then each $\pi_\lambda \circ f$ is a composition of continuous functions and is hence continuous. Conversely, suppose that $\pi_\lambda \circ f$ is continuous for each $\lambda \in \Lambda$. To show that f is continous, it suffices to show that the premiage of a subbase element $\pi_\lambda^{-1}(U_\lambda)$ is open. But note that

$$f^{-1}(\pi_\lambda^{-1}(U_\lambda)) = (\pi_\lambda \circ f)^{-1}(U_\lambda)$$

is open since $\pi_\lambda \circ f$ is continuous. Thus f is continuous. □

8 Coproduct Topology

Let Λ be a set and let $(X_\lambda, \tau_\lambda)$ be a topological space for each $\lambda \in \Lambda$. For each $\lambda \in \Lambda$, we denote by $\pi_\lambda\colon \prod_\lambda X_\lambda \to X_\lambda$ to be the λth **projection map** defined by

$$\pi_\lambda((x_\lambda)) = x_\lambda$$

for all $(x_\lambda) \in \prod_\lambda X_\lambda$. We define the **product topology** on $\prod_\lambda X_\lambda$, denoted $\prod_\lambda \tau_\lambda$, to be the topology generated by sets of the form

$$\{\pi_\lambda^{-1}(U_\lambda) \mid \lambda \in \Lambda \text{ and } U_\lambda \in \tau_\lambda\}.$$

In particular, the product is the *weakest* topology on $\prod_\lambda X_\lambda$ which makes all of the projection maps π_λ continuous. Recall that the topology on X generated by a subcollection $\mathcal{C} \subseteq \mathcal{P}(X)$ is obtained by adjoining X and $\emptyset$ to the entire collection as well as all adjoint all arbitrary unions of finite intersections of members of $\mathcal{C}$ to the entire collection. Note that for each $\mu \in \Lambda$ and $U_\mu \in \tau_\mu$ we have

$$\pi_\mu^{-1}(U_\mu) = U_\mu \times \prod_{\lambda \in \Lambda \setminus \{\mu\}} X_\lambda,$$

and for each distinct $\mu, \kappa \in \Lambda$ and $U_\mu \in \tau_\mu$ and $U_\kappa \in \tau_\kappa$, we have

$$\pi_\mu^{-1}(U_\mu) \cap \pi_\kappa^{-1}(U_\kappa) = U_\mu \times U_\kappa \times \prod_{\lambda \in \Lambda \setminus \{\mu, \kappa\}} X_\lambda.$$

In general, a basis of $\prod_\lambda \tau_\lambda$ consists of sets of the form

$$\prod_{\lambda_0 \in \Lambda_0} U_{\lambda_0} \times \prod_{\lambda \in \Lambda \setminus \Lambda_0} X_\lambda,$$

where Λ_0 is a finite subset of Λ and U_{λ_0} is an open subset of X_{λ_0} for each $\lambda_0 \in \Lambda_0$.

Proposition 8.1. *Let Λ be a set, let $(X_\lambda, \tau_\lambda)$ be a topological space for each $\lambda \in \Lambda$, let Y be a topological space, and let $f\colon Y \to \prod_\lambda X_\lambda$ be a function. Then f is continuous if and only if $\pi_\lambda \circ f\colon Y \to X_\lambda$ is continuous for each $\lambda \in \Lambda$.*

Proof. If f is continous, then each $\pi_\lambda \circ f$ is a composition of continuous functions and is hence continuous. Conversely, suppose that $\pi_\lambda \circ f$ is continuous for each $\lambda \in \Lambda$. To show that f is continous, it suffices to show that the premiage of a subbase element $\pi_\lambda^{-1}(U_\lambda)$ is open. But note that

$$f^{-1}(\pi_\lambda^{-1}(U_\lambda)) = (\pi_\lambda \circ f)^{-1}(U_\lambda)$$

is open since $\pi_\lambda \circ f$ is continuous. Thus f is continuous. □

Part II
Linear Analysis

9 Inner-Product Spaces

Definition 9.1. Let V be a vector space over $\mathbb{C}$. A map $\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{C}$ is called an **inner-product on** V if it satisfies the following properties:

1. Linearity in the first argument: $\langle x, z \rangle + \langle y, z \rangle = \langle x + y, z \rangle$ and $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$ for all $x, y, z \in V$ and $\lambda \in \mathbb{C}$.
2. Conjugate symmetric: $\langle x, y \rangle = \overline{\langle y, x \rangle}$ for all $x, y \in V$.
3. Positive definite: $\langle x, x \rangle > 0$ for all nonzero $x \in V$.

A vector space equipped with an inner-product is called an **inner-product space**. We often write $\mathcal{V}$ to denote an inner-product space.

Proposition 9.1. *Let $\mathcal{V}$ be an inner-product space. Then*

1. $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$ *for all* $x, y, z \in \mathcal{V}$;
2. $\langle x, \lambda y \rangle = \overline{\lambda} \langle x, y \rangle$ *for all* $x, y \in \mathcal{V}$ *and* $\lambda \in \mathbb{C}$
3. $\langle x, 0 \rangle = 0 = \langle 0, x \rangle$ *for all* $x \in \mathcal{V}$;
4. *Let* $x, y \in \mathcal{V}$. *If* $\langle x, z \rangle = \langle y, z \rangle$ *for all* $z \in \mathcal{V}$, *then* $x = y$.

Proof.

1. Let $x, y, z \in \mathcal{V}$. Then
$$\begin{aligned} \langle x, y + z \rangle &= \overline{\langle y + z, x \rangle} \\ &= \overline{\langle y, x \rangle + \langle z, x \rangle} \\ &= \overline{\langle y, x \rangle} + \overline{\langle z, x \rangle} \\ &= \langle x, y \rangle + \langle x, z \rangle. \end{aligned}$$
2. Let $x, y \in \mathcal{V}$ and $\lambda \in \mathbb{C}$. Then
$$\begin{aligned} \langle x, \lambda y \rangle &= \overline{\langle \lambda y, x \rangle} \\ &= \overline{\lambda \langle y, x \rangle} \\ &= \overline{\lambda}\, \overline{\langle y, x \rangle} \\ &= \overline{\lambda} \langle x, y \rangle. \end{aligned}$$
3. Let $x \in \mathcal{V}$. Then
$$\begin{aligned} \langle x, 0 \rangle &= \langle x, 0 + 0 \rangle \\ &= \langle x, 0 \rangle + \langle x, 0 \rangle \end{aligned}$$
implies $\langle x, 0 \rangle = 0$. A similar argument gives $\langle 0, x \rangle = 0$.
4. Assuming $\langle x, z \rangle = \langle y, z \rangle$ for all $z \in \mathcal{V}$, then we have $\langle x - y, z \rangle = 0$ for all $z \in \mathcal{V}$. In particular, setting $z = x - y$, we have $\langle x - y, x - y \rangle = 0$. Since the inner-product is positive definite, we must have $x - y = 0$, and hence $x = y$.

$\square$

9.1 Examples of Inner-Product Spaces

If we are given a complex vector space V, then we can **give V the structure of an inner-product space** by equipping V with an inner-product

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}.$$

In the following examples, we give many familiar complex vector spaces the structure of an inner-product space.

9.1.1 Giving $\mathbb{C}^n$ the structure of an inner-product space

Proposition 9.2. *Let $\langle \cdot, \cdot \rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ be given by*

$$\langle x, y \rangle = \langle (x_1, \dots, x_n), (y_1, \dots, y_n) \rangle = \sum_{i=1}^{n} x_i \overline{y}_i.$$

for all $x, y \in \mathbb{C}^n$. Then the pair $(\mathbb{C}^n, \langle \cdot, \cdot \rangle)$ forms an inner-product space.

Proof. For linearity in the first argument follows from linearity, let $x, y, z \in \mathbb{C}^n$. Then

$$\begin{aligned}
\langle x + y, z \rangle &= \sum_{i=1}^{n} (x_i + y_i) \overline{z}_i \\
&= \sum_{i=1}^{n} x_i \overline{z}_i + \sum_{i=1}^{n} y_i \overline{z}_i \\
&= \langle x, z \rangle + \langle y, z \rangle.
\end{aligned}$$

For conjugate symmetry of $\langle \cdot, \cdot \rangle$, let $x, y \in \mathbb{C}^n$. Then

$$\begin{aligned}
\langle x, y \rangle &= \sum_{i=1}^{n} x_i \overline{y}_i \\
&= \sum_{i=1}^{n} \overline{\overline{x_i \overline{y}_i}} \\
&= \sum_{i=1}^{n} \overline{y_i \overline{x}_i} \\
&= \overline{\langle y, x \rangle}.
\end{aligned}$$

For positive-definiteness of $\langle \cdot, \cdot \rangle$, let $x \in \mathbb{C}^n$. Then

$$\begin{aligned}
\langle x, x \rangle &= \sum_{i=1}^{n} x_i \overline{x}_i \\
&= \sum_{i=1}^{n} |x_i|^2.
\end{aligned}$$

is a sum of its components absolute squared. This implies positive-definiteness. □

9.1.2 Giving $\mathbf{M}_{m \times n}(\mathbb{C})$ the structure of an inner-product space

Since $\mathrm{M}_{m \times n}(\mathbb{C}) \cong \mathbb{C}^{mn}$, we can give $\mathrm{M}_{m \times n}(\mathbb{C})$ the structure of an inner-product space by equipping it with the inner-product described in the previous example. In more detail, let E_{ij} be the standard matrix in $\mathrm{M}_{m \times n}(\mathbb{C})$ whose entry in the (i, j)-th component is 1 and whose entry everywhere else is 0, let e_k be the standard basis vector in $\mathbb{C}^{mn}$ whose entry in the k-th compenent is 1 and whose entry everywhere else is 0, and let $\varphi : \mathrm{M}_{m \times n}(\mathbb{C}) \to \mathbb{C}^{mn}$ be the isomorphism such that

$$\varphi(E_{ij}) = e_{n(i-1)+j}$$

for all $E_{ij} \in \mathrm{M}_{m\times n}(K)$ where $1 \le i \le m$ and $1 \le j \le n$. So under φ, we have

$$A := \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \mapsto \begin{pmatrix} a_{11} \\ \vdots \\ a_{1n} \\ \vdots \\ a_{m1} \\ \vdots \\ a_{mn} \end{pmatrix} := a$$

Using this isomorphism, we give $\mathrm{M}_{m\times n}(\mathbb{C})$ the structure of an inner-product space by defining

$$\langle \cdot, \cdot \rangle \colon \mathrm{M}_{m\times n}(\mathbb{C}) \times \mathrm{M}_{m\times n}(\mathbb{C}) \to \mathbb{C}$$

by the formula

$$\langle A, B \rangle := \mathrm{Tr}(AB^*) = \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij}\overline{b}_{ij} = \langle a, b \rangle$$

where

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \quad , B = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix}, \quad \text{and} \quad B^* = \begin{pmatrix} \overline{b}_{11} & \cdots & \overline{b}_{m1} \\ \vdots & \ddots & \vdots \\ \overline{b}_{1n} & \cdots & \overline{b}_{nm} \end{pmatrix}.$$

9.1.3 Giving $\ell^2(\mathbb{N})$ the structure of an inner-product space

Lemma 9.1. *Let a and b be nonnegative real numbers. Then we have*

$$ab \le \frac{1}{2}(a^2 + b^2) \tag{5}$$

with equality if and only if $a = b$.

Proposition 9.3. *Let $\ell^2(\mathbb{N})$ be the set of all sequence (x_n) in $\mathbb{C}$ such that*

$$\sum_{n=1}^{\infty} |x_n|^2 < \infty$$

and let $\langle \cdot, \cdot \rangle \colon \ell^2(\mathbb{N}) \times \ell^2(\mathbb{N}) \to \mathbb{C}$ be given by

$$\langle (x_n), (y_n) \rangle = \sum_{n=1}^{\infty} x_n \overline{y}_n.$$

for all $(x_n), (y_n) \in \ell^2(\mathbb{N})$. Then the pair $(\ell^2(\mathbb{N}), \langle \cdot, \cdot \rangle)$ forms an inner-product space.

Proof. We first need to show that $\ell^2(\mathbb{N})$ is indeed a vector space. In fact, we will show that $\ell^2(\mathbb{N})$ is a subspace of $\mathbb{C}^N$, the set of all sequences in $\mathbb{C}$. Let $(x_n), (y_n) \in \ell^2(\mathbb{N})$ and $\lambda \in \mathbb{C}$. Then Lemma (14.1) implies

$$\begin{aligned} \sum_{n=1}^{\infty} |\lambda x_n + y_n|^2 &\le \sum_{n=1}^{\infty} |\lambda x_n|^2 + \sum_{n=1}^{\infty} |y_n|^2 + \sum_{n=1}^{\infty} 2|\lambda x_n||y_n| \\ &\le \lambda^2 \sum_{n=1}^{\infty} |x_n|^2 + \sum_{n=1}^{\infty} |y_n|^2 + \lambda^2 \sum_{n=1}^{\infty} |x_n|^2 + |y_n|^2 \\ &< \infty. \end{aligned}$$

Therefore $(\lambda x_n + y_n) \in \ell^2(\mathbb{N})$, which implies $\ell^2(\mathbb{N})$ is a subspace of $\mathbb{C}^{\mathbb{N}}$.

Next, let us show that the inner product converges, and hence is defined everywhere. Let $(x_n), (y_n) \in \ell^2(\mathbb{N})$. Then it follows from Lemma (14.1) that

$$
\begin{aligned}
\sum_{n=1}^{\infty} |x_n \overline{y}_n| &= \sum_{n=1}^{\infty} |x_n||y_n| \\
&\leq \sum_{n=1}^{\infty} \frac{|x_n|^2 + |y_n|^2}{2} \\
&= \frac{1}{2}\sum_{n=1}^{\infty} |x_n|^2 + \frac{1}{2}\sum_{n=1}^{\infty} |y_n|^2 \\
&< \infty.
\end{aligned}
$$

Therefore $\sum_{n=1}^{\infty} x_n \overline{y}_n$ is absolutely convergent, which implies it is convergent. (We can't use Cauchy-Schwarz here since we haven't yet shown that $\langle \cdot, \cdot \rangle$ is in fact an inner-product).

Finally, let us shows that $\langle \cdot, \cdot \rangle$ is an inner-product. Linearity in the first argument follows from distrubitivity of multiplication and linearity of taking infinite sums. For conjugate symmetry, let $(x_n), (y_n) \in \ell^2(\mathbb{N})$. Then

$$
\begin{aligned}
\langle (x_n), (y_n) \rangle &= \sum_{n=1}^{\infty} x_n \overline{y}_n \\
&= \lim_{N \to \infty} \sum_{n=1}^{N} x_n \overline{y}_n \\
&= \overline{\overline{\lim_{N \to \infty} \sum_{n=1}^{N} x_n \overline{y}_n}} \\
&= \overline{\lim_{N \to \infty} \sum_{n=1}^{N} \overline{x_n \overline{y}_n}} \\
&= \overline{\lim_{N \to \infty} \sum_{n=1}^{N} y_n \overline{x}_n} \\
&= \overline{\sum_{n=1}^{\infty} y_n \overline{x}_n} \\
&= \overline{\langle (y_n), (x_n) \rangle},
\end{aligned}
$$

where we were allowed to bring the conjugate inside the limit since the conjugate function is continuous on $\mathbb{C}$. For positive-definiteness, let $(x_n) \in \ell^2(\mathbb{N})$. Then

$$
\begin{aligned}
\langle (x_n), (x_n) \rangle &= \sum_{n=1}^{\infty} x_n \overline{x}_n \\
&= \sum_{n=1}^{\infty} |x_n|^2 \\
&\geq 0.
\end{aligned}
$$

If $\sum_{n=1}^{\infty} |x_n|^2 = 0$, then clearly we must have $x_n = 0$ for all n. □

9.1.4 Giving $C[a,b]$ the structure of an inner-product space

Proposition 9.4. *Let $C[a,b]$ be the space of all continuous functions defined on the closed interval $[a,b]$ and let $\langle \cdot, \cdot \rangle \colon C[a,b] \times C[a,b] \to \mathbb{C}$ be given by*

$$
\langle f, g \rangle = \int_a^b f(x)\overline{g(x)}dx
$$

for all $f, g \in C[a,b]$. Then the pair $(C[a,b], \langle \cdot, \cdot \rangle)$ forms an inner-product space.

Proof. Exercise. □

9.2 Norm Induced by Inner-Product

Definition 9.2. The **norm** of $x \in \mathcal{V}$, denoted $\|x\|$, is defined by

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

Lemma 9.2. *(Pythagorean Theorem) Let x and y be nonzero vectors in $\mathcal{V}$ such that $\langle x, y \rangle = 0$ (we call such vectors **orthogonal** to one another). Then*

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2.$$

Proof. We have

$$\begin{aligned} \|x+y\|^2 &= \langle x+y, x+y \rangle \\ &= \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle \\ &= \langle x, x \rangle + \langle y, y \rangle \\ &= \|x\|^2 + \|y\|^2 \end{aligned}$$

□

9.2.1 Properties of Norm

Proposition 9.5. *If $x, y \in \mathcal{V}$ and $\lambda \in \mathbb{C}$, then*

1. *Positive-Definiteness: $\|x\| \geq 0$ with equality if and only if $x = 0$;*
2. *Absolutely Homogeneous: $\|\lambda x\| = |\lambda| \|x\|$;*
3. *Cauchy-Schwarz: $|\langle x, y \rangle| \leq \|x\| \|y\|$ with equality if and only if x and y are linearly dependent.*
4. *Subadditivity: $\|x + y\| \leq \|x\| + \|y\|$*

Proof.

1. This follows from positive-definiteness of $\langle \cdot, \cdot \rangle$.
2. We have

$$\begin{aligned} \|\lambda x\| &= \sqrt{\langle \lambda x, \lambda x \rangle} \\ &= \sqrt{\lambda \overline{\lambda} \langle x, x \rangle} \\ &= \sqrt{|\lambda|^2 \langle x, x \rangle} \\ &= |\lambda| \sqrt{\langle x, x \rangle} \\ &= |\lambda| \|x\|. \end{aligned}$$

3. We may assume that both x and y are nonzero, since it is trivial in this case. Let

$$z = x - \mathrm{pr}_y(x) = x - \frac{\langle x, y \rangle}{\langle y, y \rangle} y.$$

Then by linearity of the inner product in the first argument, one has

$$\begin{aligned} \langle z, y \rangle &= \left\langle x - \frac{\langle x, y \rangle}{\langle y, y \rangle} y, y \right\rangle \\ &= \langle x, y \rangle - \frac{\langle x, y \rangle}{\langle y, y \rangle} \langle y, y \rangle \\ &= 0. \end{aligned}$$

Therefore z is a vector orthogonal to the vector y. We can thus apply the Pythagorean theorem to

$$x = \frac{\langle x, y \rangle}{\langle y, y \rangle} y + z$$

which gives

$$\begin{aligned}
\|x\|^2 &= \left|\frac{\langle x,y\rangle}{\langle y,y\rangle}\right|^2 \|y\|^2 + \|z\|^2 \\
&= \frac{|\langle x,y\rangle|^{\,2}}{(\|y\|^2)^2}\|y\|^2 + \|z\|^2 \\
&= \frac{|\langle x,y\rangle|^{\,2}}{\|y\|^2} + \|z\|^2 \\
&\geq \frac{|\langle x,y\rangle|^2}{\|y\|^2},
\end{aligned}$$

and after multiplication by $\|y\|^2$ and taking square root, we get the Cauchy-Schwarz inequality. Moreover, if the relation $\geq$ in the above expression is actually an equality, then $\|z\|^2 = 0$ and hence $z = 0$; the definition of z then establishes a relation of linear dependence between x and y. On the other hand, if x and y are linearly dependent, then there exists $\lambda \in \mathbb{C}$ such that $x = \lambda y$. Then

$$\begin{aligned}
|\langle x,y\rangle| &= |\langle \lambda y,y\rangle| \\
&= |\lambda||\langle y,y\rangle| \\
&= |\lambda|\|y\|^2 \\
&= \|\lambda y\|\|y\| \\
&= \|x\|\|y\|.
\end{aligned}$$

4. By the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\|x+y\|^2 &= \langle x+y, x+y\rangle \\
&= \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle \\
&= \|x\|^2 + 2\mathrm{Re}(\langle x,y\rangle) + \|y\|^2 \\
&\leq \|x\|^2 + 2|\langle x,y\rangle| + \|y\|^2 \\
&\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 \\
&= (\|x\| + \|y\|)^2
\end{aligned}$$

now we take square roots on both sides to get the desired result.

□

Proposition 9.6. *(Parallelogram Identity) Let $x, y \in \mathcal{V}$. Then*

$$\|x-y\|^2 + \|x+y\|^2 = 2\|x\|^2 + 2\|y\|^2 \tag{6}$$

Proof. We have

$$\begin{aligned}
\|x+y\|^2 &= \langle x+y, x+y\rangle \\
&= \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle \\
&= \|x\|^2 + 2\mathrm{Re}(\langle x,y\rangle) + \|y\|^2.
\end{aligned}$$

and

$$\begin{aligned}
\|x-y\|^2 &= \langle x-y, x-y\rangle \\
&= \langle x,x\rangle + \langle x,-y\rangle + \langle -y,x\rangle + \langle -y,-y\rangle \\
&= \|x\|^2 - 2\mathrm{Re}(\langle x,y\rangle) + \|y\|^2.
\end{aligned}$$

Adding these together gives us our desired result. □

Proposition 9.7. *(Polarization Identity) Let $x, y \in \mathcal{V}$. Then*

$$4\langle x,y\rangle = \|x+y\|^2 + i\|x+iy\|^2 - \|x-y\|^2 - i\|x-iy\|^2$$

Proof. We have

$$\begin{aligned}\|x+y\|^2 &= \langle x+y, x+y\rangle \\ &= \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle,\end{aligned}$$

and

$$\begin{aligned}i\|x+iy\|^2 &= i\langle x+iy, x+iy\rangle \\ &= i\langle x,x\rangle + i\langle x,iy\rangle + i\langle iy,x\rangle + i\langle iy,iy\rangle \\ &= i\langle x,x\rangle + \langle x,y\rangle - \langle y,x\rangle + i\langle y,y\rangle,\end{aligned}$$

and

$$\begin{aligned}-\|x-y\|^2 &= -\langle x-y, x-y\rangle \\ &= -\langle x,x\rangle - \langle x,-y\rangle - \langle -y,x\rangle - \langle -y,-y\rangle \\ &= -\langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle - \langle y,y\rangle,\end{aligned}$$

and

$$\begin{aligned}-i\|x-iy\|^2 &= -i\langle x-iy, x-iy\rangle \\ &= -i\langle x,x\rangle - i\langle x,-iy\rangle - i\langle -iy,x\rangle - i\langle -iy,-iy\rangle \\ &= -i\langle x,x\rangle + \langle x,y\rangle - \langle y,x\rangle - i\langle y,y\rangle.\end{aligned}$$

Adding these together gives us our desired result. □

9.2.2 Normed Vector Spaces

Definition 9.3. Let V be a $\mathbb{C}$-vector space. A **norm** on V is a nonnegative-valued scalar function $\|\cdot\|\colon V \to [0,\infty)$ such that for all $\lambda \in \mathbb{C}$ and $x, y \in V$, we have

1. (Subadditivity) $\|x+y\| \leq \|x\| + \|x\|$,
2. (Absolutely Homogeneous) $\|\lambda x\| = |\lambda|\|x\|$,
3. (Positive-Definite) $\|x\| = 0$ if and only if $x = 0$.

We call the pair $(V, \|\cdot\|)$ a **normed vector space**.

Proposition (9.5) implies $\mathcal{V}$ is a normed vector space. This justifies our choice of the word "norm" in Definition (9.2). Proposition (9.5) also tells us that $\mathcal{V}$ satisfies an extra property which is not satisfied by other normed vector spaces, namely the Cauchy-Schwarz inequality. In fact, it turns out that inner-product spaces are just normed vector spaces which satisfy the parallelogram law.

9.2.3 Metric Induced By Norm

Definition 9.4. A **metric** on a set X is a function $d\colon X \times X \to \mathbb{R}$ which satisfies the following three properties:

1. (Identity of Indiscernibles) $d(x,y) = 0$ if and only if $x = y$ for all $x, y \in X$,
2. (Symmetric) $d(x,y) = d(y,x)$ for all $x, y \in X$,
3. (Triangle Inequality) $d(x,z) \leq d(x,y) + d(y,z)$ for all $x, y, z \in X$.

A set X together with a choice of a metric d is called a **metric space** and is denoted (X, d), or just denoted X if the metric is understood from context.

Remark 15. Given the three axioms above, we also have positive-definiteness: $d(x,y) \geq 0$ with equality if and only if $x = y$ for all $x, y \in X$. Indeed,

$$\begin{aligned}0 &= d(x,x) \\ &\leq d(x,y) + d(y,x) \\ &= d(x,y) + d(x,y) \\ &= 2d(x,y).\end{aligned}$$

This implies $d(x,y) \geq 0$.

Proposition 9.8. *Let $(V, \|\cdot\|)$ be a normed vector space. Define $d\colon V \times V \to \mathbb{R}$ by $d(x,y) = \|x-y\|$ for all $(x,y) \in V \times V$. Then (V,d) is a metric space.*

Proof. Let us first check that d satisfies the identity of indiscernibles property. Since $\|\cdot\|$ is positive-definite, $d(x,y) = 0$ implies $\|x-y\| = 0$ which implies $x = y$. On the other hand, suppose $x = y$. Then since $\|\cdot\|$ is absolutely homogeneous, we have $\|0\| = |0|\|0\| = 0$, and so $d(x,y) = \|0\| = 0$.

Next we check that d is symmetric. For all $(x,y) \in V \times V$, we have

$$\begin{aligned} d(x,y) &= \|x-y\| \\ &= \|-(y-x)\| \\ &= |-1|\|y-x\| \\ &= \|y-x\| \\ &= d(y,x). \end{aligned}$$

Finally, triangle inequality for d follows from subadditivity of $\|\cdot\|$. Indeed, for all $x,y,z \in V$, we have

$$\begin{aligned} d(x,y) + d(y,z) &= \|x-y\| + \|y-z\| \\ &\geq \|x-z\| \\ &= d(x,z). \end{aligned}$$

□

The distance between points $x,y \in \mathcal{V}$ is measured using the metric induced by the norm:

$$d(x,y) := \|x-y\|.$$

Definition 9.5. A sequence (x_n) in $\mathcal{V}$ is said to converge to a point $x \in \mathcal{V}$ if for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$n \geq N \text{ implies } \|x_n - x\| < \varepsilon.$$

In this case we write $x_n \to x$ as $n \to \infty$ or more simply $x_n \to x$. We also write

$$\lim_{n\to\infty} x_n = x.$$

Proposition 9.9. *Let (x_n) and (y_n) be two sequences in $\mathcal{V}$ and let (λ_n) be a sequence in $\mathbb{C}$. Then the following statements hold:*

1. *There exists $M \geq 0$ such that $\|x_n\| \leq M$ for all $n \in \mathbb{N}$.*
2. *If $x_n \to x$ and $y_n \to y$, then $x_n + y_n \to x + y$.*
3. *If (λ_n) is a sequence in $\mathbb{C}$ and $\lambda_n \to \lambda$, then $\lambda_n x_n \to \lambda x$.*
4. *If $x_n \to x$ and $y_n \to y$, then $\langle x_n, y_n\rangle \to \langle x, y\rangle$. In particular, $\|x_n\| \to \|x\|$.*

Proof.

1. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies $\|x_n - x\| \leq 1$. Now set M to be
$$M = \max\{\|x_1\|, \ldots, \|x_{N-1}\|, \|x\| + 1\}$$
2. Let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies $\|x_n - x\| < \varepsilon/2$ and $\|y_n - y\| < \varepsilon/2$. Then $n \geq N$ implies
$$\begin{aligned} \|x_n + y_n - (x+y)\| &\leq \|x_n - x\| + \|y_n - y\| \\ &< \varepsilon/2 + \varepsilon/2 \\ &= \varepsilon. \end{aligned}$$

3. Since $x_n \to x$, there exists $M \geq 0$ such that $\|x_n\| \leq M$ for all $n \in \mathbb{N}$. Choose such an M and let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies $\|x_n - x\| < \varepsilon/2|\lambda|$ and $|\lambda_n - \lambda| < \varepsilon/2M$. Then $n \geq N$ implies

$$\begin{aligned}
\|\lambda_n x_n - \lambda x\| &= \|\lambda_n x_n - \lambda x_n + \lambda x_n - \lambda x\| \\
&\leq \|\lambda_n x_n - \lambda x_n\| + \|\lambda x_n - \lambda x\| \\
&\leq \|(\lambda_n - \lambda) x_n\| + \|\lambda(x_n - x)\| \\
&= |\lambda_n - \lambda| \|x_n\| + |\lambda| \|x_n - x\| \\
&\leq |\lambda_n - \lambda| M + |\lambda| \|x_n - x\| \\
&< \varepsilon/2 + \varepsilon/2 \\
&= \varepsilon.
\end{aligned}$$

4. Since $y_n \to y$, there exists $M \geq 0$ such that $\|y_n\| \leq M$ for all $n \in \mathbb{N}$. Choose such an M and let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies $\|x_n - x\| < \varepsilon/2M$ and $\|y_n - y\| < \varepsilon/2\|x\|$. Then $n \geq N$ implies

$$\begin{aligned}
|\langle x_n, y_n \rangle - \langle x, y \rangle| &= |\langle x_n, y_n \rangle - \langle x, y_n \rangle + \langle x, y_n \rangle - \langle x, y \rangle| \\
&\leq |\langle x_n, y_n \rangle - \langle x, y_n \rangle| + |\langle x, y_n \rangle - \langle x, y \rangle| \\
&= |\langle x_n - x, y_n \rangle| + |\langle x, y_n - y \rangle| \\
&\leq \|x_n - x\| \|y_n\| + \|x\| \|y_n - y\| \\
&\leq \|x_n - x\| M + \|x\| \|y_n - y\| \\
&< \varepsilon/2 + \varepsilon/2 \\
&= \varepsilon.
\end{aligned}$$

   To see that $\|x_n\| \to \|x\|$, we just set $y_n = x_n$. Then

$$\begin{aligned}
\|x_n\| &= \sqrt{\langle x_n, x_n \rangle} \\
&\to \sqrt{\langle x, x \rangle} \\
&= \|x\|,
\end{aligned}$$

   where we were allowed to take limits inside the square root function since the square root function is continuous on $\mathbb{R}_{\geq 0}$.

□

Definition 9.6. Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be two norms on a vector space V. We say $\|\cdot\|_1$ is **stronger** than $\|\cdot\|_2$ (or $\|\cdot\|_2$ is **weaker** than $\|\cdot\|_1$) if there exists a constant $C > 0$ such that

$$\|x\|_2 \leq C \|x\|_1$$

for all $x \in V$.

Remark 16. Observe if a sequence (x_n) in V converges to x in the metric spaced induced by the $\|\cdot\|_1$ norm, then it also converges in the metric space induced by the $\|\cdot\|_2$ norm. Indeed, let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies $\|x_n - x\|_1 < \varepsilon/C$. Then $n \geq N$ implies

$$\begin{aligned}\|x_n - x\|_2 &\leq C\|x_n - x\|_1 \\ &< \varepsilon.\end{aligned}$$

Thus a sequence (x_n) converging in the $\|\cdot\|_1$ norm is a *stronger* condition than the sequence (x_n) converging in the $\|\cdot\|_2$ norm. An alternative way of thinking about this is that the topology induced by the $\|\cdot\|_1$ norm is *finer* than the topology induced by the $\|\cdot\|_2$ norm. Indeed, if $B_r^2(0)$ is the open ball of radius r centered at 0 in the $\|\cdot\|_2$ norm and $B_{r/C}^1(0)$ is the open ball of radius r/C in the $\|\cdot\|_1$ norm, then we have

$$B_r^2(0) \subseteq B_{r/C}^1(0).$$

More generally, if $a \in V$, then

$$\begin{aligned}B_r^2(a) &= a + B_r^2(0) \\ &\subseteq a + B_{r/C}^1(0) \\ &= B_{r/C}^1(a).\end{aligned}$$

This implies the topology induced by $\|\cdot\|_1$ is finer than the topology induced by $\|\cdot\|_2$.

9.3 Closure

Definition 9.7. Let $A \subseteq \mathcal{V}$. A point $x \in \mathcal{V}$ is said to be a **closure point** of A if every open neighborhood of x meets A. This means that for any $\varepsilon > 0$, we have $B_\varepsilon(x) \cap A \neq \emptyset$, where

$$B_\varepsilon(x) := \{y \in \mathcal{V} \mid \|x - y\| < \varepsilon\}.$$

The **closure of** A is the set of all closure points of A and is denoted $\overline{A}$.

Remark 17. If every open neighborhood of a point $x \in \mathcal{V}$ meets A, then we can find a sequence (x_n) of points in A such that $x_n \to x$. Conversely, if (x_n) is a sequence of points in A such that $x_n \to x$, then every open neighborhood of x meets A. Thus, an equivalent condition for x to be a closure point of A is that there exists a sequence (x_n) of points in A such that $x_n \to x$. We will prove this in a moment.

Proposition 9.10. *Let $A, B \subseteq \mathcal{V}$. Then*

1. $A \subseteq \overline{A}$;
2. *If* $A \subseteq B$, *then* $\overline{A} \subseteq \overline{B}$;
3. $\overline{A \cup B} = \overline{A} \cup \overline{B}$;
4. $\overline{A \cap B} \subseteq \overline{A} \cap \overline{B}$;
5. $\overline{\overline{A}} = \overline{A}$.

Proof.

1. Let $x \in A$. Then every open neighborhood of x meets A since, in particular, every open neighborhood meets $x \in A$. Therefore $x \in \overline{A}$.

2. Let $x \in \overline{A}$. Then every open neighborhood of x meets B since, in particular, each open neighborhood meets A. Therefore $x \in \overline{B}$.

3. First note that $\overline{A \cup B} \supseteq \overline{A} \cup \overline{B}$ follows from 2. For the reverse inclusion, let $x \in \overline{A \cup B}$ and assume for a contradiction that $x \notin \overline{A} \cup \overline{B}$. Then there exists a open neighborhood U of x such that $U \cap A = \emptyset$ and an open neighborhood V of x such that $V \cap A = \emptyset$. Choose such open neighborhoods U and V and set $W = U \cap V$. Then W is a open neighborhood of x such that

$$W \cap (A \cup B) = (W \cap A) \cup (W \cap B) = \emptyset,$$

which contradicts the fact that $x \in \overline{A \cup B}$.

4. This follows from 2. To see why we do not necessarily get the reverse equality, consider $A = (0, 1)$ and $B = (1, 2)$ in $\mathbb{R}$. Then $\overline{A} \cap \overline{B} = \{1\}$ but $\overline{A \cap B} = \overline{\emptyset} = \emptyset$.

5. Exercise.

$\square$

Proposition 9.11. *$x \in \overline{A}$ if and only if there exists a sequence (x_n) of elements in A such that $x_n \to x$.*

Proof. Assume $x \in \overline{A}$. For each $n \in \mathbb{N}$, choose $x_n \in B_{1/n}(x) \cap A$ (this set is nonempty since $x \in \overline{A}$). Then one readily checks that $x_n \to x$.

Conversely, suppose that (x_n) in A such that $x_n \to x$ and assume for a contradiction that $x \notin \overline{A}$. Then there exists an open neighborhood $B_\varepsilon(x)$ of x such that $B_\varepsilon(x) \cap A = \emptyset$. Choose such an open neighborhood and choose $N \in \mathbb{N}$ such that $1/N < \varepsilon$. Then $n \geq N$ implies

$$x_n \in B_{1/N}(x) \cap A \subseteq B_\varepsilon(x) \cap A,$$

which is a contradiction. $\square$

Definition 9.8. A set $A \subseteq \mathcal{V}$ is said to be a **closed set** if $A = \overline{A}$.

In finite dimensional inner-product spaces every subspace is a closed set. However this is no longer true in infinite dimensional spaces.

Example 9.1. Let $\ell_0(\mathbb{N})$ be the subset of $\ell^2(\mathbb{N})$ consisting of all square summable sequences (x_n) with only finitely many nonzero terms. It is easy to see that $\ell_0(\mathbb{N})$ is in fact a subspace of $\ell^2(\mathbb{N})$. However $\ell_0(\mathbb{N})$ is not a *closed* subspace of $\ell^2(\mathbb{N})$. Indeed, consider the sequence of elements in $\ell_0(\mathbb{N})$ given by

$$\begin{aligned} x^1 &= (1, 0, 0, 0, \cdots) \\ x^2 &= (1, 1/2, 0, 0, \cdots) \\ x^3 &= (1, 1/2, 1/3, 0, \dots) \\ &\vdots \end{aligned}$$

Then the sequence (x^n) of sequences x^n converges to the sequence $(1/n) \notin \ell_0(\mathbb{N})$. Therefore $\ell_0(\mathbb{N})$ is not closed.

To see why we have $(x^n) \to (1/n)$, let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $\sum_{k=N}^{\infty} 1/k^2 < \varepsilon$ (there exists such an N since $\sum_{k=0}^{\infty} 1/k^2 < \infty$). Then for all $n \geq N$, we have

$$\|(x^n) - (1/n)\| \leq \sum_{k=N}^{\infty} \frac{1}{k^2} < \varepsilon.$$

Theorem 9.3. *Let $\mathcal{U}$ be a subspace of $\mathcal{V}$. Then $\overline{\mathcal{U}}$ is a subspace of $\mathcal{V}$.*

Proof. Let $x, y \in \overline{\mathcal{U}}$ and $\lambda \in \mathbb{C}$. Let (x_n) and (y_n) be two sequences of elements in $\mathcal{U}$ such that $x_n \to x$ and $y_n \to y$. Then $(\lambda x_n + y_n)$ is a sequence of elements in $\mathcal{U}$ such that $\lambda x_n + y_n \to \lambda x + y$. Therefore $\lambda x + y \in \overline{\mathcal{U}}$, which implies $\overline{\mathcal{U}}$ is a subspace of $\mathcal{V}$. $\square$

Proposition 9.12. *If $\mathcal{U}$ is a subspace of $\mathcal{V}$, then the closure $\overline{\mathcal{V}}$ is a closed subspace.*

Proof. Clearly $\overline{\mathcal{U}}$ is a closed set. Therefore it suffices to show that $\overline{\mathcal{U}}$ is a subspace of $\mathcal{V}$. Let $x, y \in \overline{\mathcal{U}}$ and let $\lambda \in \mathbb{C}$. Choose sequence (x_n) and (y_n) in $\mathcal{U}$ such that $x_n \to x$ and $y_n \to y$. Then since $\lambda x_n + y_n \in \mathcal{U}$ for each $n \in \mathbb{N}$ and

$$\lim_{n \to \infty} (\lambda x_n + y_n) = \lambda x + y,$$

we have $\lambda x + y \in \overline{\mathcal{U}}$. Therefore $\overline{\mathcal{U}}$ is a subspace of $\mathscr{V}$. $\square$

10 Hilbert Spaces

Let $\mathcal{V}$ be an inner-product space. A sequence $(x_n) \in \mathcal{V}$ is said to be a **Cauchy sequence** if for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$n, m \geq N \text{ implies } \|x_n - x_m\| < \varepsilon.$$

Every convergent sequence in $\mathcal{V}$ is a Cauchy sequence. Indeed:

Proposition 10.1. *Let $\mathcal{V}$ be an inner-product space and let (x_n) be a convergent sequence in $\mathcal{V}$. Then (x_n) is a Cauchy sequence.*

Proof. Let $x \in \mathcal{V}$ be the limit of (x_n) and let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies $\|x_n - x\| < \varepsilon/2$. Then $m, n \geq N$ implies

$$\begin{aligned} \|x_n - x_m\| &= \|x_n - x + x - x_m\| \\ &\leq \|x_n - x\| + \|x - x_m\| \\ &< \varepsilon/2 + \varepsilon/2 \\ &= \varepsilon. \end{aligned}$$

It follows that (x_n) is Cauchy. □

Though every convergent sequence is Cauchy, the converse need not hold. For instance, in $\mathbb{Q}$, we can construct a sequence of rational numbers which gets closer and closer to π. Namely, such a sequence starts out as

$$(3, 3.1, 3.14, 3.141, \dots). \tag{7}$$

It's easy to see that such a sequence is in fact a Cauchy sequence. However, by construction, this sequence converges to π, which is not a rational number. On the other hand, since π is a real number, the Cauchy sequence (7) of real numbers converges to a real number. Is every Cauchy sequence in $\mathbb{R}$ convergent? It turns out that the answer to this is, yes! We give a special name to inner-product spaces which share this property with $\mathbb{R}$.

Definition 10.1. Let $\mathcal{V}$ be an inner-product space. We say $\mathcal{V}$ is a **Hilbert space** if every Cauchy sequence in $\mathcal{V}$ converges to a limit in $\mathcal{V}$.

The most basic examples with Hilbert spaces are $\mathbb{R}$, $\mathbb{R}^n$, and $\mathbb{C}^n$ (with their usual inner-product). We will show later on that $\ell^2(\mathbb{N})$ is also a Hilbert space. Nonexamples of Hilbert spaces include $\mathbb{Q}$ and $C[a, b]$. Hilbert spaces are usually denoted by $\mathcal{H}$ and $\mathcal{K}$.

10.1 Distances

Definition 10.2. Let $\mathcal{V}$ be an inner-product space, let $A, B \subseteq \mathcal{V}$, and let $x \in \mathcal{V}$. We define the **distance from x to A**, denoted $\mathrm{d}(x, A)$, by

$$\mathrm{d}(x, A) := \inf\{\|x - a\| \mid a \in A\}.$$

More generally, we defined the **distance from A to B**, denoted $\mathrm{d}(A, B)$, by

$$\mathrm{d}(A, B) := \inf\{\|a - b\| \mid a \in A \quad \text{and} \quad b \in B\}.$$

Proposition 10.2. *With the notation as in Definition (10.2), we have $\mathrm{d}(x, A) = 0$ if and only if $x \in \overline{A}$.*

Proof. Suppose $\mathrm{d}(x, A) = 0$. Then for each $n \in \mathbb{N}$ we can choose $a_n \in A$ such that

$$\mathrm{d}(x, A) \leq \|x - a_n\| < \mathrm{d}(x, A) + \frac{1}{n}.$$

In other words, since $\mathrm{d}(x, A) = 0$, we can find a sequence (a_n) in A such that

$$\|x - a_n\| < \frac{1}{n}$$

for all $n \in \mathbb{N}$. This implies $x \in \overline{A}$.

Conversely, suppose $x \in \overline{A}$. Then we can find a sequence (a_n) in A such that

$$\mathrm{d}(x, A) \leq \|x - a_n\| < \frac{1}{n}$$

for all $n \in \mathbb{N}$. Taking $n \to \infty$ gives us $\mathrm{d}(x, A) = 0$. □

Remark 18. A common technique that we do is we choose a sequence (a_n) of elements in A such that

$$\|x - a_n\| < d(x, A) + \frac{1}{n}$$

for all $n \in \mathbb{N}$. Such a sequence must exist since otherwise there would exist an $n \in \mathbb{N}$ such that

$$d(x, A) + \frac{1}{n} \leq \|x - a\|$$

for all $a \in A$, and this contradicts the fact that $d(x, A)$ is the infimum.

10.1.1 Absolute Homogeneity of Distances

Proposition 10.3. *Let $\mathcal{V}$ be an inner-product space, let $\mathcal{A}$ be a subspace of $\mathcal{V}$, let $x \in \mathcal{V}$, and let $\lambda \in \mathbb{C}$. Then*

$$d(\lambda x, \mathcal{A}) = |\lambda| d(x, \mathcal{A}).$$

Proof. Choose a sequence (y_n) of elements in $\mathcal{A}$ such that

$$\|x - y_n\| < d(x, \mathcal{A}) + \frac{1}{|\lambda n|}$$

for all $n \in \mathbb{N}$. Then since $\mathcal{A}$ is a subspace, we have

$$\begin{aligned} d(\lambda x, \mathcal{A}) &\leq \|\lambda x - \lambda y_n\| \\ &= |\lambda| \|x - y_n\| \\ &< |\lambda| \left(d(x, \mathcal{A}) + \frac{1}{|\lambda n|} \right) \\ &= |\lambda| d(x, \mathcal{A}) + \frac{1}{n} \end{aligned}$$

for all $n \in \mathbb{N}$. In particular, this implies $d(\lambda x, \mathcal{A}) \leq |\lambda| d(x, \mathcal{A})$.

Conversely, choose a sequence (z_n) of elements in $\mathcal{A}$ such that

$$\|\lambda x - z_n\| < d(\lambda x, \mathcal{A}) + \frac{1}{n}$$

Then since $\mathcal{A}$ is a subspace, we have

$$\begin{aligned} |\lambda| d(x, \mathcal{A}) &\leq |\lambda| \|x - z_n / |\lambda|\| \\ &= \|\lambda x - z_n\| \\ &< d(\lambda x, \mathcal{A}) + \frac{1}{n} \end{aligned}$$

for all $n \in \mathbb{N}$. In particular, this implies $|\lambda| d(x, \mathcal{A}) \leq d(\lambda x, \mathcal{A})$. □

10.1.2 Subadditivity of Distances

Proposition 10.4. *Let $\mathcal{V}$ be an inner-product space, let $\mathcal{A}$ be a subspace of $\mathcal{V}$, and let $x, y \in \mathcal{V}$. Then*

$$d(x + y, \mathcal{A}) \leq d(x, \mathcal{A}) + d(y, \mathcal{A}).$$

Proof. Choose a sequences (w_n) and (z_n) of elements in $\mathcal{A}$ such that

$$\|x - w_n\| < d(x, \mathcal{A}) + \frac{1}{2n} \quad \text{and} \quad \|y - z_n\| < d(y, \mathcal{A}) + \frac{1}{2n}$$

for all $n \in \mathbb{N}$. Then since $\mathcal{A}$ is a subspace, we have

$$\begin{aligned} d(x + y, \mathcal{A}) &\leq \|(x + y) - (w_n + z_n)\| \\ &\leq \|x - w_n\| + \|y - z_n\| \\ &< d(x, \mathcal{A}) + \frac{1}{2n} + d(y, \mathcal{A}) + \frac{1}{2n} \\ &= d(x, \mathcal{A}) + d(y, \mathcal{A}) + \frac{1}{n} \end{aligned}$$

for all $n \in \mathbb{N}$. In particular, this implies $d(x + y, \mathcal{A}) \leq d(x, \mathcal{A}) + d(y, \mathcal{A})$. □

10.2 Orthogonal Projection

Let K be a 2-dimensional subspace in $\mathbb{R}^3$. Such a subspace corresponds to a plane in $\mathbb{R}^3$ which passes through the origin. One of the main tools that we use in this setting is the concept of projecting onto K. For instance, if K corresponds to the plane $\{z = 0\}$ in $\mathbb{R}^3$, then the projection of the vector $(1,2,1)^\top$ onto K gives the vector $(1,2,0)^\top$. In fact, $(1,2,0)^\top$ is the *closest* vector to $(1,2,1)^\top$ which belongs to K. In other words, if $(a,b,0)^\top \in K$, then

$$\begin{aligned}\sqrt{(1-a)^2+(2-b)^2+1} &= \|(1,2,1)^\top - (a,b,0)^\top\| \\ &\geq \|(1,2,1)^\top - (1,2,0)^\top\| \\ &= 1.\end{aligned}$$

We wish to generalize these concepts from $\mathbb{R}^3$ to an arbitrary Hilbert space.

Theorem 10.1. *Let $\mathcal{H}$ be a Hilbert space and let $\mathcal{K}$ be a closed subspace of $\mathcal{H}$. Suppose $x \in \mathcal{H}\backslash\mathcal{K}$. Then there exists a unique $a \in \mathcal{K}$ such that $\mathrm{d}(x,\mathcal{K}) = \|x-a\|$.*

Proof. Choose a sequence (a_n) of elements in $\mathcal{K}$ such that

$$\mathrm{d}(x,\mathcal{K}) \leq \|x-a_n\| < \mathrm{d}(x,\mathcal{K}) + 1/n. \tag{8}$$

for all $n \in \mathbb{N}$. We claim that (a_n) is a Cauchy sequence. The key to showing this is to use the parallelogram identity, which says

$$\|y-z\|^2 + \|y+z\|^2 = 2\|y\|^2 + 2\|z\|^2 \tag{9}$$

for all $y,z \in \mathcal{H}$. Setting $y = x - a_m$ and $z = x - a_n$ in (9) and rearranging terms, we have

$$\begin{aligned}\|a_m - a_n\|^2 &= 2\|x-a_m\|^2 + 2\|x-a_n\|^2 - \|(x-a_m)+(x-a_n)\|^2 \\ &< 2(\mathrm{d}(x,\mathcal{K})+1/m)^2 + 2(\mathrm{d}(x,\mathcal{K})+1/n)^2 - \|(2(x-(a_n-a_m)/2))\|^2 \\ &= 4\mathrm{d}(x,\mathcal{K})^2 + (4/m+4/n)\mathrm{d}(x,\mathcal{K}) + 2/m^2 + 2/n^2 - 4\|(x-(a_n-a_m)/2\|^2 \\ &\leq 4\mathrm{d}(x,\mathcal{K})^2 + (4/m+4/n)\mathrm{d}(x,\mathcal{K}) + 2/m^2 + 2/n^2 - 4\mathrm{d}(x,\mathcal{K})^2 \\ &= (4/n+4/m)\mathrm{d}(x,\mathcal{K}) + 2/n^2 + 2/m^2.\end{aligned}$$

Thus if $\varepsilon > 0$, then we choose $N \in \mathbb{N}$ such that $n,m \geq N$ implies

$$(4/n+4/m)\mathrm{d}(x,\mathcal{K}) + 2/n^2 + 2/m^2 < \varepsilon^2.$$

which implies $\|a_m - a_n\| < \varepsilon$. Therefore the sequence (a_n) is a Cauchy sequence, and since we are in a Hilbert space, (a_n) must be convergent, say $a_n \to a$, with $a \in \mathcal{K}$ since $\mathcal{K}$ is closed. Then taking the limit of (8) as $n \to \infty$ gives us $\mathrm{d}(x,\mathcal{K}) = \|x-a\|$.

To see uniqueness of a, let b be another point in $\mathcal{K}$ such that $\|x-b\| = \mathrm{d}(x,\mathcal{K})$. We again use the parallelogram identity. We have

$$\begin{aligned}\|b-a\|^2 &= \|(x-a)-(x-b)\|^2 \\ &= 2\|x-a\|^2 + 2\|x-b\|^2 - \|(x-a)+(x-b)\|^2 \\ &= 4\mathrm{d}(x,\mathcal{K})^2 - \|2x-(a+b)\|^2 \\ &= 4\mathrm{d}(x,\mathcal{K})^2 - 4\|x-(a+b)/2\|^2 \\ &\leq 0,\end{aligned}$$

which implies $a = b$. □

Definition 10.3. With the notation in Theorem (10.1), we shall denote $a = \mathrm{P}_{\mathcal{K}}x$ and call this point the **orthogonal projection** of x onto $\mathcal{K}$. More generally we obtain a map $\mathrm{P}_{\mathcal{K}}\colon \mathcal{H} \to \mathcal{K}$ called the **orthogonal projection** map of $\mathcal{H}$ onto $\mathcal{K}$.

Theorem 10.2. *Let $\mathcal{H}$ be a Hilbert space, let $\mathcal{K}$ be a closed subspace of $\mathcal{H}$, and let $x \in \mathcal{H}$. Then $\mathrm{P}_{\mathcal{K}}x$ is the unique point in $\mathcal{K}$ such that $\langle x - \mathrm{P}_{\mathcal{K}}x, y\rangle = 0$ for all $y \in \mathcal{K}$.*

Proof. Let $y \in \mathcal{K}$ and define a function $f \colon \mathbb{R} \to \mathbb{R}$ by

$$f(t) = \|x - \mathrm{P}_{\mathcal{K}}x + ty\|^2$$

for all $t \in \mathbb{R}$. Then $f(0) = \mathrm{d}(x, \mathcal{K})^2$ and $f(t) > \mathrm{d}(x, \mathcal{K})^2$ for all $t \neq 0$ (note we have a *strict* inequality here by uniqueness of $\mathrm{P}_{\mathcal{K}}x$). For each $t \in \mathbb{R}$, we have

$$\begin{aligned} f(t) &= \|x - \mathrm{P}_{\mathcal{K}}x + ty\|^2 \\ &= \langle x - \mathrm{P}_{\mathcal{K}}x + ty, x - \mathrm{P}_{\mathcal{K}}x + ty\rangle \\ &= \|x - \mathrm{P}_{\mathcal{K}}x\|^2 + 2\mathrm{Re}(\langle x - \mathrm{P}_{\mathcal{K}}x, ty\rangle) + \|ty\|^2 \\ &= \mathrm{d}(x, \mathcal{K})^2 + 2t\mathrm{Re}(\langle x - \mathrm{P}_{\mathcal{K}}x, y\rangle) + t^2\|y\|^2. \end{aligned}$$

So the function f is just a quadratic function in t. In particular, it is differentiable at $t = 0$, and since it has a global minimum at $t = 0$, we have

$$\begin{aligned} 0 &= f'(t)\Big|_{t=0} \\ &= \left(2\mathrm{Re}(\langle x - \mathrm{P}_{\mathcal{K}}x, y\rangle) + 2t\|y\|^2\right)\Big|_{t=0} \\ &= 2\mathrm{Re}(\langle x - \mathrm{P}_{\mathcal{K}}x, y\rangle). \end{aligned}$$

Therefore $\mathrm{Re}(\langle x - \mathrm{P}_{\mathcal{K}}x, y\rangle) = 0$ for all $y \in \mathcal{K}$. Note that this also implies

$$\begin{aligned} \mathrm{Im}(\langle x - \mathrm{P}_{\mathcal{K}}x, y\rangle) &= -\mathrm{Re}(i\langle x - \mathrm{P}_{\mathcal{K}}x, y\rangle) \\ &= -\mathrm{Re}(\langle x - \mathrm{P}_{\mathcal{K}}x, -iy\rangle) \\ &= 0 \end{aligned}$$

for all $y \in \mathcal{K}$. Combining these together gives us $\langle x - \mathrm{P}_{\mathcal{K}}x, y\rangle = 0$ for all $y \in \mathcal{K}$. This proves existence.

To prove uniqueness, let $a \in \mathcal{K}$ such that $\langle x - a, y\rangle = 0$ for all $y \in \mathcal{K}$. Then

$$\begin{aligned} 0 &= \langle x - a, y\rangle \\ &= \langle x - \mathrm{P}_{\mathcal{K}}x + \mathrm{P}_{\mathcal{K}}x - a, y\rangle \\ &= \langle x - \mathrm{P}_{\mathcal{K}}x, y\rangle + \langle \mathrm{P}_{\mathcal{K}}x - a, y\rangle \\ &= \langle \mathrm{P}_{\mathcal{K}}x - a, y\rangle \end{aligned}$$

for all $y \in \mathcal{K}$. In particular, setting $y = \mathrm{P}_{\mathcal{K}} - a$ gives us $\|\mathrm{P}_{\mathcal{K}}x - a\|^2 = 0$, which implies $\mathrm{P}_{\mathcal{K}}x = a$. □

10.2.1 The Orthogonal Projection Map

Recall that the map $\mathrm{P}_{\mathcal{K}}$ is called the orthogonal projection of $\mathcal{H}$ onto $\mathcal{K}$. with respect to $\mathcal{K}$. Actually in linear analysis, the words "orthogonal projection" describe a certain class of linear maps, so we should make sure that our terminology is consistent here. First, let's recall the definition of an orthgonal projection:

Definition 10.4. Let $\mathcal{V}$ be an inner-product space and let $P \colon \mathcal{V} \to \mathcal{V}$ be a linear map. We say P is a **projection** if $P^2 = P$. We say P is an **orthogonal projection** if it is a projection and it satisfies

$$\langle Px, y\rangle = \langle x, Py\rangle$$

for all $x, y \in V$.

We now want to show that $\mathrm{P}_{\mathcal{K}}$ is an orthononal projection in the sense of Definition (10.4), so that our terminology is consistent. This will be established along with other properties in the next proposition.

Proposition 10.5. *Let $\mathcal{H}$ be a Hilbert space and let $\mathcal{K}$ be a closed subspace of $\mathcal{H}$. Then*

1. $\mathrm{P}_{\mathcal{K}}$ *is $\mathbb{C}$-linear, that is,* $\mathrm{P}_{\mathcal{K}}(\lambda x + \mu y) = \lambda \mathrm{P}_{\mathcal{K}}x + \mu \mathrm{P}_{\mathcal{K}}y$ *for all $x, y \in \mathcal{H}$ and $\lambda, \mu \in \mathbb{C}$.*
2. $\|\mathrm{P}_{\mathcal{K}}x\| \leq \|x\|$ *for all $x \in \mathcal{H}$.*
3. *Let $x \in \mathcal{H}$. Then $\mathrm{P}_{\mathcal{K}}x = x$ if and only if $x \in \mathcal{K}$.*
4. $\mathrm{P}_{\mathcal{K}}$ *is an orthogonal projection, that is,* $\mathrm{P}_{\mathcal{K}}(\mathrm{P}_{\mathcal{K}}x) = \mathrm{P}_{\mathcal{K}}x$ *and* $\langle \mathrm{P}_{\mathcal{K}}x, y\rangle = \langle x, \mathrm{P}_{\mathcal{K}}y\rangle$ *for all $x, y \in \mathcal{H}$.*

Proof. 1. Let $x, y \in \mathcal{H}$ and $\lambda, \mu \in \mathbb{C}$. Then for all $z \in \mathcal{K}$, we have

$$\begin{aligned}\langle \lambda x + \mu y - \lambda \mathrm{P}_{\mathcal{K}} x - \mu \mathrm{P}_{\mathcal{K}} y, z \rangle &= \langle \lambda x - \lambda \mathrm{P}_{\mathcal{K}} x, z \rangle + \langle \mu y - \mu \mathrm{P}_{\mathcal{K}} y, z \rangle \\ &= \lambda \langle x - \mathrm{P}_{\mathcal{K}} x, z \rangle + \mu \langle y - \mathrm{P}_{\mathcal{K}} y, z \rangle \\ &= \lambda \cdot 0 + \mu \cdot 0 \\ &= 0.\end{aligned}$$

In particular, this implies

$$\lambda \mathrm{P}_{\mathcal{K}} x + \mu \mathrm{P}_{\mathcal{K}} y = \mathrm{P}_{\mathcal{K}}(\lambda x + \mu y)$$

by Proposition (10.2).

2. Let $x \in \mathcal{H}$. By the Pythagorean theorem, we have

$$\begin{aligned}\|x\|^2 &= \|\mathrm{P}_{\mathcal{K}} x\|^2 + \|\mathrm{P}_{\mathcal{K}} x - x\|^2 \\ &\geq \|\mathrm{P}_{\mathcal{K}} x\|^2,\end{aligned}$$

which implies $\|\mathrm{P}_{\mathcal{K}} x\| \leq \|x\|$.

3. Suppose $\mathrm{P}_{\mathcal{K}} x = x$. Then $x \in \mathcal{K}$ since $\mathrm{P}_{\mathcal{K}} x \in \mathcal{K}$. Conversely, suppose $x \in \mathcal{K}$. Then

$$\begin{aligned}0 &= \|x - x\| \\ &\geq \mathrm{d}(x, \mathcal{K}) \\ &= \|x - \mathrm{P}_{\mathcal{K}} x\| \\ &\geq 0.\end{aligned}$$

It follows that $\|x - \mathrm{P}_{\mathcal{K}} x\| = 0$, which implies $x = \mathrm{P}_{\mathcal{K}} x$.

4. We first show $\mathrm{P}_{\mathcal{K}}$ is a projection. Let $x \in \mathcal{H}$. Then since $\mathrm{P}_{\mathcal{K}} x \in \mathcal{K}$, we have $\mathrm{P}_{\mathcal{K}}(\mathrm{P}_{\mathcal{K}} x) = \mathrm{P}_{\mathcal{K}} x$ by part 3. Thus $\mathrm{P}_{\mathcal{K}}$ is a projection. Now we show it is an orthogonal projection. Let $x, y \in \mathcal{H}$. Then

$$\begin{aligned}0 &= \langle x - \mathrm{P}_{\mathcal{K}} x, \mathrm{P}_{\mathcal{K}} y \rangle \\ &= \langle x, \mathrm{P}_{\mathcal{K}} y \rangle - \langle \mathrm{P}_{\mathcal{K}} x, \mathrm{P}_{\mathcal{K}} y \rangle \\ &= \langle x, \mathrm{P}_{\mathcal{K}} y \rangle - \langle \mathrm{P}_{\mathcal{K}} x, (\mathrm{P}_{\mathcal{K}} y - y) + y \rangle \\ &= \langle x, \mathrm{P}_{\mathcal{K}} y \rangle - \langle \mathrm{P}_{\mathcal{K}} x, \mathrm{P}_{\mathcal{K}} y - y \rangle - \langle \mathrm{P}_{\mathcal{K}} x, y \rangle \\ &= \langle x, \mathrm{P}_{\mathcal{K}} y \rangle - \langle \mathrm{P}_{\mathcal{K}} x, y \rangle,\end{aligned}$$

which implies $\langle x, \mathrm{P}_{\mathcal{K}} y \rangle = \langle \mathrm{P}_{\mathcal{K}} x, y \rangle$. Thus $\mathrm{P}_{\mathcal{K}}$ is an orthogonal projection. $\square$

10.2.2 Orthogonal Complements

Definition 10.5. Let $\mathcal{V}$ be an inner-product space, $x, y \in \mathcal{V}$, and let $A, B \subseteq \mathcal{V}$.

1. We say x is **orthonal to** y, denoted $x \perp y$, if $\langle x, y \rangle = 0$.
2. We say x is **orthogonal to** A, denoted $x \perp A$, if $\langle x, a \rangle = 0$ for all $a \in A$.
3. We say A is **orthogonal to** B, denoted $A \perp B$, if $\langle a, b \rangle = 0$ for all $a \in A$ and $b \in B$.
4. The **orthogonal complement of** A, denoted $A^{\perp}$, is defined to be

$$A^{\perp} := \{z \in \mathcal{V} \mid z \perp A\}.$$

Theorem 10.3. *Let $\mathcal{H}$ be a Hilbert space and let $\mathcal{K} \subseteq \mathcal{L} \subseteq \mathcal{H}$. Then*

1. *we have $\mathcal{K}^{\perp} \supseteq \mathcal{L}^{\perp}$.*
2. *$\mathcal{K}^{\perp}$ is a closed subspace of $\mathcal{H}$.*
3. *If $\mathcal{K}$ is a closed subspace of $\mathcal{H}$, then every $x \in \mathcal{H}$ can be decomposed in a* unique *way as a sum of a vector in $\mathcal{K}$ and a vector in $\mathcal{K}^{\perp}$. In other words, we have $\mathcal{H} = \mathcal{K} \oplus \mathcal{K}^{\perp}$.*
4. *If $\mathcal{K}$ is a closed subspace of $\mathcal{H}$, then $(\mathcal{K}^{\perp})^{\perp} = \mathcal{K}$.*

Proof. 1. We have

$$\begin{aligned} x \in \mathcal{L}^\perp &\implies \langle x, y\rangle = 0 \text{ for all } y \in \mathcal{L} \\ &\implies \langle x, y\rangle = 0 \text{ for all } y \in \mathcal{K} \\ &\implies x \in \mathcal{K}^\perp. \end{aligned}$$

Thus $\mathcal{K}^\perp \supseteq \mathcal{L}^\perp$.

2. First we show that $\mathcal{K}^\perp$ is a subspace of $\mathcal{V}$. Let $x, z \in \mathcal{K}^\perp$ and $\lambda \in \mathbb{C}$. Then

$$\begin{aligned} \langle x + \lambda z, y\rangle &= \langle x, y\rangle + \lambda\langle z, y\rangle \\ &= 0 \end{aligned}$$

for all $y \in \mathcal{K}$. This implies $\mathcal{K}^\perp$ is a subspace of $\mathcal{V}$. Now we will show that $\mathcal{K}^\perp$ is closed. Let (x_n) be a sequence of points in $\mathcal{K}^\perp$ such that $x_n \to x$ for some $x \in \mathcal{H}$. Then since $\langle x_n, y\rangle = 0$ for all $n \in \mathbb{N}$ and $y \in \mathcal{K}$, we have

$$\begin{aligned} \langle x, y\rangle &= \lim_{n\to\infty} \langle x_n, y\rangle \\ &= \lim_{n\to\infty} 0 \\ &= 0. \end{aligned}$$

for all $y \in \mathcal{K}$. Therefore $x \in \mathcal{K}^\perp$, which implies $\mathcal{K}^\perp$ is closed.

3. Let $x \in \mathcal{H}$. Then

$$x = \mathrm{P}_{\mathcal{K}}x + (x - \mathrm{P}_{\mathcal{K}}x),$$

where $\mathrm{P}_{\mathcal{K}}x \in \mathcal{K}$ and where $x - \mathrm{P}_{\mathcal{K}}x \in \mathcal{K}^\perp$ by Theorem (10.2). This establishes existence. For uniqueness, first note that $\mathcal{K} \cap \mathcal{K}^\perp = \{0\}$. Indeed, if $y \in \mathcal{K} \cap \mathcal{K}^\perp$, then we must have $\langle y, y\rangle = 0$, which implies $y = 0$. Now suppose that

$$x = y + z$$

is another decomposition of x where $y \in \mathcal{K}$ and $z \in \mathcal{K}^\perp$. Then we have

$$\mathrm{P}_{\mathcal{K}}x + (x - \mathrm{P}_{\mathcal{K}}x) = x = y + z,$$

which implies

$$\mathrm{P}_{\mathcal{K}}x - y = (x - \mathrm{P}_{\mathcal{K}}x) - z.$$

In particular, we see that

$$\mathrm{P}_{\mathcal{K}}x - y \in \mathcal{K} \cap \mathcal{K}^\perp = \{0\} \quad \text{and} \quad (x - \mathrm{P}_{\mathcal{K}}x) - z \in \mathcal{K} \cap \mathcal{K}^\perp = \{0\}.$$

So $\mathrm{P}_{\mathcal{K}}x = y$ and $(x - \mathrm{P}_{\mathcal{K}}x) = z$. This establishes uniqueness.

4. Let $x \in \mathcal{K}$. Then $\langle x, y\rangle = 0$ for all $y \in \mathcal{K}^\perp$. Thus $x \in (\mathcal{K}^\perp)^\perp$, and so $\mathcal{K} \subseteq (\mathcal{K}^\perp)^\perp$. Conversely, let $x \in (\mathcal{K}^\perp)^\perp$. Then $\langle x, y\rangle = 0$ for all $y \in \mathcal{K}^\perp$. In particular, we have

$$\begin{aligned} \|x - \mathrm{P}_{\mathcal{K}}x\|^2 &= \langle x - \mathrm{P}_{\mathcal{K}}x, x - \mathrm{P}_{\mathcal{K}}x\rangle \\ &= \langle x, x - \mathrm{P}_{\mathcal{K}}x\rangle - \langle \mathrm{P}_{\mathcal{K}}x, x - \mathrm{P}_{\mathcal{K}}x\rangle \\ &= 0 - 0 \\ &= 0, \end{aligned}$$

which implies $x = \mathrm{P}_{\mathcal{K}}x$. This implies $x \in \mathcal{K}$, and hence $(\mathcal{K}^\perp)^\perp \subseteq \mathcal{K}$. $\square$

10.3 Separable Hilbert Spaces

Definition 10.6. Let $\mathcal{V}$ be an inner-product space and let $\mathcal{H}$ be a Hilbert space.

1. A sequence (e_n) of vectors in $\mathcal{V}$ is said to be an **orthonormal sequence** if

$$\langle e_m, e_n \rangle = \begin{cases} 1 & \text{if } m = n \\ 0 & \text{else.} \end{cases}$$

2. A sequence (x_n) of vectors in $\mathcal{H}$ is said to be **complete** if

$$\overline{\text{span}}(\{x_n \mid n \in \mathbb{N}\}) = \mathcal{H}.$$

3. A sequence (e_n) of vectors in $\mathcal{H}$ is said to be an **orthonormal basis** if it is both orthonormal and complete. If $\mathcal{H}$ contains an orthonormal basis, then we say $\mathcal{H}$ is **separable**.

To give motivation for what follows, let $\mathcal{H}$ be a separable Hilbert space and let (e_n) be an orthonormal basis for $\mathcal{H}$. We will show that every $x \in \mathcal{H}$ can be represented as

$$x = \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n. \tag{10}$$

where the $\langle x, e_n \rangle$ are uniquely determined (with respect to the orthonormal basis (e_n)). Moreover we will show that the sequence $(\langle x, e_n \rangle)$ of complex numbers is square summable, and so $(\langle x, e_n \rangle) \in \ell^2(\mathbb{N})$. We will arrive at this in the following steps:

10.3.1 Orthonormal Sequences

We first show that if $(a_n) \in \ell^2(\mathbb{N})$ and (e_n) is an orthonormal sequence of vectors in $\mathcal{H}$, then the infinite sum

$$\sum_{n=1}^{\infty} a_n e_n$$

converges, and hence (10) converges as long as $(\langle x, e_n \rangle) \in \ell^2(\mathbb{N})$.

Proposition 10.6. *Let $\mathcal{H}$ be a Hilbert space. Suppose $(a_n) \in \ell^2(\mathbb{N})$ and (e_n) is an orthonormal sequence of vectors in $\mathcal{H}$. Then the sequence of partial sums (s_N), where $s_N = \sum_{n=1}^{N} a_n e_n$, converges in $\mathcal{H}$ and the limit, which we denote by $\sum_{n=1}^{\infty} a_n e_n$, satisfies*

$$\left\| \sum_{n=1}^{\infty} a_n e_n \right\|^2 = \sum_{n=1}^{\infty} |a_n|^2. \tag{11}$$

Proof. We first show that the sequence of partial sums (s_N) converges in $\mathcal{H}$. To do this, we will show (s_N) is Cauchy. Let $\varepsilon > 0$. Since the sequence of partial sums (t_N) converges, where $t_N = \sum_{n=1}^{N} |a_n|^2$, there exists $N_0 \in \mathbb{N}$ such that $M, N \geq N_0$ (with $M \leq N$) implies $|t_N - t_M|^2 < \varepsilon$. Choose such an N_0. Then $M, N \geq N_0$ (with $M \leq N$) implies

$$\begin{aligned} \|s_N - s_M\|^2 &= \left\| \sum_{n=M}^{N} a_n e_n \right\|^2 \\ &= \sum_{n=M}^{N} |a_n| \, \|e_n\|^2 \\ &= \sum_{n=M}^{N} |a_n|^2 \\ &= |t_N - t_M|^2 \\ &< \varepsilon, \end{aligned}$$

where we used the Pythagorean Theorem to get from the first line to the second line. This implies (s_N) is a Cauchy sequence. To show (11), write

$$\begin{aligned}
\left\| \sum_{n=1}^{\infty} a_n e_n \right\|^2 &= \left\| \lim_{N\to\infty} \sum_{n=1}^{N} a_n e_n \right\|^2 \\
&= \lim_{N\to\infty} \left\| \sum_{n=1}^{N} a_n e_n \right\|^2 \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} |a_n|^2. \\
&= \sum_{n=1}^{\infty} |a_n|^2.
\end{aligned}$$

□

Proposition 10.7. *Suppose (e_n) is an orthonormal sequence in a Hilbert space $\mathcal{H}$. Then for every $x \in \mathcal{H}$ we have*

1. *(Bessel's Inequality) $\sum_{n=1}^{\infty} |\langle x, e_n \rangle|^2 \leq \|x\|^2$ (in particular $(\langle x, e_n \rangle) \in \ell^2(\mathbb{N})$);*
2. *$\sum_{n=1}^{\infty} \langle x, e_n \rangle e_n$ exists.*

Proof. 1. For each $N \in \mathbb{N}$, we have

$$\begin{aligned}
0 &\leq \left\| x - \sum_{n=1}^{N} \langle x, e_n \rangle e_n \right\|^2 \\
&= \left\langle x - \sum_{n=1}^{N} \langle x, e_n \rangle e_n, x - \sum_{n=1}^{N} \langle x, e_n \rangle e_n \right\rangle \\
&= \|x\|^2 - 2\mathrm{Re} \left\langle x, \sum_{n=1}^{N} \langle x, e_n \rangle e_n \right\rangle + \left\| \sum_{n=1}^{N} \langle x, e_n \rangle e_n \right\|^2 \\
&= \|x\|^2 - 2\mathrm{Re} \sum_{n=1}^{N} \overline{\langle x, e_n \rangle} \langle x, e_n \rangle + \sum_{n=1}^{N} |\langle x, e_n \rangle|^2 \\
&= \|x\|^2 - 2 \sum_{n=1}^{N} |\langle x, e_n \rangle|^2 + \sum_{n=1}^{N} |\langle x, e_n \rangle|^2 \\
&= \|x\|^2 - \sum_{n=1}^{N} |\langle x, e_n \rangle|^2,
\end{aligned}$$

where we applied the Pythagorean Theorem to get to the fourth line from the third line. Since this holds for all $N \in \mathbb{N}$, this gives us Bessel's inequality

$$\sum_{n=1}^{\infty} |\langle x, e_n \rangle|^2 \leq \|x\|^2.$$

2. This follows from 1 and Proposition (10.6) with $(a_n) = (\langle x, x_n \rangle)$.

□

10.3.2 Complete Sequences

Here is a criterion to determine if a sequence in a Hilbert space is complete.

Proposition 10.8. *A sequence (x_n) in a Hilbert space $\mathcal{H}$ is complete if and only if the only $x \in \mathcal{H}$ with the property $\langle x, x_n \rangle = 0$ for all $n \in \mathbb{N}$ is $x = 0$.*

Proof. Let $E = \{x_n \mid n \in \mathbb{N}\}$. We first observe that

$$\begin{aligned}
\langle x, x_n \rangle = 0 \text{ for all } n \in \mathbb{N} &\iff x \in \mathrm{span}(E)^{\perp} \\
&\iff x \in \overline{\mathrm{span}}(E)^{\perp}.
\end{aligned}$$

Thus we are trying to show that $\mathcal{H} = \overline{\text{span}}(E)$ if and only if $\overline{\text{span}}(E)^{\perp} = 0$. If $\mathcal{H} = \overline{\text{span}}(E)$, then

$$\begin{aligned} 0 &= \mathcal{H}^{\perp} \\ &= \overline{\text{span}}(E). \end{aligned}$$

Conversely, if $\overline{\text{span}}(E)^{\perp} = 0$, then

$$\begin{aligned} \mathcal{H} &= 0^{\perp} \\ &= (\overline{\text{span}}(E)^{\perp})^{\perp} \\ &= \overline{\text{span}}(E), \end{aligned}$$

where the last equality follows from the fact that $\overline{\text{span}}(E)$ is a closed subspace. □

10.3.3 Unique Representations of Vectors in a Separable Hilbert Spaces

Theorem 10.4. *Let $\mathcal{H}$ be a separable Hilbert space and let (e_n) be an orthonormal basis for $\mathcal{H}$. Then for every $x \in \mathcal{H}$ we have*

$$x = \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n \quad \text{and} \quad \|x\|^2 = \sum_{n=1}^{\infty} |\langle x, e_n \rangle|^2.$$

In addition,

$$\langle x, y \rangle = \sum_{n=1}^{\infty} \langle x, e_n \rangle \overline{\langle y, e_n \rangle}$$

for all $x, y \in \mathcal{H}$.

Proof. Consider $y = x - \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n \in \mathcal{H}$. Then for all $k \in \mathbb{N}$, we have

$$\begin{aligned} \langle y, e_k \rangle &= \left\langle x - \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n, e_k \right\rangle \\ &= \langle x, e_k \rangle - \left\langle \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n, e_k \right\rangle \\ &= \langle x, e_k \rangle - \left\langle \lim_{N \to \infty} \sum_{n=1}^{N} \langle x, e_n \rangle e_n, e_k \right\rangle \\ &= \langle x, e_k \rangle - \lim_{N \to \infty} \sum_{n=1}^{N} \langle x, e_n \rangle \langle e_n, e_k \rangle \\ &= \langle x, e_k \rangle - \langle x, e_k \rangle \\ &= 0. \end{aligned}$$

Therefore $y = 0$ by Proposition (10.8).

Next, let $x, y \in \mathcal{H}$. Then

$$\begin{aligned} \langle x, y \rangle &= \left\langle \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n, \sum_{k=1}^{\infty} \langle y, e_k \rangle e_k \right\rangle \\ &= \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \langle x, e_n \rangle \overline{\langle y, e_k \rangle} \langle e_n, e_k \rangle \\ &= \sum_{n=1}^{\infty} \langle x, e_n \rangle \overline{\langle y, e_n \rangle}. \end{aligned}$$

To see that the $\langle x, e_n \rangle$ are uniquely determined. Suppose we have another representation of x, say

$$x = \sum_{n=1}^{\infty} \lambda_n e_n.$$

Then o

$$
\begin{aligned}
0 &= \|0\| \\
&= \|x - \sum_{n=1}^{\infty} \lambda_n e_n\| \\
&= \| \sum_{n=1}^{\infty} \langle x, e_n \rangle x - \sum_{n=1}^{\infty} \lambda_n e_n\| \\
&=
\end{aligned}
$$

□

10.3.4 Gram-Schmidt

We want to show that every inner-product space contains an orthonormal sequence.

Proposition 10.9. *Let $\{x_n \mid n \in \mathbb{N}\}$ be a linearly independent set of vectors in an inner-product space $\mathcal{V}$. Consider the so called Gram-Schmidt process: set $e_1 = \frac{1}{\|x_1\|} x_1$. Proceed inductively. If $e_1, e_2, \ldots, e_{n-1}$ are computed, compute e_n in two steps by*

$$
f_n := x_n - \sum_{k=1}^{n-1} \langle x_n, e_k \rangle e_k, \text{ and then set } e_n := \frac{1}{\|f_n\|} f_n.
$$

Then

1. *for every $N \in \mathbb{N}$ we have span$\{x_1, x_2, \ldots, x_N\} =$ span$\{e_1, e_2, \ldots, e_N\}$;*
2. *the set $\{e_n \mid n \in \mathbb{N}\}$ is an orthonormal set in $\mathcal{H}$;*

Remark 19. Note that $\|f_n\| \neq 0$ follows from linear independence of $\{x_n \mid n \in \mathbb{N}\}$.

Proof.

1. Let $N \in \mathbb{N}$. Then for each $1 \leq n \leq N$, we have

$$
x_n = \sum_{k=1}^{n-1} \langle x_n, e_k \rangle e_k + \|f_n\| e_n.
$$

This implies $\text{span}\{x_1, x_2, \ldots, x_N\} \subseteq \text{span}\{e_1, e_2, \ldots, e_N\}$. We show the reverse inclusion by induction on n such that $1 \leq n \leq N$. The base case $n = 1$ being $\text{span}\{x_1\} \supseteq \text{span}\{e_1\}$, which holds since $e_1 = \frac{1}{\|x_1\|} x_1$. Now suppose for some n such that $1 \leq n < N$ we have

$$
\text{span}\{x_1, x_2, \ldots, x_k\} \supseteq \text{span}\{e_1, e_2, \ldots, e_k\} \tag{12}
$$

for all $1 \leq k \leq n$. Then

$$
e_{n+1} = \frac{1}{\|f_n\|} x_n - \sum_{k=1}^{n} \frac{1}{\|f_n\|} \langle x_n, e_k \rangle e_k \in \text{span}\{x_1, x_2, \ldots, x_n\}.
$$

where we used the induction step (12) on the e_k's ($1 \leq k \leq n$). Therefore

$$
\text{span}\{x_1, x_2, \ldots, x_k\} \supseteq \text{span}\{e_1, e_2, \ldots, e_k\}
$$

for all $1 \leq k \leq n+1$, and this proves our claim.

2. By construction, we have $\langle e_n, e_n \rangle = 1$ for all $n \in \mathbb{N}$. Thus, it remains to show that $\langle e_m, e_n \rangle = 0$ whenever $m \neq n$. We prove by induction on $n \geq 2$ that $\langle e_n, e_m \rangle = 0$ for all $m < n$. Proving this also give us $\langle e_m, e_n \rangle = 0$ for all $m < n$, since

$$
\begin{aligned}
\langle e_m, e_n \rangle &= \overline{\langle e_n, e_m \rangle} \\
&= \overline{0} \\
&= 0.
\end{aligned}
$$

The base case is

$$\begin{aligned}
\langle e_2, e_1 \rangle &= \frac{1}{\|x_1\|\|f_2\|} \left\langle \left(x_2 - \frac{\langle x_2, x_1 \rangle}{\langle x_1, x_1 \rangle} x_1 \right), x_1 \right\rangle \\
&= \frac{1}{\|x_1\|\|f_2\|} \left(\langle x_2, x_1 \rangle - \langle x_2, x_1 \rangle \right) \\
&= 0
\end{aligned}$$

Now suppose that $n > 2$ and that $\langle e_n, e_m \rangle = 0$ for all $m < n$. Then

$$\begin{aligned}
\langle e_{n+1}, e_m \rangle &= \frac{1}{\|f_{n+1}\|} \langle x_{n+1} - \sum_{k=1}^{n} \langle x_{n+1}, e_k \rangle e_k, e_m \rangle \\
&= \frac{1}{\|f_{n+1}\|} \left(\langle x_{n+1}, e_m \rangle - \sum_{k=1}^{n} \langle x_{n+1}, e_k \rangle \langle e_k, e_m \rangle \right) \\
&= \frac{1}{\|f_{n+1}\|} \left(\langle x_{n+1}, e_m \rangle - \langle x_{n+1}, e_m \rangle \langle e_m, e_m \rangle \right) \\
&= \frac{1}{\|f_{n+1}\|} \left(\langle x_{n+1}, e_m \rangle - \langle x_{n+1}, e_m \rangle \right) \\
&= 0,
\end{aligned}$$

for all $m < n + 1$, where we used the induction hypothesis to get from the second line to the third line. This proves the induction step, which finishes the proof of part 2 of the proposition.

3. By 2, we know that $\{e_n \mid n \in \mathbb{N}\}$ is an orthonormal set. Thus, it suffices to show that $\{e_n \mid n \in \mathbb{N}\}$ is complete. To do this, we use the criterion that the set $\{e_n \mid n \in \mathbb{N}\}$ is complete if and only if the only $x \in \mathcal{H}$ such that $\langle x, e_n \rangle = 0$ for all $n \in \mathbb{N}$ is $x = 0$.

Let $x \in \mathcal{H}$ and suppose $\langle x, e_n \rangle = 0$ for all $n \in \mathbb{N}$. Then

$$\begin{aligned}
\langle x, x_n \rangle &= \left\langle x, \sum_{k=1}^{n-1} \langle x_n, e_k \rangle e_k + \|f_n\| e_n \right\rangle \\
&= \sum_{k=1}^{n-1} \langle x_n, e_k \rangle \langle x, e_k \rangle + \|f_n\| \langle x, e_n \rangle \\
&= 0
\end{aligned}$$

for all $n \in \mathbb{N}$. Since $\{x_n \mid n \in \mathbb{N}\}$ is complete, this implies $x = 0$. Therefore $\{e_n \mid n \in \mathbb{N}\}$ is complete. $\square$

11 Operators

In linear analysis, an **operator** is just another word for a linear map. We will stick to tradition and often refer to linear maps as operators.

11.1 Unitary Operators

Definition 11.1. An operator $U\colon \mathcal{V}_1 \to \mathcal{V}_2$ between inner-product spaces $\mathcal{V}_1$ and $\mathcal{V}_2$ is said to be **unitary** if it is an isomorphism as $\mathbb{C}$-vector spaces and it satisfies

$$\langle U(x), U(y) \rangle = \langle x, y \rangle$$

for all $x, y \in \mathcal{V}_1$. We say $\mathcal{V}_1$ and $\mathcal{V}_2$ are **unitarily equivalent** if there exists a unitary map $U\colon \mathcal{V}_1 \to \mathcal{V}_2$ which is unitary. In this case, the inverse map $U^{-1}\colon \mathcal{V}_2 \to \mathcal{V}_1$ is necessarily unitary (as one *should* check!).

Corollary 4. *Let $\mathcal{H}$ be an infinite dimensional separable Hilbert space. Then $\mathcal{H}$ is unitarily equivalent to $\ell^2(\mathbb{N})$.*

Proof. Let (e_n) be an orthonormal basis for $\mathcal{H}$. Define $U\colon \mathcal{H} \to \ell^2(\mathbb{N})$ by $U(x) = (\langle x, e_n \rangle)$ for all $x \in \mathcal{H}$. The proof that U is linear is very easy. That U is injective follows from completeness of (e_n) and Proposition (10.8).

To show that U is surjective, let $(a_n) \in \ell^2(\mathbb{N})$. Since (e_n) is an orthonormal sequence, the series $x = \sum_{n=1}^{\infty} a_n e_n$ converges in $\mathcal{H}$ by Proposition (10.6). Then

$$\begin{aligned} U(x) &= (\langle x, e_n \rangle) \\ &= \left(\left\langle \sum_{m=1}^{\infty} a_m e_m, e_n \right\rangle \right) \\ &= \left(\sum_{m=1}^{\infty} a_n \langle e_m, e_n \rangle \right) \\ &= (a_n) \end{aligned}$$

implies U is surjective. Finally, we need to show that

$$\langle U(x), U(y) \rangle = \langle x, y \rangle$$

for all $x, y \in \mathcal{H}$. To see this, let $x, y \in \mathcal{H}$. Then

$$\begin{aligned} \langle U(x), U(y) \rangle &= \langle (\langle x, e_n \rangle), (\langle y, e_n \rangle) \rangle \\ &= \sum_{n=1}^{\infty} \langle x, e_n \rangle \overline{\langle y, e_n \rangle} \\ &= \langle x, y \rangle, \end{aligned}$$

by Theorem (10.4). □

Corollary 5. *$\ell^2(\mathbb{N})$ is a Hilbert space.*

Proof. Let $(a^k)_{k=1}^{\infty}$ be a Cauchy sequence in $\ell^2(\mathbb{N})$. Then for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $i, j \geq N$ implies $\|a^i - a^j\| < \varepsilon$. By the previous corollary, for each $k \in \mathbb{N}$ there exists $x_k \in \mathcal{H}$ such that $U(x_k) = a^k$. Then

$$\begin{aligned} \|a^i - a^j\| &= \|U(x_i) - U(x_j)\| \\ &= \|U(x_i - x_j)\| \\ &= \|x_i - x_j\| \end{aligned}$$

which implies $\|a^i - a^j\| = \|x_i - x_j\|$ for all $i, j \in \mathbb{N}$, and hence $i, j \geq N$ implies $\|x_i - x_j\| < \varepsilon$. So (x_n) is a Cauchy sequence in $\mathcal{H}$. Since $\mathcal{H}$ is a Hilbert space, the sequence (x_n) must converge, say $x_n \to x$. Then using the same argument that we just used it is easy to show $a^k \to U(x)$ as $k \to \infty$. So (a^k) is convergent, which implies $\ell^2(\mathbb{N})$ is a Hilbert space. □

Proposition 11.1. *Let $\mathcal{H}$ be a Hilbert space and let $\mathcal{K}$ be a closed subspace. Suppose (e_n) is an orthonormal basis for $\mathcal{K}$. Then*

$$P_{\mathcal{K}}(x) = \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n$$

for all $x \in \mathcal{H}$.

Proof. We'll use the following fact that for any set $E \subseteq H$, we have $\overline{\text{span}}(E)^{\perp} = E^{\perp}$. Recall that $\mathrm{P}_{\mathcal{K}} x$ is the unique vector such that

$$\langle x - \mathrm{P}_{\mathcal{K}} x, y \rangle = 0$$

for all $y \in \mathcal{K}$. So by the uniqueness, it suffices to show that $\sum_{n=1}^{\infty} \langle x, e_n \rangle e_n$ has the same property. Then

$$\begin{aligned} \left\langle x - \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n, e_k \right\rangle &= \langle x, e_k \rangle - \left\langle \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n, e_k \right\rangle \\ &= \langle x, e_k \rangle - \sum_{n=1}^{\infty} \langle x, e_n \rangle \langle e_n, e_k \rangle \\ &= \langle x, e_k \rangle - \langle x, e_k \rangle \\ &= 0. \end{aligned}$$

for all $k \in \mathbb{N}$. In particular, this implies

$$\left\langle x - \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n, y \right\rangle = 0$$

for all $y \in \mathcal{K}$, which implies the proposition. □

11.2 Bounded Operators

Definition 11.2. Let $\mathcal{U}$ and $\mathcal{V}$ be normed linear spaces. A **bounded operator** $T \colon \mathcal{U} \to \mathcal{V}$ is a linear map such that

$$\sup \{\|Tx\| \mid x \in \mathcal{U} \text{ and } \|x\| \leq 1\} < \infty.$$

In this case we define

$$\|T\| := \sup \{\|Tx\| \mid x \in \mathcal{U} \text{ and } \|x\| \leq 1\}$$

and call it the **operator norm** of T.

11.2.1 Composition of Bounded Operators is Bounded

Proposition 11.2. *Let $T \colon \mathcal{H}_1 \to \mathcal{H}_2$ and $S \colon \mathcal{H} \to \mathcal{H}$ be two bounded operators. Then ST is bounded*

1. *TS is bounded and $\|TS\| \leq \|T\| \|S\|$;*
2. *$(TS)^* = S^* T^*$.*

Proof.

1. Let $x \in \mathcal{H}$ such that $\|x\| = 1$. Then

$$\begin{aligned} \|TSx\| &\leq \|T\| \|Sx\| \\ &\leq \|T\| \|S\| \|x\| \\ &= \|T\| \|S\|. \end{aligned}$$

Thus TS is bounded and $\|TS\| \leq \|T\| \|S\|$.

2. Let $y \in \mathcal{H}$. Then for all $x \in \mathcal{H}$, we have

$$\begin{aligned} \langle x, (TS)^* y \rangle &= \langle TSx, y \rangle \\ &= \langle Sx, T^* y \rangle \\ &= \langle x, S^* T^* y \rangle. \end{aligned}$$

In particular, this implies $(TS)^* y = S^* T^* y$ for all $y \in \mathcal{H}$, which implies $(TS)^* = S^* T^*$. □

Proposition 11.3. *Let $T \colon \mathcal{U} \to \mathcal{V}$ be a bounded operator. Then*

$$\|Tx\| \leq \|T\| \|x\| \tag{13}$$

for all $x \in \mathcal{U}$. *Moreover, let $M > 0$ and suppose there exists some $x_0 \in \mathcal{U}$ such that*

$$\|Tx_0\| \geq M \|x_0\|.$$

Then $M \leq \|T\|$.

Proof. Let $x \in \mathcal{U}$. If $x = 0$, then (13) is obvious, so assume $x \neq 0$. Then

$$\begin{aligned} \frac{\|Tx\|}{\|x\|} &= \left\| T \left(\frac{x}{\|x\|} \right) \right\| \\ &\leq \|T\|, \end{aligned}$$

and this implies (13).

For the latter statement, let $x_0 \in \mathcal{U}$ be such an element. Then

$$\begin{aligned} \|T\| &\geq \left\| T\left(\frac{x_0}{\|x_0\|}\right) \right\| \\ &= \frac{\|Tx_0\|}{\|x_0\|} \\ &\geq M. \end{aligned}$$

$\square$

Proposition 11.4. *Let $\mathcal{H}$ be a Hilbert space and let $T\colon \mathcal{H} \to \mathcal{H}$ be a bounded operator. Then*

$$\|T\| = \sup\{\|Tx\| \mid \|x\| = 1\}.$$

Proof. First note that

$$\begin{aligned} \sup\{\|Tx\| \mid \|x\| = 1\} &\leq \sup\{\|Tx\| \mid \|x\| \leq 1\} \\ &= \|T\|. \end{aligned}$$

We prove the reverse inequality by contradiction. Assume that $\|T\| > \sup\{\|Tx\| \mid \|x\| = 1\}$. Choose $\varepsilon > 0$ such that

$$\|T\| - \varepsilon > \sup\{\|Tx\| \mid \|x\| = 1\} \tag{14}$$

Next, choose $x \in \mathcal{H}$ such that $\|x\| \leq 1$ and $\|Tx\| \geq \|T\| - \varepsilon$. Then since $\|x\| \leq 1$ and $\left\| \frac{x}{\|x\|} \right\| = 1$, we have

$$\begin{aligned} \|T\| &\geq \left\| T\left(\frac{x}{\|x\|}\right) \right\| \\ &= \frac{\|Tx\|}{\|x\|} \\ &\geq \|Tx\| \\ &> \|T\| - \varepsilon, \end{aligned}$$

and this contradicts (14). $\square$

Exercise 1. Let $T\colon \mathcal{U} \to \mathcal{V}$ be a bounded operator and let $M \geq 0$. The following statements are equivalent.

1. $\|Tx\| \leq M\|x\|$ for all $x \in \mathcal{V}$ if and only if $\|T\| \leq M$.
2. $\|Tx\| > M\|x\|$ for some $x \in \mathcal{V}$ if and only if $\|T\| > M$.
3. $\|Tx\| \geq M\|x\|$ for some $x \in \mathcal{V}$ if and only if $\|T\| \geq M$.

11.2.2 Bounded Linear Operators and Normed Vector Spaces

We now want to justify our choice in terminology.

Definition 11.3. Let $\mathcal{V}$ and $\mathcal{W}$ be inner-product spaces. We define

$$\mathrm{Hom}(\mathcal{V}, \mathcal{W}) := \{T\colon \mathcal{V} \to \mathcal{W} \mid T \text{ is a bounded linear operator}\}.$$

$\mathrm{Hom}(\mathcal{V}, \mathcal{W})$ has the structure of a $\mathbb{C}$-vector space, where addition and scalar multiplication are defined by

$$(T + U)(x) = Tx + Ux \quad \text{and} \quad (\lambda T)(x) = T(\lambda x)$$

for all $T, U \in \mathrm{Hom}(\mathcal{V}, \mathcal{W})$, $\lambda \in \mathbb{C}$, and $v \in \mathcal{V}$.

Proposition 11.5. *Let $\mathcal{V}$ and $\mathcal{W}$ be inner-product spaces. Then $(Hom(\mathcal{V}, \mathcal{W}), \|\cdot\|)$ is a normed vector space, where $\|\cdot\|$ is the map which sends a bounded linear operator T to its norm $\|T\|$.*

Proof. An easy exercise in linear algebra shows that $\mathrm{Hom}(\mathcal{V},\mathcal{W})$ has the structure of a $\mathbb{C}$-vector space, where addition and scalar multiplication are defined by

$$(T+U)(x) = T(x) + U(x) \quad \text{and} \quad (\lambda T)(x) = T(\lambda x)$$

for all $T, U \in \mathrm{Hom}(\mathcal{V},\mathcal{W})$, $\lambda \in \mathbb{C}$, and $v \in \mathcal{V}$. The details of this are left as an exercise. We are more interested in the fact that $\mathrm{Hom}(\mathcal{V},\mathcal{W})$ is a *normed* vector space. We just need to check that $\|\cdot\|$ satisfies the conditions laid out in Definition (25.1).

We first check for subadditivity. Let $T, U \in \mathrm{Hom}(\mathcal{V},\mathcal{W})$. Then

$$\begin{aligned}
\|(T+U)(x)\| &= \|Tx + Ux\| \\
&\leq \|Tx\| + \|Ux\| \\
&\leq \|T\|\|x\| + \|U\|\|x\| \\
&= (\|T\| + \|U\|)\|x\|
\end{aligned}$$

for all $x \in \mathcal{V}$. In particular, this implies $\|T+U\| \leq \|T\| + \|U\|$. Thus we have subadditivity.

Next we check that $\|\cdot\|$ is absolutely homogeneous. Let $T \in \mathrm{Hom}(\mathcal{V},\mathcal{W})$ and let $\lambda \in \mathbb{C}$. Then

$$\begin{aligned}
\|(\lambda T)(x)\| &= \|T(\lambda x)\| \\
&= \|\lambda T x\| \\
&= |\lambda|\|Tx\|
\end{aligned}$$

for all $x \in \mathcal{V}$. In particular, this implies $\|\lambda T\| = |\lambda|\|T\|$. Thus $\|\cdot\|$ is absolutely homogeneous.

Finally we check for positive-definitiness. Let $T \in \mathrm{Hom}(\mathcal{V},\mathcal{W})$. Clearly $\|T\|$ is greater than or equal to 0 since it is the supremum of terms which are greater than or equal to 0. Suppose $\|T\| = 0$. Then

$$\begin{aligned}
\|Tx\| &\leq \|T\|\|x\| \\
&= 0 \cdot \|x\| \\
&= 0
\end{aligned}$$

for all $x \in \mathcal{V}$. In particular, this implies $Tx = 0$ for all $x \in \mathcal{V}$ (by positive-definiteness of the norm for $\mathcal{W}$). Therefore $T = 0$ since they agree on all $x \in \mathcal{V}$. □

11.2.3 Bounded Linear Operators and Continuity

Theorem 11.1. *Let $T \colon \mathcal{U} \to \mathcal{V}$ be a linear operator. Then the following are equivalent.*

1. *T is uniformly continuous;*
2. *T is continuous at $0 \in \mathcal{U}$;*
3. *T is bounded.*

Proof. We may assume that $T \neq 0$ since the theorem is obvious in this case. That 1 implies 2 is clear. Let us show that 2 implies 3. Assume T is continous at 0. Choose $\delta > 0$ such that $\|x\| < \delta$ implies $\|Tx\| < 1$ (we can do this since T is continuous at 0). Then for any nonzero $y \in \mathcal{U}$ such that $\|y\| \leq 1$, we have

$$\begin{aligned}
\|Ty\| &= \frac{2\delta}{2\delta}\|Ty\| \\
&= \frac{2}{\delta}\|T(\delta y/2)\| \\
&< \frac{2}{\delta}.
\end{aligned}$$

It follows that $\|T\| < 2/\delta$ which implies T is bounded.

To finish the proof, we just need to show 3 implies 1. Assume T is bounded. Let $\varepsilon > 0$ and choose $\delta = \varepsilon/\|T\|$. Then $\|x - y\| < \delta$ implies

$$\begin{aligned}
\|Tx - Ty\| &= \|T(x-y)\| \\
&\leq \|T\|\|x-y\| \\
&< \|T\|\frac{\varepsilon}{\|T\|} \\
&= \varepsilon.
\end{aligned}$$

It follows that T is uniformly continuous. □

Proposition 11.6. *Let $T\colon \mathcal{U} \to \mathcal{V}$ be a bounded linear operator. Then Ker(T) is a closed linear subspace of $\mathcal{U}$.*

Proof. We first show that $\mathrm{Ker}(T)$ is a linear subspace of $\mathcal{U}$. Let $x, y \in \mathrm{Ker}(T)$ and let $\lambda \in \mathbb{C}$. Then

$$\begin{aligned} T(x + \lambda y) &= T(x) + \lambda T(y) \\ &= 0 + \lambda \cdot 0 \\ &= 0 \end{aligned}$$

implies $x + \lambda y \in \mathrm{Ker}(T)$. Thus, $\mathrm{Ker}(T)$ is a linear subspace of $\mathcal{U}$.

To see that $\mathrm{Ker}(T)$ is closed, let (x_n) be a sequence of elements in $\mathrm{Ker}(T)$ such that $x_n \to x$ where $x \in \mathcal{U}$. Since T is bounded, it is uniformly continuous (and in particular continuous at x). Therefore

$$\begin{aligned} 0 &= \lim_{n\to\infty} 0 \\ &= \lim_{n\to\infty} T(x_n) \\ &= T(\lim_{n\to\infty} x_n) \\ &= T(x), \end{aligned}$$

which implies $x \in \mathrm{Ker}(T)$. Thus $\mathrm{Ker}(T)$ is closed in $\mathcal{U}$. □

Remark 20. Note that $\mathrm{Im}(T)$ is not always closed.

11.3 Examples of Bounded Operators

11.3.1 Multiplication by Continuous Function is Bounded Operator

Proposition 11.7. *Let $k \in C[a,b]$. Then the operator $T\colon C[a,b] \to C[a,b]$ defined by*

$$Tf = kf$$

for all $f \in C[a,b]$ is bounded. It's norm will be explicitly computed in the proof below.

Proof. We first show it is linear. Let $f, g \in C[a,b]$ and let $\lambda, \mu \in \mathbb{C}$. Then we have

$$\begin{aligned} T(\lambda f + \mu g) &= k(\lambda f + \mu g) \\ &= \lambda k f + \mu k g \\ &= \lambda T(f) + \mu T(g). \end{aligned}$$

Thus, T is linear.

Next we show it is bounded. If $k = 0$, then $\|T\| = 0$, so assume $k \neq 0$. Since k is continuous on the compact interval $[a,b]$, there exists $c \in [a,b]$ such that $|k(x)| \leq |k(c)|$ for all $x \in [a,b]$. Choose such a $c \in [a,b]$ and let $f \in C[a,b]$ such that $\|f\| \leq 1$. Then

$$\begin{aligned} \|Tf\| &= \|kf\| \\ &= \sqrt{\int_a^b |k(x)|^2 |f(x)|^2 dx} \\ &\leq |k(c)| \sqrt{\int_a^b |f(x)|^2 dx} \\ &\leq |k(c)|. \end{aligned}$$

implies $\|T\| \leq |k(c)|$, and hence T is bounded.

To find the norm of T, let $\varepsilon > 0$ such that $\varepsilon < |k(c)|$. Without loss of generality, assume that $c < b$ (if $c = b$, then we swap the role of b with a in the argument which follows). Choose $c' \in (c, b)$ such that $|k(x)| \geq |k(c)| - \varepsilon$ for all $x \in (c, c')$ (such a c' must exist since k is continuous) and choose f to be a nonzero continuous function in $C[a,b]$ which vanishes outside the interval (c, c'). Then

$$|k(x)||f(x)| \geq (|k(c)| - \varepsilon)|f(x)|$$

for all $x \in (a, b)$. In particular, this implies

$$\begin{aligned}
\|Tf\| &= \|kf\| \\
&= \sqrt{\int_a^b |k(x)f(x)|^2 dx} \\
&\geq \sqrt{\int_a^b (|k(c)| - \varepsilon)|f(x)|^2 dx} \\
&= (|k(c)| - \varepsilon)\sqrt{\int_a^b |f(x)|^2 dx} \\
&= (|k(c)| - \varepsilon)\|f\|.
\end{aligned}$$

Therefore $\|T(f/\|f\|)\| \geq |k(c)| - \varepsilon$, and this implies

$$\|T\| \geq |k(c)| - \varepsilon \tag{15}$$

Since (34) holds for all $\varepsilon > 0$, we must have $\|T\| \geq |k(c)|$. Thus $\|T\| = |k(c)|$. □

11.3.2 Orthogonal Projections are Bounded

Example 11.1. Let $\mathcal{H}$ be a Hilbert space and let $\mathcal{K}$ be a closed subspace of $\mathcal{H}$. Then the orthogonal projection map $\mathrm{P}_{\mathcal{K}} \colon \mathcal{H} \to \mathcal{K}$ is a bounded operator. This is because $\mathrm{P}_{\mathcal{K}}$ is a linear map and $\|\mathrm{P}_{\mathcal{K}} x\| \leq \|x\|$ for all $x \in \mathcal{H}$. In particular, If $\mathcal{K} \neq 0$, then $\|\mathrm{P}_{\mathcal{K}}\| \leq 1$. In fact, we have equality here: choose any nonzero $x \in \mathcal{K}$ such that $\|x\| = 1$. Then $\|\mathrm{P}_{\mathcal{K}} x\| = 1$.

11.3.3 Unitary Operators are Bounded

Example 11.2. Let $U \colon \mathcal{H}_1 \to \mathcal{H}_2$ be a unitary operator. Then U is bounded. Indeed, this is because

$$\begin{aligned}
\|U\| &= \sup\{\|Ux\| \mid \|x\| \leq 1\} \\
&= \sup\{\|x\| \mid \|x\| \leq 1\} \\
&= 1.
\end{aligned}$$

11.3.4 Diagonal Operator is Bounded Operator

Example 11.3. Let (a_n) be a bounded sequence of complex numbers. Let $M = \sup(|a_n|)$. Define $T \colon \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$ by

$$T((x_n)) = (a_n x_n)$$

for all $(x_n) \in \ell^2(\mathbb{N})$. Then T is a linear map and note that

$$\begin{aligned}
\|T((x_n))\| &= \|(a_n x_n)\| \\
&= \sqrt{\sum_{n=1}^{\infty} |a_n x_n|^2} \\
&\leq \sqrt{\sum_{n=1}^{\infty} M^2 |x_n|^2} \\
&= M\sqrt{\sum_{n=1}^{\infty} |x_n|^2} \\
&= M\|(x_n)\|.
\end{aligned}$$

Thus

$$\begin{aligned}
\|T\| &= \sup\{\|T((x_n))\| \mid \|(x_n)\| \leq 1\} \\
&\leq \sup\{M\|(x_n)\| \mid \|(x_n)\| \leq 1\} \\
&= M.
\end{aligned}$$

We claim that $\|T\| = M$. Assume for a contradiction that $||T|| < M$. Choose $k \in \mathbb{N}$ such that $|a_k| > \|T\|$. Let $e^k = (0, 0, \ldots, 0, 1, 0, \ldots,)$. Then $\|e^k\| = 1$ and $\|Te^k\| = |a_k| > \|T\|$. Contradiction.

11.3.5 Shift Operator is Bounded Operator

Let $S\colon \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$ be the unique linear map such that $S(e_n) = e_{n+1}$ for all $n \in \mathbb{N}$, where e_n is the standard orthonormal basis vector with 1 in its nth coordinate and 0 everywhere else. Then

$$\begin{aligned}\|Sx\| &= \left\| S\left(\sum_{n=1}^{\infty} x_n e_n\right)\right\| \\ &= \left\|\sum_{n=1}^{\infty} x_n e_{n+1}\right\| \\ &= \sqrt{\sum_{n=1}^{\infty} |x_n|^2} \\ &= \|x\|\end{aligned}$$

for all $x \in \ell^2(\mathbb{N})$ implies S is bounded and $\|S\| = 1$.

11.3.6 Operator on $(C[0,1], \langle\cdot,\cdot\rangle)$

Let $T\colon C[0,1] \to \mathbb{C}$ be defined by

$$T(f) = \int_0^1 f(x)dx$$

for all $f \in C[0,1]$. Then T is linear since the integral is linear. Furthermore

$$\begin{aligned}\|Tf\| &= |Tf| \\ &= \left|\int_0^1 f(x)dx\right| \\ &\leq \sqrt{\int_0^1 |f(x)|^2}\sqrt{\int_0^1 |1|^2 dx} \\ &\leq \sqrt{\int_0^1 |f(x)|^2} \\ &= \|f\|,\end{aligned}$$

where we applied Cauchy-Schwarz. Thus, T is bounded and $\|T\| \leq 1$. Moreover, $T(1) = 1$, and so $\|T(1)\| = 1$. In particular, this implies $\|T\| = 1$.

11.4 Riesz Representation Theorem

Theorem 11.2. *(Riesz representation theorem) Let $\mathcal{H}$ be a Hilbert space and let $\ell\colon \mathcal{H} \to \mathbb{C}$ be a bounded operator. Then there exists a unique vector $y \in \mathcal{H}$ such that $\ell = \ell_y$. In other words, we have*

$$\ell(x) = \langle x, y\rangle$$

for all $x \in \mathcal{H}$. Moreover, we have $\|\ell\| = \|y\|$.

Proof. If $\ell = 0$ then the theorem is clear, so assume $\ell \neq 0$. Denote $\mathcal{K} = \ker \ell$. Then $\mathcal{K}$ is a closed proper subspace of $\mathcal{H}$. Choose a nonzero vector z in $\mathcal{K}^\perp$. Note that $\ell(z) \neq 0$ since $\mathcal{K} \cap \mathcal{K}^\perp = \{0\}$. Now for any $x \in \mathcal{H}$, we can express it as

$$x = \left(x - \frac{\ell(x)}{\ell(z)}z\right) + \frac{\ell(x)}{\ell(z)}z \tag{16}$$

where $x - (\ell(x)/\ell(z))z \in \mathcal{K}$ and where $(\ell(x)/\ell(z))z \in \mathcal{K}^\perp$. Applying $\langle \cdot, z \rangle$ to both sides of (**??**) gives us

$$\begin{aligned}\langle x, z \rangle &= \left\langle x - \frac{\ell(x)}{\ell(z)} z, z \right\rangle + \left\langle \frac{\ell(x)}{\ell(z)} z, z \right\rangle \\ &= \left\langle \frac{\ell(x)}{\ell(z)} z, z \right\rangle \\ &= \frac{\ell(x)}{\ell(z)} \|z\|^2.\end{aligned}$$

In particular, we see that

$$\begin{aligned}\ell(x) &= \frac{\ell(z)}{\|z\|^2} \langle x, z \rangle \\ &= \left\langle x, \frac{\overline{\ell(z)}}{\|z\|^2} z \right\rangle.\end{aligned}$$

So setting $y = (\overline{\ell(z)}/\|z\|^2)z$, it follows that that $\ell = \ell_y$.

This proves the existence of $y \in \mathcal{H}$. To show uniqueness, suppose $y' \in \mathcal{H}$ such that $\ell_y = \ell = \ell_{y'}$. Then

$$\langle x, y \rangle = \ell(x) = \langle x, y' \rangle$$

for all $x \in \mathcal{H}$. However since $\langle \cdot, \cdot \rangle$ is positive-definite, this implies $y = y'$.

Finally, we need to check that $\|\ell\| = \|y\|$. Let $x \in \mathcal{H}$ such that $\|x\| \leq 1$. Then

$$\begin{aligned}|\ell(x)| &= |\langle x, y \rangle| \\ &\leq \|x\| \|y\| \\ &= \|y\|.\end{aligned}$$

Thus $\|\ell\| \leq \|y\|$. To see that equality is achieved, consider $x = y/\|y\|$. In this case, we have

$$\begin{aligned}\left| \ell \left(\frac{y}{\|y\|} \right) \right| &= \frac{1}{\|y\|} |\ell(y)| \\ &= \frac{1}{\|y\|} |\langle y, y \rangle| \\ &= \frac{1}{\|y\|} \|y\|^2 \\ &= \|y\|.\end{aligned}$$

It follows that $\|\ell\| = \|y\|$. □

11.4.1 Schur's test for boundedness

Let $T\colon \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$ be the operator given by

$$T(x)_i = \sum_{j=1}^{\infty} a_{ij}.$$

Then T is bounded if

$$\sup_i \left\{ \sum_{j=1}^{\infty} |a_{ij}| \right\} \sup_j \left\{ \sum_{i=1}^{\infty} |a_{ij}| \right\} < \infty$$

Set $\alpha := \sup_i \left\{ \sum_{j=1}^{\infty} |a_{ij}| \right\}$ and $\beta := \sup_j \left\{ \sum_{i=1}^{\infty} |a_{ij}| \right\}$. Then in this case, we have

$$\|T\| \leq \sqrt{\alpha\beta}.$$

Indeed, let $x \in \ell^2(\mathbb{N})$ such that $\|x\| \leq 1$. Then

$$\|Tx\| = \sup \left\{ \langle Tx, y \rangle \mid \|y\| \leq 1 \right\}.$$

Take $y \in \ell^2(\mathbb{N})$ such that $\|y\| \leq 1$. Then

$$\begin{aligned}
|\langle Tx, y\rangle| &= |\langle \sum_{i=1}^{\infty} a_{ij}x_j), y_i\rangle| \\
&= \left| \sum_{i=1}^{\infty} \left(\sum_{j=1}^{\infty} a_j x_j \right) \overline{y}_i \right| \\
&\leq \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |a_{ij}||x_i||y_j| \\
&\leq \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |a_{ij}| \frac{c|x_j|^2 + \frac{1}{c}|y_i|^2}{2} \\
&= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{c}{2}|a_{ij}||x_j|^2 + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{1}{2c}|a_{ij}||y_i|^2 \\
&= \sum_{j=1}^{\infty} \frac{c}{2}|x_j|^2 \sum_{i=1}^{\infty} |a_{ij}| + \sum_{i=1}^{\infty} \frac{1}{2c}|y_i|^2 \sum_{j=1}^{\infty} |a_{ij}| \\
&\leq \sum_{j=1}^{\infty} \frac{c}{2}|x_j|^2 \beta + \sum_{i=1}^{\infty} \frac{1}{2c}|y_i|^2 \alpha \\
&= \frac{c\beta}{2} \sum_{j=1}^{\infty} |x_j|^2 + \frac{\alpha}{2c} \sum_{i=1}^{\infty} |y_i|^2 \\
&= \frac{c\beta}{2}\|x\|^2 + \frac{\alpha}{2c}\|y\|^2 \\
&\leq \frac{c\beta}{2} + \frac{\alpha}{2c}
\end{aligned}$$

where $c > 0$. Now choose $c = \sqrt{\frac{\alpha}{\beta}}$ (this makes the inequality minimal). Then we get

$$|\langle Tx, y\rangle| = \sqrt{\alpha\beta}$$

for all x such that $\|x\| \leq 1$. In particular, this implies $\|T\| \leq \sqrt{\alpha\beta}$.

11.5 Adjoint of an Operator

Theorem 11.3. *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be Hilbert spaces and let $T\colon \mathcal{H}_1 \to \mathcal{H}_2$ be a bounded operator. There exists a unique operator $T^*\colon \mathcal{H}_2 \to \mathcal{H}_1$ such that*

$$\langle Tx, y\rangle_{\mathcal{H}_2} = \langle x, T^*y\rangle_{\mathcal{H}_1}$$

for all $x \in \mathcal{H}_1$ and $y \in \mathcal{H}_2$.

Proof. We define $T^*\colon \mathcal{H}_2 \to \mathcal{H}_1$ as follows: Let $z \in \mathcal{H}_2$ and set $\ell = \ell_z \circ T$. Thus $\ell\colon \mathcal{H}_2 \to \mathbb{C}$ is defined by

$$\ell(x) = \langle Tx, z\rangle_{\mathcal{H}_2}$$

for all $x \in \mathcal{H}_1$. Since ℓ is a composition of bounded operators, it must also be a bounded operator with

$$\|\ell\| \leq \|T\|\|\ell_z\| = \|T\|\|z\|.$$

By the Riesz representation theorem, there exists a unique $y \in \mathcal{H}_1$ such that $\ell = \ell_y$. We set

$$T^*z = y.$$

Observe that T^* is well-defined because of the uniqueness part of the Riesz representation theorem!

Let us show that T^* is $\mathbb{C}$-linear. Let $\alpha, \alpha' \in \mathbb{C}$ and let $z, z' \in \mathcal{H}_2$. Then we have

$$\begin{aligned}
\langle x, T^*(\alpha z + \alpha' z')\rangle &= \langle Tx, \alpha z + \alpha' z'\rangle \\
&= \overline{\alpha}\langle Tx, z\rangle + \overline{\alpha}'\langle Tx, z'\rangle \\
&= \overline{\alpha}\langle x, T^*z\rangle + \overline{\alpha}'\langle x, T^*z'\rangle \\
&= \langle x, \alpha T^*z + \alpha' T^*z'\rangle
\end{aligned}$$

for all $x \in \mathcal{H}_1$. It follows that
$$T^*(\alpha z + \alpha' z') = \alpha T^* z + \alpha' T^* z',$$
which implies T^* is $\mathbb{C}$-linear.

Next let us show that T^* is bounded. For any $z \in \mathcal{H}_2$, we have

$$\begin{aligned}
\|T^*z\|^2 &= \langle T^*z, T^*z\rangle \\
&= \langle TT^*z, z\rangle \\
&\leq \|T(T^*z)\|\|z\| \\
&\leq \|T\|\|T^*z\|\|z\|.
\end{aligned}$$

It follows that $\|T^*z\| \leq \|T\|\|z\|$ which implies T^* is bounded with $\|T^*\| \leq \|T\|$. □

Definition 11.4. The operator T^* given in Theorem (11.3) is called the **adjoint** of T.

Proposition 11.8. *Let $T\colon \mathcal{H}_1 \to \mathcal{H}_2$ be a bounded operator between Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$. Then $(T^*)^* = T$ and $\|T\| = \|T^*\|$.*

Proof. Let $x \in \mathcal{H}_1$. Then

$$\begin{aligned}
\langle x, (T^*)^*y\rangle &= \langle T^*x, y\rangle \\
&= \overline{\langle y, T^*x\rangle}. \\
&= \overline{\langle Ty, x\rangle} \\
&= \langle x, Ty\rangle.
\end{aligned}$$

Thus $(T^*)^*y = Ty$ for all $y \in \mathcal{H}_2$.

Now

$$\begin{aligned}
\|T\| &= \|(T^*)^*\| \\
&\leq \|T^*\|
\end{aligned}$$

implies $\|T\| \leq \|T^*\|$. Combining this with the previous theorem gives us $\|T^*\| \leq \|T\|$. □

Example 11.4. Let $S\colon \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$ be the shift operator, defined by

$$S((x_1, x_2, \dots)) = (0, x_1, x_2, \dots),$$

We compute S^*. We have

$$\langle S((x_n)), (y_n)\rangle = \langle ((x_n)), S^*(y_n)\rangle$$

if and ony if

$$\sum_{n=1}^{\infty} x_n y_{n+1} = \sum_{n=1}^{\infty} x_n \overline{(S^*((y_n))_n)}$$

Choose $(x_n) = e^k = (0, 0, \dots, 1, 0, 0, \dots)$. Then

$$\overline{y}_{k+1} = \overline{(S^*((y_n))}_k$$

which implies $S^*((y_n))_k = y_{k+1}$. So
$$S^*((y_1, y_2, \dots,)) = (y_2, y_3, \dots,).$$
We call S^* the **backwards shift operator**.

11.6 Self-Adjoint Operators

Definition 11.5. An operator $T\colon \mathcal{H} \to \mathcal{H}$ is said to be **self-adjoint** (or **symmetric, Hermitian**) if $T^* = T$, i.e.

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

for all $x, y \in \mathcal{H}$.

11.6.1 Examples and Non-examples of Self-Adjoint Operators

Example 11.5. Any orthogonal projection is self-adjoint.

Example 11.6. Let (a_n) be a sequence of complex numbers and let $T\colon \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$ be given by

$$T((x_n)) = (a_n x_n)$$

for all $(x_n) \in \ell^2(\mathbb{N})$. Then T is self-adjoint if and only if (a_n) is a sequence of real numbers. Indeed, if T is self-adjoint, then

$$\begin{aligned} a_n &= \langle a_n e^n, e^n \rangle \\ &= \langle T(e^n), e^n \rangle \\ &= \langle e^n, T(e^n) \rangle \\ &= \langle e^n, a_n e^n \rangle \\ &= \overline{a}_n \end{aligned}$$

for all $n \in \mathbb{N}$. Thus each a_n is real. Conversely, if each a_n is real, then

$$\begin{aligned} \langle T((x_n)), (y_n) \rangle &= \sum_{n=1}^{\infty} a_n x_n \overline{y}_n \\ &= \sum_{n=1}^{\infty} x_n \overline{a}_n \overline{y}_n \\ &= \langle (x_n), T((y_n)) \rangle. \end{aligned}$$

Example 11.7. The shift operator is not self-adjoint.

Example 11.8. Unitary operators are not self-adjoint.

Proposition 11.9. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be self-adjoint. Then*

$$\|T\| = \sup \{ |\langle Tx, x \rangle| \mid \|x\| \leq 1 \}.$$

Proof. First note that by Cauchy-Schwarz, we have

$$\begin{aligned} \sup \{ |\langle Tx, x \rangle| \mid \|x\| \leq 1 \} &\leq \sup \{ \|Tx\| \mid \|x\| \leq 1 \} \\ &= \|T\|. \end{aligned}$$

Conversely, let

$$M = \sup \{ |\langle Tx, x \rangle| \mid \|x\| \leq 1 \}.$$

Let $x, y \in \mathcal{H}$ with $\|x\|, \|y\| \leq 1$. Then

$$\begin{aligned} \langle T(x+y), x+y \rangle - \langle T(x-y), x-y \rangle &= \langle Tx, x \rangle + \langle Tx, y \rangle + \langle Ty, x \rangle + \langle Ty, y \rangle - \langle Tx, x \rangle + \langle Tx, y \rangle + \langle Ty, x \rangle - \langle Ty, y \rangle \\ &= \langle Tx, y \rangle + \langle Ty, x \rangle + \langle Tx, y \rangle + \langle Ty, x \rangle \\ &= 2\langle Tx, y \rangle + 2\overline{\langle x, Ty \rangle} \\ &= 2\langle Tx, y \rangle + 2\overline{\langle Tx, y \rangle} \\ &= 4\mathrm{Re}\langle Tx, y \rangle. \end{aligned}$$

Now observe that for any $z \in \mathcal{H}$ we have

$$\begin{aligned} |\langle Tz, z \rangle| &= |\langle \|z\| T(\frac{z}{\|z\|})|, \|z\| \frac{z}{\|z\|} \rangle \\ &= \|z\|^2 \langle T\left(\frac{z}{\|z\|}\right), \frac{z}{\|z\|} \rangle \\ &\leq \|z\|^2 M. \end{aligned}$$

Therefore

$$\begin{aligned}
4\text{Re}\langle Tx, y\rangle &\le |\langle T(x+y), x+y\rangle - \langle T(x-y), x-y\rangle| \\
&\le M(\|x+y\|^2 + \|x-y\|^2) \\
&= 2M(\|x\|^2 + \|y\|^2) \\
&\le 4M.
\end{aligned}$$

Thus for any $x, y \in \mathcal{H}$ with $\|x\|, \|y\| \le 1$, we proved

$$\text{Re}\langle Tx, y\rangle \le M. \tag{17}$$

Setting $y = Tx/\|Tx\|$, Then plugging in y into (17), we obtain

$$\begin{aligned}
\|Tx\| &\le= \frac{1}{\|Tx\|}\text{Re}\langle \|Tx\|^2\rangle \\
&= \text{Re}\langle Tx, \frac{Tx}{\|Tx\|}\rangle \\
&\le M.
\end{aligned}$$

Thus $\|T\| \le M$. □

12 Compactness

In topology, one studies a class of spaces called **compact** spaces. Recall that a toplogical space X is said to be compact if it satisfies the following property: every open cover of X contains a finite subcover of X. In other words, X is compact if for all open covers $\{U_i\}_{i\in I}$ of X, where by an open cover we mean each U_i is an open subset of X and

$$X = \bigcup_{i\in I} U_i,$$

then there exists $U_{i_1}, \dots, U_{i_n} \in \{U_i\}_{i\in I}$ such that $\{U_{i_1}, \dots, U_{i_n}\}$ is an open cover of X, that is

$$X = U_{i_1} \cup \cdots \cup U_{i_n}.$$

There is an analagous notion of compactness called **sequential compactness**. Recall that a topological space X is said to be sequentially compact if every sequence of points in X has a convergent subsequence converging to a point in X. It turns out that for metric space, compactness and sequintial compactness are equivalent. Since we are always talking about sequences in linear analysis, it makes sense for us to study compact spaces as sequentially compact spaces. In particular, we make the following definitions:

Definition 12.1. Let $\mathcal{V}$ be an inner-product space and let $K \subseteq \mathcal{V}$.

1. We say K is **precompact** if every sequence in K has a convergent sequence.
2. We say K is **compact** if every sequence in K has a convergent sequence with a limit in K.

Precompactness can be thought of as "almost" compact. In fact, we have the following proposition.

Proposition 12.1. *Let $\mathcal{H}$ be an inner-product space and let $A \subseteq \mathcal{H}$. Then A is precompact if and only if $\overline{A}$ is compact.*

Proof. Suppose A is precompact. Let (a_n) be a sequence in $\overline{A}$. For each $n \in \mathbb{N}$ choose $b_n \in A$ such that

$$\|a_n - b_n\| < \frac{1}{n}.$$

Since A is precompact, there exists a convergent subsequence of (b_n), say $(b_{\pi(n)})$. We claim that the subsequence $(a_{\pi(n)})$ of (a_n) is Cauchy. Indeed, let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $\pi(n) \ge \pi(m) \ge N$ implies

$$\|b_{\pi(n)} - b_{\pi(m)}\| < \frac{\varepsilon}{3} \quad \text{and} \quad \frac{1}{\pi(m)} < \frac{\varepsilon}{3}.$$

Then $\pi(n) \geq \pi(m) \geq N$ implies

$$\begin{aligned}
\|a_{\pi(n)} - a_{\pi(m)}\| &= \|a_{\pi(n)} - b_{\pi(n)} + b_{\pi(n)} - b_{\pi(m)} + b_{\pi(m)} - a_{\pi(m)}\| \\
&\leq \|a_{\pi(n)} - b_{\pi(n)}\| + \|b_{\pi(n)} - b_{\pi(m)}\| + \|b_{\pi(m)} - a_{\pi(m)}\| \\
&< \frac{1}{\pi(n)} + \frac{\varepsilon}{3} + \frac{1}{\pi(m)} \\
&\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \\
&= \varepsilon.
\end{aligned}$$

This implies $(a_{\pi(n)})$ is Cauchy. Finally, since $(a_{\pi(n)})$ is Cauchy and since $\mathcal{H}$ is a Hilbert space, we must have $a_{\pi(n)} \to a$ for some $a \in \overline{A}$. This implies $\overline{A}$ is compact.

Conversely, suppose $\overline{A}$ is compact. Let (a_n) be a sequence in A. Then (a_n) is a sequence in $\overline{A}$. Since $\overline{A}$ is compact, the sequence (a_n) has a convergent subsequence. This implies A is precompact. □

12.0.1 Compactness = Sequential Compactness in Metric Spaces

As we mentioned before, sequential compactness and compactness are equivalent notions when it comes to metric spaces. It takes some work to show that sequential compactness implies compactness, so we will save that for later. On the other hand, we can prove that compactness implies sequential compactness relatively easily:

Proposition 12.2. *Let K be a compact subspace of $\mathcal{V}$. Then K is sequentially compact.*

Proof. Let (x_n) be a sequence in K. Assume for a contradiction that (x_n) does not have a convergent subsequence with a limit in K. We claim that for each $x \in K$, we can find an open neighborhood U_x of x such that only finitely many elements in the sequence (x_n) belongs to U_x. Indeed, assume for a contradiction that every open neighborhood of x contains infinitely many terms of (x_n). Then for each $n \in \mathbb{N}$, we can choose an $x_{\pi(n)}$ such that $x_{\pi(n)} \in B_{1/n}(x)$. Then the subsequence $(x_{\pi(n)})$ is a Cauchy sequence which converges to x, and this contradicts our first assumption.

Thus for each $x \in X$ we may choose an open neighborhood U_x of x such that only finitely many elements in the sequence (x_n) belongs to U_x. Since K is compact, the open cover $\{U_x\}_{x \in X}$ of K has a finite subcover, say $\{U_{x_1}, \dots, U_{x_k}\}$. But then each U_{x_i} contains only finitely many elements in the sequence (x_n), so the sequence only contains finitely many distinct elements. Such a sequence clearly has a convergent subsequence with a limit in K. This gives us our desired contradiction. □

12.0.2 Failure of Heine-Borel Theorem in the Infinite-Dimensional Setting

As we've mentioned before, sequential compactness and compactness are equivalent notions when it comes to metric spaces. It turns out that there is another description of compactness when it comes to spaces like $\mathbb{R}$, $\mathbb{C}$, $\mathbb{R}^n$, and $\mathbb{C}^n$. More generally, if $\mathcal{H}$ is a finite-dimensional Hilbert space and K is a subset of $\mathcal{H}$, then K is compact if and only if it is closed and bounded. This is essentially the content of the Heine-Borel Theorem. On the other hand, closed and bound subspaces of infinite-dimensional Hilbert spaces need not be compact subspaces. To see what goes wrong, consider the closed unit ball $\mathrm{B}_1[0]$ (centered at 0 and of radius 1) in $\ell^2(\mathbb{N})$. Clearly, $\mathrm{B}_1[0]$ is closed and bounded. However it is not compact. Indeed, the sequence $(\mathbf{e}_n)$ of standard coordinate vectors in $\mathrm{B}_1[0]$ cannot have a convergent subsequence. Indeed, the distance between any two standard coordinate vectors $\mathbf{e}_n$ and $\mathbf{e}_m$ is $\sqrt{2}$. Thus, any subsequence of $(\mathbf{e}_n)$ will fail to be Cauchy, and hence will not converge.

Even though closed and bound subspaces of infinite-dimensional Hilbert spaces need not be compact subspaces, the converse still holds. In other words, being closed and bounded is a necessary condition for a subspace to be compact, though it is not sufficient.

Proposition 12.3. *Let $\mathcal{H}$ be a Hilbert space and let $A \subseteq \mathcal{H}$ be a compact subspace. Then A is closed and bounded.*

Proof. In any Hausdorff space X (for example a metric space), a compact subspace $K \subseteq X$ is necessarily a closed subset of X. We leave the proof of this as an exercise for the reason. Let us prove that A is a bounded. Assume for a contradiction that A is not bounded. For each $a \in A$, let $U_a = \mathrm{B}_1(a)$ be the open ball of radius 1 centered at a. Then $\{U_a\}_{a \in A}$ forms an open cover of A. Since A is compact, there exists a finite subcover of $\{U_a\}$, say $\{U_{a_1}, \dots, U_{a_n}\}$. For each $1 \leq i, j \leq n$, set

$$L_{ij} = \mathrm{d}(U_{a_i}, U_{a_j}) = \sup\left\{\|x_i - x_j\| \mid x_i \in U_{a_i} \text{ and } x_j \in U_{a_j}\right\}$$

Clearly L_{ij} is finite since $L_{ij} \leq \|a_i - a_j\| + 2$. Setting $L = \max_{1 \leq i,j \leq n}\{L_{ij}\}$, we see that for all $x, x' \in A$, we must have $\|x - x'\| \leq L$. Thus, A is bounded. □

12.1 Weak Convergence

As we've seen, the sequence $(\mathbf{e}_n)$ in $\ell^2(\mathbb{N})$ has not convergent subsequences, so obviously the sequence $(\mathbf{e}_n)$ doesn't converge. On the other hand, there is a weaker notion of convergence in which the sequence $(\mathbf{e}_n)$ does converge. Let us discuss this weakened version of convergence now.

Definition 12.2. Let $\mathcal{V}$ be an inner-product space and let (x_n) be a sequence in $\mathcal{V}$. We say (x_n) **converges weakly** to an element $x \in \mathcal{V}$, which we denote by $x_n \xrightarrow{w} x$, if the sequence of complex numbers $(\langle x_n, y\rangle)$ converges (in the usual sense) to $\langle x, y\rangle \in \mathbb{C}$ for all $y \in \mathcal{V}$. In this case, we call x the **weak limit** of (x_n).

Remark 21. If $\langle x_n, y\rangle \to \langle x, y\rangle$, then $\langle y, x_n\rangle \to \langle y, x\rangle$. This is because

$$\begin{aligned}
\lim_{n\to\infty}\langle y, x_n\rangle &= \lim_{n\to\infty}\overline{\langle x_n, y\rangle} \\
&= \overline{\lim_{n\to\infty}\langle x_n, y\rangle} \\
&= \overline{\langle x, y\rangle} \\
&= \langle y, x\rangle,
\end{aligned}$$

where we are allowed to pull the conjugation outside of the limit operator since the conjugation function is continuous everywhere. A similar argument shows that if $\langle y, x_n\rangle \to \langle y, x\rangle$, then $\langle x_n, y\rangle \to \langle x, y\rangle$. Thus $\langle x_n, y\rangle \to \langle x, y\rangle$ if and only if $\langle y, x_n\rangle \to \langle y, x\rangle$.

Remark 22. Note that weak limits are unique. Indeed, this follows from positive-definiteness of the inner-product. If $\langle x_n, y\rangle \to \langle x, y\rangle$ and $\langle x_n, y\rangle \to \langle x', y\rangle$ for all $y \in \mathcal{V}$, then $\langle x, y\rangle = \langle x', y\rangle$ for all $y \in \mathcal{V}$, which implies $x = x'$ by positive-definiteness of the inner-product.

Example 12.1. Let $\mathcal{H}$ be an infinite-dimensional separable Hilbert space and let (e_n) be an orthonormal sequence in $\mathcal{H}$. We claim that $e_n \xrightarrow{w} 0$. Indeed, let $y \in \mathcal{H}$. Then since $\sum_{n=1}^{\infty}\langle y, e_n\rangle e_n$ is convergent, we must in particular have

$$\lim_{n\to\infty}\langle y, e_n\rangle = 0 = \langle y, 0\rangle.$$

Note that we need

Remark 23. Any orthonormal sequence is weakly convergent to 0.

We will prove that any bounded sequence in an infinite dimensional Hilbert space has a weakly convergent subsequence.

12.1.1 Weak Convergence Properties

Proposition 12.4. *Let $\mathcal{H}$ be an infinite dimensional Hilbert space. Then $x_n \to x$ implies $x_n \xrightarrow{w} x$. Conversely, if $x_n \xrightarrow{w} x$ and $\|x_n\| \to \|x\|$, then $x_n \to x$.*

Remark 24. In finite dimensional spaces we have $x_n \to x$ if and only if $x_n \xrightarrow{w} x$.

Proof. Suppose $x_n \to x$. Then for all $y \in \mathcal{H}$, we have

$$\begin{aligned}
|\langle x_n, y\rangle - \langle x, y\rangle| &= |\langle x_n - x, y\rangle| \\
&\leq \|x_n - x\|\|y\| \\
&\to 0
\end{aligned}$$

as $n \to \infty$. It follows that $x_n \xrightarrow{w} x$.

Converesely, suppose $x_n \xrightarrow{w} x$ and $\|x_n\| \to \|x\|$. Since $x_n \xrightarrow{w} x$, we have in particular $\langle x_n, x\rangle \to \langle x, x\rangle$ and $\langle x, x_n\rangle \to \langle x, x\rangle$. Thus

$$\begin{aligned}
\|x_n - x\|^2 &= \langle x_n, x_n\rangle - \langle x_n, x\rangle - \langle x, x_n\rangle + \langle x, x\rangle \\
&\to \|x\|^2 - \langle x, x\rangle - \langle x, x\rangle + \langle x, x\rangle \\
&= 0.
\end{aligned}$$

It follows that $x_n \to x$. $\square$

Note that we really do need both $x_n \xrightarrow{w} x$ *and* $\|x_n\| \to \|x\|$ for the converse to be true. Indeed, let (x_n) be any orthonormal sequence in $\mathcal{H}$. Then for any $y \in \mathcal{H}$ we have the Bessel inequality

$$\sum_{n=1}^{\infty} |\langle y, x_n \rangle|^2 \leq \|y\|^2.$$

In particular, the series on the left is convergent. Therefore $\langle x_n, y \rangle \to 0 = \langle 0, y \rangle$, and hence $x_n \xrightarrow{w} 0$. However, clearly (x_n) is not a convergent sequence since $\|x_n\| = 1$ for all $n \in \mathbb{N}$.

12.1.2 Weak convergence in Separable Hilbert Space

Lemma 12.1. *Let $\mathcal{H}$ be a separable Hilbert space, let (e_m) be an othonormal sequence in $\mathcal{H}$, let (x_n) be a bounded sequence in $\mathcal{H}$, and let $x \in \mathcal{H}$. Then $x_n \xrightarrow{w} x$ if and only if $\langle x_n, e_m \rangle \to \langle x, e_m \rangle$ for all $m \in \mathbb{N}$.*

Proof. Suppose $x_n \xrightarrow{w} x$. Then $\langle x_n, y \rangle \to \langle x, y \rangle$ for all $y \in \mathcal{H}$, so certainly $\langle x_n, e_m \rangle \to \langle x, e_m \rangle$ for all $m \in \mathbb{N}$. Conversely, suppose $\langle x_n, e_m \rangle \to \langle x, e_m \rangle$ for all $m \in \mathbb{N}$.

Step 1: We first show that $\langle x_n, y \rangle \to \langle x, y \rangle$ for all $y \in E$, where $E = \operatorname{span}\{e_m \mid m \in \mathbb{N}\}$, so let $y \in E$. Then

$$y = \lambda_1 e_{m_1} + \cdots + \lambda_r e_{m_r}$$

for some (unique) $\lambda_1, \ldots, \lambda_r \in \mathbb{C}$ and $m_1, \ldots, m_r \in \mathbb{N}$. Then

$$\begin{aligned}
\lim_{n\to\infty} \langle x_n, y \rangle &= \lim_{n\to\infty} \langle x_n, \lambda_1 e_{m_1} + \cdots + \lambda_r e_{m_r} \rangle \\
&= \overline{\lambda}_1 \lim_{n\to\infty} \langle x_n, e_{m_1} \rangle + \cdots + \overline{\lambda}_r \lim_{n\to\infty} \langle x_n, e_{m_r} \rangle \\
&= \overline{\lambda}_1 \langle x, e_{m_1} \rangle + \cdots + \overline{\lambda}_r \langle x, e_{m_r} \rangle \\
&= \langle x, \lambda_1 e_{m_1} + \cdots + \lambda_r e_{m_r} \rangle \\
&= \langle x, y \rangle.
\end{aligned}$$

Step 2: Now we will show that $\langle x_n, z \rangle \to \langle x, z \rangle$ for all $z \in \mathcal{H}$, so let $z \in \mathcal{H}$. Let $\varepsilon > 0$. Choose $M \geq 0$ such that $\|x_n\| \leq M$ for all $n \in \mathbb{N}$ (we can do this by the assumption that (x_n) is bounded). Choose an element $y \in \operatorname{span}\{e_m \mid m \in \mathbb{N}\}$ such that $\|y - z\| < \frac{\varepsilon}{3\max(\|x\|, M)}$ (we can do this since $\mathcal{H}$ is separable). Finally, choose $N \in \mathbb{N}$ such that $n \geq N$ implies $|\langle x_n - x, y \rangle| < \varepsilon/3$ (we can do this by step 1). Then $n \geq N$ implies

$$\begin{aligned}
|\langle x_n, z \rangle - \langle x, z \rangle| &= |\langle x_n, z \rangle - \langle x_n, y \rangle + \langle x_n, y \rangle - \langle x, y \rangle + \langle x, y \rangle - \langle x, z \rangle| \\
&= |\langle x_n, z - y \rangle + \langle x_n - x, y \rangle + \langle x, y - z \rangle| \\
&\leq |\langle x_n, z - y \rangle| + |\langle x_n - x, y \rangle| + |\langle x, y - z \rangle| \\
&\leq \|x_n\| \|z - y\| + |\langle x_n - x, y \rangle| + \|x\| \|y - z\| \\
&< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 \\
&= \varepsilon.
\end{aligned}$$

$\square$

12.2 Baby Version of the Uniform Boundedness Principle and its Consequences

12.2.1 Baby Version of the Uniform Boundedness Principle

Lemma 12.2. *Let (x_n) be a sequence in $\mathcal{H}$. Assume there exists an $M > 0$ and a closed ball $\mathrm{B}_r[a] \subseteq \mathcal{H}$ such that*

$$|\langle x_n, y \rangle| \leq M$$

for all $n \in \mathbb{N}$ and for all $y \in \mathrm{B}_r[a]$. Then the sequence (x_n) is bounded.

Proof. The key is to translate everything from the closed ball $\mathrm{B}_r[a]$ to the closed ball $\mathrm{B}_1[0]$. Let $z \in \mathrm{B}_1[0]$. Then

$$\begin{aligned}
|\langle x_n, z\rangle| &= \frac{1}{r}|\langle x_n, rz\rangle| \\
&= \frac{1}{r}|\langle x_n, rz + a - a\rangle| \\
&= \frac{1}{r}|\langle x_n, rz + a\rangle - \langle x_n, a\rangle| \\
&\leq \frac{1}{r}|\langle x_n, rz + a\rangle| + \frac{1}{r}|\langle x_n, a\rangle \\
&\leq \frac{1}{r}M + \frac{1}{r}M \\
&= \frac{2M}{r}
\end{aligned}$$

for all $n \in \mathbb{N}$. In particular, fixing x_n and setting $z = x_n/\|x_n\|$, we have $\|x_n\| \leq 2M/r$. Since x_n was arbitrary, we have $\|x_n\| \leq 2M/r$ for all $n \in \mathbb{N}$. □

Theorem 12.3. *(Uniform Boundedness Principle) Let $\mathcal{H}$ be a Hilbert space and let (x_n) be a sequence in $\mathcal{H}$. Assume that for every $y \in \mathcal{H}$ there exists an $M_y > 0$ such that*

$$|\langle x_n, y\rangle| \leq M_y \tag{18}$$

for all $n \in \mathbb{N}$. Then the sequence (x_n) is bounded.

Proof. We claim that there exists $M > 0$ and a closed ball $\mathrm{B}_r[a] \subseteq \mathcal{H}$ such that

$$|\langle x_n, y\rangle| \leq M$$

for all $n \in \mathbb{N}$ and for all $y \in \mathrm{B}_r[a]$. This will imply (x_n) is bounded by Lemma (12.2).

For each $m \in \mathbb{N}$, we set

$$E_m = \{y \in \mathcal{H} \mid |\langle x_n, y\rangle| \leq m \text{ for all } n \in \mathbb{N}\}.$$

Observe that (E_m) is an ascending sequence of closed sets. Indeed, it is clearly ascending, to see that E_m is closed, view it as an infinite intersection of closed sets, namely

$$E_m = \bigcap_{n=1}^{\infty} \{y \in \mathcal{H} \mid |\langle x_n, y\rangle| \leq m\}.$$

Moreover, for any $y \in \mathcal{H}$, (18) implies $y \in \bigcup_{m=1}^{\infty} E_m$. Thus

$$\mathcal{H} = \bigcup_{m=1}^{\infty} E_m.$$

Now to prove the claim, it suffices to show that one of the E_m's contains an open ball of the form $\mathrm{B}_r(a)$ (the closure $\mathrm{B}_r[a]$ will then also belong to E_m since E_m is itself closed). Let $\mathrm{B}_{r_1}(a_1)$ be any open ball. If $\mathrm{B}_{r_1}(a_1) \subseteq E_1$, then we are done, so assume $\mathrm{B}_{r_1}(a_1) \cap E_1^c \neq \emptyset$. Choose $a_2 \in \mathcal{H}$ and $r_2 > 0$ such that $r_2 < r_1/2$ and $\mathrm{B}_{r_2}[a_2] \subseteq \mathrm{B}_{r_1}(a_1) \cap E_1^c$: we can find an open ball $\mathrm{B}_{r_2}(a_2)$ which is contained in $\mathrm{B}_{r_1}(a_1) \cap E_1^c$ since the intersection is a nonempty open set, and by replacing r_2 by a smaller positive number if necessary, we may assume that $\mathrm{B}_{r_2}[a_2]$ is contained in $\mathrm{B}_{r_2}[a_2] \cap E_1^c$). Continuing this process, we will either stop after finitely many iterations (and we'd be done!) or we can construct a sequence (a_n) in $\mathcal{H}$ and a sequence $(r_n) \in \mathbb{R}$ such that

$$r_{n+1} < r_n/2 < \cdots < r_1/2^n \quad \text{and} \quad \mathrm{B}_{r_{n+1}}[a_{n+1}] \subseteq \mathrm{B}_{r_n}(a_n) \cap E_n^c$$

for all $n \in \mathbb{N}$.

Assume for a contradiction that such a sequence has been constructed. We claim that (a_n) is a Cauchy sequence. Indeed, let $\varepsilon > 0$. Since $r_n \to 0$, there exists $N \in \mathbb{N}$ such that $n \geq N$ implies $2r_n < \varepsilon$. Then $n, m \geq N$ implies

$$\begin{aligned}
\|a_n - a_m\| &\leq \|a_n - a_N\| + \|a_N - a_m\| \\
&< r_N + r_N \\
&= 2r_N \\
&< \varepsilon.
\end{aligned}$$

since $a_n, a_m \in \mathrm{B}_{r_N}[a_N]$ for all $n, m \geq N$. Thus (a_n) is a Cauchy sequence and hence converges (since we are in a Hilbert space) say to $a \in \mathcal{H}$. Now observe that for any $m \in \mathbb{N}$, we have $a \in \mathrm{B}_{r_m}[a_m] \subseteq E_{m-1}^c$. In particular

$$\begin{aligned} a &\in \bigcap_{m=1}^{\infty} E_m^c \\ &= \left(\bigcup_{m=1}^{\infty} E_m \right)^c \\ &= \mathcal{H}^c \\ &= \emptyset, \end{aligned}$$

which is a contradiction. □

Corollary 6. *Let $\mathcal{H}$ be a Hilbert space, let (x_n) be a sequence of elements in $\mathcal{H}$, and let $x \in \mathcal{H}$. If $x_n \xrightarrow{w} x$, then the sequence (x_n) is bounded. Moreover, we have*

$$\|x\| \leq \liminf_{n \to \infty} \|x_n\|. \tag{19}$$

Proof. Let $y \in \mathcal{H}$. Since the sequence $(\langle x_n, y \rangle)$ of complex numbers is convergent, it must be bounded. Thus there exists $M_y > 0$ such that $|\langle x_n, y \rangle| \leq M_y$ for all $n \in \mathbb{N}$. It follows from Theorem (12.3) that the sequence (x_n) is bounded. For the last part, we have

$$\begin{aligned} \|x\|^2 &= |\langle x, x \rangle| \\ &= \lim_{n \to \infty} |\langle x_n, x \rangle| \\ &\leq \liminf_{n \to \infty} \|x_n\| \|x\| \\ &= \|x\| \liminf_{n \to \infty} \|x_n\|, \end{aligned}$$

which implies (19). □

12.2.2 Weak Convergence Plus Convergence Implies Convergence

Proposition 12.5. *Let $\mathcal{H}$ be a Hilbert space. If $x_n \xrightarrow{w} x$ and $y_n \to y$. Then $\langle x_n, y_n \rangle \to \langle x, y \rangle$.*

Proof. Choose $M \geq 0$ such that $\|x_n\| \leq M$ for all $n \in \mathbb{N}$ (we can do this by the previous theorem). Let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies $\|y - y_n\| < \varepsilon/2M$ and $|\langle x_n, y \rangle - \langle x, y \rangle| < \varepsilon/2$. We have

$$\begin{aligned} |\langle x_n, y_n \rangle - \langle x, y \rangle| &\leq |\langle x_n, y_n \rangle - \langle x_n, y \rangle| + |\langle x_n, y \rangle - \langle x, y \rangle| \\ &= |\langle x_n, y - y_n \rangle| + |\langle x_n, y \rangle - \langle x, y \rangle| \\ &\leq \|x_n\| \|y - y_n\| + \frac{\varepsilon}{2} \\ &< M \frac{\varepsilon}{2M} + \frac{\varepsilon}{2} \\ &= \varepsilon. \end{aligned}$$

□

12.2.3 Every Bounded Sequence in Hilbert Space has Weakly Convergent Subsequence

Lemma 12.4. *Let $\mathcal{H}$ be a separable Hilbert space. Then there exists a countable dense subset of $\mathcal{H}$.*

Proof. Choose an orthonormal basis (e_n) of $\mathcal{H}$. Let

$$\mathcal{Y} = \operatorname{span}_{\mathbb{Q}(i)}\{e_n \mid n \in \mathbb{N}\} = \{\lambda_1 e_{n_1} + \cdots + \lambda_k e_{n_k} \mid \lambda_1, \ldots, \lambda_k \in \mathbb{Q}(i)\}.$$

Then $\mathcal{Y}$ is countable since $\mathbb{Q}(i)$ is countable. Moreover, $\mathcal{Y}$ is dense in $\mathcal{H}$ since $\mathbb{Q}(i)$ is dense in $\mathbb{C}$ and since

$$\operatorname{span}_{\mathbb{C}}\{e_n \mid n \in \mathbb{N}\} = \{\lambda_1 e_{n_1} + \cdots + \lambda_k e_{n_k} \mid \lambda_1, \ldots, \lambda_k \in \mathbb{C}\}$$

is dense in $\mathcal{H}$. □

Remark 25. In general, a topological space X is said to be **separable** if it contains a countable dense subset. Every continuous function on a separable space whose image is a subset of a Hausdorff space is determined by its values on the countable dense subset.

Theorem 12.5. *Let $\mathcal{H}$ be an infinite dimensional separable Hilbert space. Then any bounded sequence in $\mathcal{H}$ has a weakly convergent subsequence.*

Remark 26. If $\mathcal{H}$ is finite-dimensional, then every bounded sequence has a convergent subsequence. This is a consequence of the Bolzano-Weierstrass Theorem. However this theorem is no longer true in infinite dimensions.

Proof. Let (x_n) be a bounded sequence in $\mathcal{H}$. Choose a countably dense susbset $\mathcal{Y}$ of $\mathcal{H}$ and order the elements in $\mathcal{Y}$ as a sequence, say $\mathcal{Y} = (y_m)$. Choose $M \geq 0$ such that $\|x_n\| \leq M$ for all $n \in \mathbb{N}$. Consider the sequence $(\langle y_1, x_n \rangle)$ of complex numbers. Since

$$\begin{aligned} |\langle y_1, x_n \rangle| &\leq \|y_1\| \|x_n\| \\ &\leq \|y_1\| M \end{aligned}$$

for all $n \in \mathbb{N}$, we see that the sequence $(\langle y_1, x_n \rangle)$ is bounded. By the Bolzano-Weierstrass Theorem, it must have a convergent subsequence, say $(\langle y_1, x_{\pi_1(n)} \rangle)$[1] where $\langle y_1, x_{\pi_1(n)} \rangle \to \lambda_1$ for some (uniquely determined) $\lambda_1 \in \mathbb{C}$. Repeating the same argument for the bounded sequence $(\langle y_2, x_{\pi_1(n)} \rangle)$, we can find a convergent subsequence, say $(\langle y_2, x_{\pi_2(n)} \rangle)$ where $\langle y_2, x_{\pi_2(n)} \rangle \to \lambda_2$ for some (uniquely determined) $\lambda_2 \in \mathbb{C}$ *and* $\langle y_1, x_{\pi_2(n)} \rangle \to \lambda_1$ since $(\langle y_1, x_{\pi_2(n)} \rangle)$ is a subsequence of $(\langle y_1, x_{\pi_1(n)} \rangle)$ and hence must converge to the same limit. Repeating this process for each $m \in \mathbb{N}$, we can find a subsequence $(\langle y_m, x_{\pi_m(n)} \rangle)$ such that $\langle y_m, x_{\pi_m(n)} \rangle \to \lambda_m$ for some (uniquely determined) $\lambda_m \in \mathbb{C}$ *and* $\langle y_k, x_{\pi_m(n)} \rangle \to \lambda_k$ for all $1 \leq k < m$. Now we apply the diagonal trick. Define $\pi \colon \mathbb{N} \to \mathbb{N}$ by $\pi(n) = \pi_n(n)$ for all $n \in \mathbb{N}$ and consider the sequence $(x_{\pi(n)})$. Observe that the sequence $(x_{\pi(n)})$ is a subsequence of (x_n). Moreover, $(x_{\pi(n)})$ is essentially a subsequence of $(x_{\pi_m(n)})$ (minus the first m terms) for all $m \in \mathbb{N}$. Therefore $\langle y_m, x_{\pi(n)} \rangle \to \lambda_m$ for all $m \in \mathbb{N}$. In particular, for each $m \in \mathbb{N}$, the sequence $(\langle y_m, x_{\pi(n)} \rangle)$ is Cauchy.

Let $y \in \mathcal{H}$. We claim that the sequence $(\langle y, x_{\pi(n)} \rangle)$ of complex numbers is Cauchy. Indeed, let $\varepsilon > 0$. Choose $m_0 \in \mathbb{N}$ such that

$$\|y_{m_0} - y\| < \frac{\varepsilon}{3M}$$

(we can do this since $\mathcal{Y}$ is dense in $\mathcal{H}$). Choose $N \in \mathbb{N}$ such that $m, n \geq N$ implies

$$|\langle y_{m_0}, x_{\pi(n)} \rangle - \langle y_{m_0}, x_{\pi(m)} \rangle| < \frac{\varepsilon}{3}$$

(we can do this since the sequence $(\langle y_{m_0}, x_{\pi(n)} \rangle)$ is Cauchy). Then $m, n \geq N$ implies

$$\begin{aligned} |\langle y, x_{\pi(n)} \rangle - \langle y, x_{\pi(m)} \rangle| &= |\langle y, x_{\pi(n)} \rangle - \langle y_{m_0}, x_{\pi(n)} \rangle + \langle y_{m_0}, x_{\pi(n)} \rangle - \langle y_{m_0}, x_{\pi(m)} \rangle + \langle y_{m_0}, x_{\pi(m)} \rangle - \langle y, x_{\pi(m)} \rangle| \\ &\leq |\langle y - y_{m_0}, x_{\pi(n)} \rangle| + |\langle y_{m_0}, x_{\pi(n)} \rangle - \langle y_{m_0}, x_{\pi(m)} \rangle| + |\langle y_{m_0} - y, x_{\pi(m)} \rangle| \\ &\leq \|y - y_{m_0}\| M + |\langle y_{m_0}, x_{\pi(n)} \rangle - \langle y_{m_0}, x_{\pi(m)} \rangle| + \|y_{m_0} - y\| M \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \\ &= \varepsilon. \end{aligned}$$

Thus the sequence $(\langle y, x_{\pi(n)} \rangle)$ is Cauchy.

Since the sequence $(\langle y, x_{\pi(n)} \rangle)$ is a Cauchy sequence of complex numbers for each $y \in \mathcal{H}$, we are justified in defining $\ell \colon \mathcal{H} \to \mathbb{C}$ by

$$\ell(y) = \lim_{n \to \infty} \langle y, x_{\pi(n)} \rangle$$

for all $y \in \mathcal{H}$. The map ℓ is linear since the limit operator is linear and since the inner-product is linear in the first argument. The map ℓ is also bounded since

$$\begin{aligned} |\ell(y)| &= |\lim_{n \to \infty} \langle y, x_{\pi(n)} \rangle| \\ &= \lim_{n \to \infty} |\langle y, x_{\pi(n)} \rangle| \\ &\leq \lim_{n \to \infty} \|y\| M \\ &= \|y\| M \end{aligned}$$

[1] Here we view π_1 as a strictly increasing function from $\mathbb{N}$ to $\mathbb{N}$ whose range consists of the indices in the subsequence.

for all $y \in \mathcal{H}$. By the Riesz Representation Theorem, there is a unique $x \in \mathcal{H}$ such that

$$\begin{aligned}\langle y, x \rangle &= \ell(y) \\ &= \lim_{n\to\infty} \langle y, x_{\pi(n)} \rangle\end{aligned}$$

for all $y \in \mathcal{H}$. In other words, there exists a (necessarily unique) $x \in \mathcal{H}$ such that $x_{\pi(n)} \xrightarrow{w} x$. □

Corollary 7. *Let $\mathcal{H}$ be an infinite-dimensional Hilbert space. Then every bounded sequence has a weakly convergent subsequence.*

Proof. Let (x_n) be a bounded sequence in $\mathcal{H}$. Define

$$\mathcal{H}_0 := \overline{\operatorname{span}}\{x_n \mid n \in \mathbb{N}\}.$$

Then (x_n) is a bounded sequence in the separable Hilbert space $\mathcal{H}_0$. Therefore by Theorem (12.5), there exists a bounded subsequence, say $(x_{\pi(n)})$, of the sequence (x_n) which weakly converges to some element, say x, in $\mathcal{H}_0$. Since $\mathcal{H}_0$ is a subspace of $\mathcal{H}$, it follows that $(x_{\pi(n)})$ is a subsequence of (x_n) which weakly converges to the element x, where $x \in \mathcal{H}_0 \subseteq \mathcal{H}$. □

12.3 Compact Operators

Definition 12.3. An operator $T \colon \mathcal{H} \to \mathcal{H}$ is said to be **compact** (also called **completely continuous**) if for any sequence (x_n) of elements in $\mathcal{H}$ such that $x_n \xrightarrow{w} x$, we have $Tx_n \to Tx$.

Remark 27. Thus compact operators improve convergence. Note that compact operators are continuous, and hence bounded. The set of all compact operators from $\mathcal{H}_1$ to $\mathcal{H}_2$ is denoted $\mathcal{C}(\mathcal{H}_1, \mathcal{H}_2)$. Note that $\mathcal{C}(\mathcal{H}, \mathcal{H})$ is a maximal ideal in $\mathcal{B}(\mathcal{H}, \mathcal{H})$.

Definition 12.4. A bounded operator T with finite-dimensional range is called a **finite rank operator**.

Proposition 12.6. *Every **finite rank operator** is compact.*

Proof. Let $T \colon \mathcal{H} \to \mathcal{H}$ be a finite rank operator. Let (x_n) be a sequence of elements in $\mathcal{H}$ such that $x_n \xrightarrow{w} x$. Then $Tx_n \xrightarrow{w} Tx$ since T is bounded. Let $e_1, \dots, e_k$ be an orthonormal basis for $\operatorname{im}(T)$. Then

$$\begin{aligned}Tx_n &= \langle Tx_n, e_1 \rangle e_1 + \cdots + \langle Tx_n, e_k \rangle e_k \\ &\to \langle Tx, e_1 \rangle e_1 + \cdots + \langle Tx, e_k \rangle e_k \\ &= Tx.\end{aligned}$$

□

12.4 Eigenvalues

Let T be a linear map from a complex vector space V to itself. Recall from linear algebra that $\lambda \in \mathbb{C}$ is said to be an **eigenvalue of** T if there exists a nonzero $v \in V$ such that $Tv = \lambda v$. In this case, we call v an **eigenvector of** T **corresponding to the eigenvalue** λ. We denote by Λ to be the set of all eigenvalues of T and we denote by E_λ to be the set of all eigenvectors of T corresponding to λ. Observe that $E_\lambda = \ker(\lambda - T)$, hence E_λ is in fact a subspace of V. We call this subspace the **eigenspace of** T **corresponding to** λ. When context is clear, we often refer to λ, v, and E_λ as "an eigenvalue", "an eigenvector", and "an eigenspace" respectively.

We'd like to have a good understanding of Λ. If V is finite dimensional, say $\dim V = n$, then the eigenvalues of T are completely characterized by the **characteristic polynomial of** T which is defined by the equation

$$\chi_T(X) = \det(X I_n - [T]_\beta) \tag{20}$$

where $[T]_\beta$ is the matrix representation of T with respect to some basis β of V. Indeed, λ is an eigenvalue of T if and only if λ is a root of $\chi_T(X)$. Any polynomial in $\mathbb{C}[X]$ splits in $\mathbb{C}$ since $\mathbb{C}$is algebraically closed, and therefore $\chi_T(X)$ factors as

$$\chi_T(X) = \prod_{\lambda \in \Lambda} (X - \lambda)^{m_T(\lambda)} \tag{21}$$

where $m_T(\lambda)$ is called the **algebraic multiplicity of** λ. Therefore if we have a good understanding $\chi_T(X)$, then we will have a good understanding Λ as well.

If V is infinite dimensional, then the formulas (20) and (21) may no longer make sense due to convergence issues (for instance what does $\prod_{n=1}^{\infty}(X-n)$ mean?). Therefore we need to find alternative methods to better understand Λ. This is where linear analysis comes in. Throughout the rest of this subsection, let $\mathcal{H}$ be a separable Hilbert space.

12.4.1 Eigenspaces of Compact Operators

Proposition 12.7. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a compact operator. Then E_λ is finite dimensional for all $\lambda \in \Lambda\backslash\{0\}$.*

Proof. Let λ be a nonzero eigenvalue of T and let (x_n) be a bounded sequence in E_λ. Choose a subsequence $(x_{\pi(n)})$ of (x_n) such that $x_{\pi(n)} \xrightarrow{w} x$ for some $x \in \mathcal{H}$ (such a subsequence exists by the Uniform Boundedness Principle). Since T is compact, we have

$$\begin{aligned} \lambda x_{\pi(n)} &= Tx_{\pi(n)} \\ &\to Tx \\ &= \lambda x, \end{aligned}$$

and since $\lambda \neq 0$, this implies $x_{\pi(n)} \to x$. Thus $(x_{\pi(n)})$ is a convergent subsequence of (x_n). It follows from Proposition (20.7) (in HW7 Problem 6.a) that E_λ is finite dimensional. □

Remark 28. Note that the eigenspace E_0 corresponding to the eigenvalue 0 is precisely the kernel of T, and this may be *infinite* dimensional! For instance, if $T\colon \mathcal{H} \to \mathcal{H}$ is the zero map, then E_0 is all of $\mathcal{H}$.

12.4.2 Eigenspaces of Self-Adjoint Operators

Proposition 12.8. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a self-adjoint operator. Then Λ is bounded above by $\|T\|$.*

Proof. Let $\lambda \in \Lambda$ and choose an eigenvector x corresponding to the eigenvalue λ. By scaling if necessary, we may assume $\|x\| = 1$. Then

$$\begin{aligned} \|T\| &= \sup\{|\langle Ty, y\rangle| \mid \|y\| \leq 1\} \\ &\geq |\langle Tx, x\rangle| \\ &= |\langle \lambda x, x\rangle| \\ &= |\lambda|. \end{aligned}$$

Therefore $|\lambda| \leq \|T\|$ for all $\lambda \in \Lambda$. In other words, Λ is bounded above by $\|T\|$. □

Proposition 12.9. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a self-adjoint operator. Suppose λ and μ are two distinct eigenvalues of T. Then $E_\lambda \perp E_\mu$.*

Proof. Let $x \in E_\lambda$ and let $y \in E_\mu$. Then we have

$$\begin{aligned} (\lambda - \mu)\langle x, y\rangle &= \langle \lambda x, y\rangle - \langle x, \overline{\mu} y\rangle \\ &= \langle \lambda x, y\rangle - \langle x, \mu y\rangle \\ &= \langle Tx, y\rangle - \langle x, Ty\rangle \\ &= \langle Tx, y\rangle - \langle Tx, y\rangle \\ &= 0. \end{aligned}$$

Therefore $E_\lambda \perp E_\mu$. □

Proposition 12.10. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a self-adjoint operator. Then $\Lambda \subseteq \mathbb{R}$.*

Proof. Let λ be an eigenvalue of T. Choose an eigenvector of λ, say $x \in \mathcal{H}$. By scaling if necessary, we may assume $\|x\| = 1$. Then

$$\begin{aligned}\lambda &= \lambda\langle x, x\rangle \\ &= \langle \lambda x, x\rangle \\ &= \langle Tx, x\rangle \\ &= \langle x, Tx\rangle \\ &= \langle x, \lambda x\rangle \\ &= \overline{\lambda}\langle x, x\rangle \\ &= \overline{\lambda},\end{aligned}$$

which implies λ is real. Therefore $\Lambda \subseteq \mathbb{R}$. □

12.4.3 Eigenspaces of Compact Self-Adjoint Operators

Proposition 12.11. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a compact self-adjoint operator. Assume that Λ is infinite. Then 0 is an accumulation point of Λ. Moreover, the set of all accumulation points of Λ is $\Lambda \cup \{0\}$. In other words, the closure of Λ (as a subset of $\mathbb{C}$) is $\overline{\Lambda} = \Lambda \cup \{0\}$.*

Remark 29. We give two remarks before we prove this proposition.

1. We are not saying that 0 is an eigenvalue of T. We are merely saying that 0 is an accumulation point of Λ and in fact the only accumulation point of Λ. Equivalently, we are saying that there exists a sequence of distinct eigenvalues (λ_n) such that $\lambda_n \to 0$ and that any convergent sequence of distinct eigenvalues must converge to 0.

2. The proposition tells us that Λ must be countable. This is because any uncountable subset of $\mathbb{C}$ must have infinitely many accumulation points.

Proof. Since Λ is bounded above by $\|T\|$, there exists a convergent sequence of distinct eigenvalues, say (λ_n). For each $n \in \mathbb{N}$, choose an eigenvector x_n of λ_n. By scaling if necessary, we may assume that $\|x_n\| = 1$ for all $n \in \mathbb{N}$. Since (λ_n) consists of distinct eigenvalues and since T is self-adjoint, we have $\langle x_m, x_n\rangle = 0$ whenever $m \neq n$. Thus (x_n) is an orthonormal sequence. In particular, this implies $x_n \xrightarrow{w} 0$. Since T is compact, we have $Tx_n \to 0$. Thus

$$\begin{aligned}\lambda_n x_n &= Tx_n \\ &\to 0,\end{aligned}$$

Taking norms gives us $|\lambda_n| \to 0$, which implies $\lambda_n \to 0$. □

12.4.4 Spectral Theorem

Since $\mathcal{H}$ is a separable Hilbert space, we know that there exists an orthonormal basis of $\mathcal{H}$. It turns out that we can get something even better:

Theorem 12.6. *(Spectral Theorem) Let $T\colon \mathcal{H} \to \mathcal{H}$ be a compact self-adjoint operator. Then there exists an orthonormal basis of $\mathcal{H}$ consisting of eigenvectors of T.*

Remark 30. Note that the sum (23) may be finite. Also note that the sequence (λ_n) of eigenvalues corresponding to the sequence (e_n) of eigenvectors may have repeated terms.

Proof. If $T = 0$, then the theorem is clear, so assume $T \neq 0$.

Step 1: A nonzero eigenvalue of T exists: Denote by S to be the set $\{\langle Tx, x\rangle \mid \|x\| \leq 1\}$. Then S is nonempty and bounded above by $\|T\|$. Therefore the supremum of S exists. Denote by λ_0 to be this supremum. We claim that $\lambda_0 \in S$ and hence is in fact a maximum of S. Indeed, choose a sequence $(\langle Tx_n, x_n\rangle)$ in S such that $\langle Tx_n, x_n\rangle \to \lambda_0$. Since the sequence (x_n) is bounded, it has a weakly convergent subsequence, say $(x_{\pi(n)})$ where $x_{\pi(n)} \xrightarrow{w} x_0$ for some $x_0 \in \mathcal{H}$. Since T is compact, we will have $Tx_{\pi(n)} \to Tx_0$, and this implies $\langle Tx_{\pi(n)}, x_{\pi(n)}\rangle \to \langle Tx_0, x_0\rangle$ by

Proposition (12.5). Since every subsequence of a convergence subsequence is convergent and moreover converges to the same limit, we have $\langle Tx_0, x_0\rangle = \lambda_0$. Finally, since

$$\begin{aligned} \|x_0\| &\leq \liminf_{n\to\infty} \|x_{\pi(n)}\| \\ &\leq 1, \end{aligned}$$

our claim is proved. In fact, we must have $\|x_0\| = 1$, since if $\|x_0\| < 1$, then

$$\begin{aligned} \left\langle T\left(\frac{x_0}{\|x_0\|}\right), \frac{x_0}{\|x_0\|}\right\rangle &= \frac{1}{\|x_0\|}\langle Tx_0, x_0\rangle \\ &> \lambda_0, \end{aligned}$$

which contradicts the fact that λ_0 is the supremum. Thus $\|x_0\| = 1$.

We next show that x_0 is an eigenvector of T with eigenvalue λ_0. Define a function $R_T\colon \mathcal{H}\backslash\{0\} \to \mathbb{R}$ (this is called the **Rayleigh quotient**) by

$$R_T(x) = \frac{\langle Tx, x\rangle}{\|x\|^2}$$

for all $x \in \mathcal{H}\backslash\{0\}$. Observe that $R_T(x) \leq R_T(x_0)$ for all $x \in \mathcal{H}\backslash\{0\}$. Define a function $f\colon \mathbb{R} \to \mathbb{R}$ by

$$f(t) = R_T(x_0 + tw)$$

where w is an arbitrary fixed vector in $\mathcal{H}$. Then f is maximized at $t = 0$. Moreover since

$$\begin{aligned} f(t) &= \frac{\langle T(x_0+tw), x_0+tw\rangle}{\|x_0+tw\|^2} \\ &= \frac{\langle Tx_0, x_0\rangle + t(\langle Tx_0, w\rangle + \langle Tw, x_0\rangle) + t^2\langle Tw, w\rangle}{\|x_0\|^2 + 2t\mathrm{Re}\langle x_0, w\rangle + t^2\|w\|^2}, \end{aligned}$$

we see that f is differentiabe at $t = 0$. It follows that

$$\begin{aligned} 0 &= f'(0) \\ &= \frac{\langle Tx_0, w\rangle + \langle Tw, x_0\rangle}{\|x_0\|^2} - \langle Tx_0, x_0\rangle \frac{\langle x_0, w\rangle + \langle w, x_0\rangle}{\|x_0\|^4} \\ &= \langle Tx_0, w\rangle + \langle Tw, x_0\rangle - \lambda_0(\langle x_0, w\rangle + \langle w, x_0\rangle) \\ &= \langle Tx_0, w\rangle + \langle w, Tx_0\rangle - \lambda_0(\langle x_0, w\rangle + \langle w, x_0\rangle) \\ &= 2\mathrm{Re}\langle Tx_0, w\rangle - \lambda_0(2\mathrm{Re}\langle x_0, w\rangle) \\ &= 2\mathrm{Re}\langle Tx_0 - \lambda_0 x_0, w\rangle \end{aligned}$$

for all $w \in \mathcal{H}$. Plugging in $w = Tx_0 - \lambda_0 x_0$, we get $2\mathrm{Re}\|Tx_0 - \lambda_0 x_0\|^2 = 0$ which implies $\|Tx_0 - \lambda_0 x_0\| = 0$ which implies $Tx_0 = \lambda_0 x_0$. Thus λ_0 is an eigenvalue corresponding to the eigenvector x_0.

Now observe that $-\lambda_0$ is an eigenvalue of $-T$. Denote by S' to be the set $\{\langle -Tx, x\rangle \mid \|x\| \leq 1\}$. Then S' is nonempty and bounded above by $\|T\|$. Therefore the supremum of S' exists. Denote by λ_0' to be this supremum. Running through the same argument above, we find that λ_0' is an eigenvalue of $-T$ and hence $-\lambda_0'$ is an eigenvalue of T. If $\lambda_0 \geq \lambda_0'$, then it follows that

$$\begin{aligned} \lambda_0 &= \sup\{|\langle Tx, x\rangle| \mid \|x\| \leq 1\} \\ &= \|T\|. \end{aligned}$$

If $\lambda_0' \geq \lambda_0$, then we would get $\lambda_0' = \|T\|$. Thus either $\|T\|$ or $-\|T\|$ is an eigenvalue of T (and in fact the largest eigenvalue of T in absolute value).

Step 2: An orthonormal basis of T consisting of eigenvectors exists. Denote by $\Lambda_{\neq 0}$ to be the set of all nonzero eigenvalues of T. Then $\Lambda_{\neq 0} \neq \emptyset$ by step 1. For each $\lambda \in \Lambda_{\neq 0}$, let $\{e_{\lambda,i} \mid 1 \leq i \leq n_\lambda\}$ be an orthonormal basis of E_λ where $n_\lambda := \dim E_\lambda$ is finite since T is compact. The set

$$\bigcup_{\lambda\in\Lambda_{\neq 0}} \{e_{\lambda,i} \mid 1 \leq i \leq n_\lambda\} \tag{22}$$

is an orthonormal set consisting of eigenvectors. Indeed, we have $e_{\lambda,i} \perp e_{\mu,j}$ for all $\lambda, \mu \in \Lambda_{\neq 0}$ such that $\lambda \neq \mu$ and for all $1 \leq i \leq n_\lambda$ and $1 \leq j \leq n_\lambda$ since T is self-adjoint. We also have $e_{\lambda,i} \perp e_{\lambda,j}$ for all $\lambda \in \Lambda_{\neq 0}$ and for all $1 \leq i, j \leq n_\lambda$ such that $i \neq j$. By placing an order on (22) and relabeling indices based on that order, we obtain an orthonormal sequence (e_n) consisting of eigenvectors corresponding to nonzero eigenvalues of T. Let

$$\mathcal{K} = \overline{\text{span}}(\{e_n \mid n \in \mathbb{N}\}).$$

Observe that $\mathcal{K}$ is T-invariant, meaning $Tx \in \mathcal{K}$ for all $x \in \mathcal{K}$. In fact, $\mathcal{K}^\perp$ is T-invariant too. Indeed, let $y \in \mathcal{K}^\perp$. Then

$$\begin{aligned} \langle x, Ty \rangle &= \langle Tx, y \rangle \\ &= 0 \end{aligned}$$

for all $x \in \mathcal{K}$. Therefore $Ty \in \mathcal{K}^\perp$. Thus the restriction of T to $\mathcal{K}^\perp$ lands in $\mathcal{K}^\perp$, denote this restriction by $T|_{\mathcal{K}^\perp} \colon \mathcal{K}^\perp \to \mathcal{K}^\perp$. The operator $T|_{\mathcal{K}^\perp}$ is compact and self-adjoint since it inherits these properties from T. We claim that $T|_{\mathcal{K}^\perp}$ is the zero operator. Indeed, assume for a contradiction that $T|_{\mathcal{K}^\perp}$ is not the zero operator. Then from what we just proved, the operator $T|_{\mathcal{K}^\perp}$ must have a nonzero eigenvalue say λ. Choose an eigenvector $x \in \mathcal{K}^\perp$ of $T|_{\mathcal{K}^\perp}$ corresponding to the eigenvalue λ. Then $x \in \mathcal{K}^\perp$ is also an eigenvector of T. Therefore $x \in \mathcal{K}^\perp \cap \mathcal{K} = \{0\}$, which is a contradiction since eigenvectors are nonzero. Thus $T|_{\mathcal{K}^\perp}$ must be the zero operator. In particular this implies $\mathcal{K}^\perp \subseteq \ker(T)$. In fact, we already have the reverse inclusion. Indeed, let $x \in \ker(T)$ and let $n \in \mathbb{N}$. Then

$$\begin{aligned} \lambda_n \langle e_n, x \rangle &= \langle \lambda_n e_n, x \rangle \\ &= \langle Te_n, x \rangle \\ &= \langle e_n, Tx \rangle \\ &= \langle e_n, 0 \rangle \\ &= 0, \end{aligned}$$

and since $\lambda_n \neq 0$, this implies $\langle e_n, x \rangle = 0$. Therefore $\langle e_n, x \rangle = 0$ for all $n \in \mathbb{N}$, and hence $x \in \mathcal{K}^\perp$. Finally, choose an orthonormal basis for $\ker T = \mathcal{K}^\perp$ and combine the orthonormal basis for $\mathcal{K}$ to get an orthonormal basis of $\mathcal{H} = \mathcal{K} \oplus \mathcal{K}^\perp$ consisting of eigenvectors of T. □

Corollary 8. *Let $T \colon \mathcal{H} \to \mathcal{H}$ be a compact self-adjoint operator and let (e_n) be an orthonormal basis of $\mathcal{H}$ consisting of eigenvectors. Then*

$$Tx = \sum_{n=1}^{\infty} \lambda_n \langle x, e_n \rangle e_n \tag{23}$$

for all $x \in \mathcal{H}$, where λ_n is the eigenvalue corresponding to the eigenvector e_n for all $n \in \mathbb{N}$.

Proof. Let $x \in \mathcal{H}$. Then

$$\begin{aligned} Tx &= T\left(\sum_{n=1}^{\infty} \langle x, e_n \rangle e_n\right) \\ &= \sum_{n=1}^{\infty} \langle x, e_n \rangle Te_n \\ &= \sum_{n=1}^{\infty} \langle x, e_n \rangle \lambda_n e_n \\ &= \sum_{n=1}^{\infty} \lambda_n \langle x, e_n \rangle e_n. \end{aligned}$$

□

12.5 Functional Calculus

Let $T \colon \mathcal{H} \to \mathcal{H}$ be a positive compact self-adjoint operator. Thus $\langle Tx, x \rangle \geq 0$ for all $x \in \mathcal{H}$. This implies

$$\begin{aligned} 0 &\leq \langle Te_n, e_n \rangle \\ &= \langle \lambda_n e_n, e_n \rangle \\ &= \lambda_n \|e_n\|^2, \end{aligned}$$

which implies $\lambda_n \geq 0$ for all $n \in \mathbb{N}$. Define $S\colon \mathcal{H} \to \mathcal{H}$ by

$$Sx = \sum_{n=1}^{\infty} \sqrt{\lambda_n}\langle x, e_n\rangle e_n$$

for all $x \in \mathcal{H}$. Then

$$\begin{aligned}
S^2x &= S(Sx) \\
&= \sum_{n=1}^{\infty} \sqrt{\lambda_n}\langle Sx, e_n\rangle e_n \\
&= \sum_{n=1}^{\infty} \sqrt{\lambda_n}\left\langle \sum_{m=1}^{\infty} \sqrt{\lambda_n}\langle x, e_m\rangle e_m, e_n \right\rangle e_n \\
&= \sum_{n=1}^{\infty} \sqrt{\lambda_n} \sum_{m=1}^{\infty} \sqrt{\lambda_n}\langle x, e_m\rangle \langle e_m, e_n\rangle e_n \\
&= \sum_{n=1}^{\infty} \sqrt{\lambda_n}\sqrt{\lambda_n}\langle x, e_n\rangle e_n \\
&= \sum_{n=1}^{\infty} \lambda_n\langle x, e_n\rangle e_n \\
&= Tx.
\end{aligned}$$

More generally, let $f\colon D \subseteq \mathbb{R} \to \mathbb{R}$ be a bounded function. If $\{0\} \cup \Lambda \subseteq D$ where Λ is the set of all eigenvalues of T, then we can define $f(T)\colon \mathcal{H}' \to \mathcal{H}$ by

$$f(T)x = \sum_{n=1}^{\infty} f(\lambda_n)\langle x, e_n\rangle e_n.$$

Then $f(T) + g(T) = (f+g)(T)$ and $f(T) \circ g(T) = (f \circ g)(T)$ and moreover we will have

$$\|f(T)\| = \sup_{x \in D} |f(x)| \cdot \|T\|.$$

12.6 Singular-Value Decomposition

Recall that the polarization decomposition of a complex number

$$z = |z|e^{i\theta} \text{ where } |z|^2 = \overline{z}z.$$

We want to find an analogue of this for compact operators. Let $T\colon \mathcal{H} \to \mathcal{H}$ be a compact operator. Then T^*T is compact, self-adjoint, and positive. Then there exists a unique positive compact self-adjoint operator S such that $S^2 = T^*T$. [2] We denote this operator S by $|T|$. Thus

$$|T|x = \sum_{n=1}^{\infty} \sqrt{\lambda_n}\langle x, e_n\rangle e_n$$

for all $x \in \mathcal{H}$ where (λ_n) is a sequence consisting of all eigenvalues of T^*T and (e_n) a corresponding orthonormal basis of eigenvectors. Let $U\colon \mathcal{H} \to \mathcal{H}$ be defined by

$$Ux = \sum_{n=1}^{\infty} \frac{1}{\sqrt{\lambda_n}}\langle x, e_n\rangle Te_n,$$

[2] In fact, if T is not compact, then it turns out that we still obtain a unique positive compact self-adjoint operator S such that $S^2 = T^*T$, but this requires some measure theory which is outside the scope of this class.

where if $\lambda_n = 0$, then it is understood that $1/\sqrt{\lambda_n} = 0$. Then observe that

$$\begin{aligned}
U(|T|x) &= \sum_{n=1}^{\infty} \frac{1}{\sqrt{\lambda_n}} \langle |T|x, e_n \rangle Te_n \\
&= \sum_{n=1}^{\infty} \frac{1}{\sqrt{\lambda_n}} \left\langle \sum_{m=1}^{\infty} \sqrt{\lambda_m} \langle x, e_m \rangle e_m, e_n \right\rangle Te_n \\
&= \sum_{n=1}^{\infty} \frac{1}{\sqrt{\lambda_n}} \sum_{m=1}^{\infty} \sqrt{\lambda_m} \langle x, e_m \rangle \langle e_m, e_n \rangle Te_n \\
&= \sum_{n=1}^{\infty} \frac{1}{\sqrt{\lambda_n}} \sqrt{\lambda_n} \langle x, e_n \rangle Te_n \\
&= \sum_{n=1}^{\infty} \langle x, e_n \rangle Te_n \\
&= T\left(\sum_{n=1}^{\infty} \langle x, e_n \rangle e_n \right) \\
&= Tx
\end{aligned}$$

for all $x \in \mathcal{H}$. Therefore $U|T| = T$. We call this the **polar decomposition of** T. The numbers $\sqrt{\lambda_n}$ are called the **singular values of** T. Observe that

$$\begin{aligned}
Tx &= U|T|x \\
&= U\left(\sum_{n=1}^{\infty} \sqrt{\lambda_n} \langle x, e_n \rangle e_n \right) \\
&= \sum_{n=1}^{\infty} \sqrt{\lambda_n} \langle x, e_n \rangle Ue_n
\end{aligned}$$

for all $x \in \mathcal{H}$.

13 Normed Linear Spaces

Definition 13.1. Let $\mathcal{X}$ be a vector space (over $\mathbb{C}$). A function $\|\cdot\| \colon \mathcal{X} \to \mathbb{R}$ satisfying

1. (Positive-Definite) $\|x\| \geq 0$ for all $x \in \mathcal{X}$ with equality if and only if $x = 0$.
2. (Absolutely Homogeneous) $\|\alpha x\| = |\alpha| \|x\|$ for all $x \in \mathcal{X}$ and $\alpha \in \mathbb{C}$.
3. (Subadditivity) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathcal{X}$

is called a **norm** on $\mathcal{X}$. We call the pair $(\mathcal{X}, \|\cdot\|)$ a **normed linear space**.

Example 13.1. Every inner-product space is a normed linear space.

Example 13.2. Let $1 \leq p < \infty$. Define $\|\cdot\|_p \colon \mathbb{C}^n \to \mathbb{R}$ by

$$\|(x_1, \dots, x_n)^\top\|_p = \sqrt[p]{|x_1|^p + \cdots + |x_n|^p}$$

for all $(x_1, \dots, x_n)^\top \in \mathbb{C}^n$. Then $\|\cdot\|_p$ is a norm on $\mathbb{C}^n$. More generally, we define

$$\ell^p(\mathbb{N}) := \left\{ (x_n) \in \mathbb{C}^\infty \mid \sum_{n=1}^{\infty} |x_n|^2 < \infty \right\}$$

and we define $\|\cdot\|_p \colon \ell^p(\mathbb{N}) \to \mathbb{R}$ by

$$\|(x_n)\|_p = \sqrt{\sum_{n=1}^{\infty} |x_n|^p}$$

Example 13.3. Define $\|\cdot\|_\infty\colon \mathbb{C}^n \to \mathbb{R}$ by

$$\|(x_1,\dots,x_n)^\top\|_\infty = \max\{|x_1|,\dots,|x_n|\}.$$

for all $(x_1,\dots,x_n)^\top \in \mathbb{C}^n$. Then $\|\cdot\|_\infty$ is a norm on $\mathbb{C}^n$. In fact we have $\|(x_1,\dots,x_n)^\top\|_p \to \|(x_1,\dots,x_n)^\top\|_\infty$ as $p\to\infty$ for all $(x_1,\dots,x_n)^\top \in \mathbb{C}^n$. More generally, we define

$$\ell^\infty(\mathbb{N}) := \{(x_n) \in \mathbb{C}^\infty \mid \sup\{|x_n| \mid n \in \mathbb{N}\} < \infty\}$$

and we define $\|\cdot\|_\infty\colon \ell^\infty(\mathbb{N}) \to \mathbb{R}$ by

$$\|(x_n)\|_\infty = \sup\{|x_n| \mid n \in \mathbb{N}\}.$$

Example 13.4. Define $\|\cdot\|_{\sup}\colon C[a,b] \to \mathbb{R}$ by

$$\|f\|_{\sup} = \sup\{|f(x)| \mid x \in [a,b]\}.$$

Example 13.5. Let $\mathcal{H}$ be a Hilbert space and

$$\mathcal{L}(\mathcal{H}) = \{T\colon \mathcal{H}\to\mathcal{H} \mid T \text{ is bounded}\}.$$

We want to prove that $\|\cdot\|_p$ is a norm. We first prove it in the finite dimensional case. For the first property we have

$$\|x\|_p = \sqrt[p]{|x_1|^p+\cdots+|x_n|^p} \geq 0$$

with equality if and only if $x_1=\cdots=x_p=0$. For the second property, we have

$$\begin{aligned}\|\alpha x\|_p &= \sqrt[p]{|\alpha x_1|^p+\cdots+|\alpha x_n|^p}\\ &= |\alpha|\|x\|_p.\end{aligned}$$

For the triangle inequality, we need to show that

$$\sqrt[p]{|x_1+y_1|^p+\cdots+|x_n+y_n|^p} \leq \sqrt[p]{|x_1|^p+\cdots+|x_n|^p} + \sqrt[p]{|y_1|^p+\cdots+|y_n|^p}. \tag{24}$$

To prove this we will use the following analog of the Cauchy-Schwarz inequality called the **Hölder inequality**: for positive numbers $a_1,\dots,a_n$ and $b_1,\dots,b_n$, we have

$$\sum_{i=1}^n a_ib_i \leq \sqrt[p]{\sum_{i=1}^n a_i^p}\sqrt[q]{\sum_{i=1}^n b_i^q} \tag{25}$$

where $1\leq p,q<\infty$ such that $1=\frac{1}{p}+\frac{1}{q}$. To prove Hölder inequality we will use

$$ab \leq \frac{a^p}{p}+\frac{b^q}{q} \tag{26}$$

for all $a,b\geq 0$. If $\sum_{i=1}^n a_i^p = 0$ or if $\sum_{i=1}^n b_i^p = 0$ then the inequality (25) is trivial. So assume both are nonzero. In this case, the inequality (25) is equivalent to

$$\sum_{i=1}^n \frac{a_i}{A}\frac{b_i}{B} \leq 1. \tag{27}$$

where $A=\sqrt[p]{\sum_{i=1}^n a_i^p}$ and $B=\sqrt[q]{\sum_{i=1}^n b_i^q}$. Applying (26), we obtain

$$\begin{aligned}\sum_{i=1}^n \frac{a_i}{A}\frac{b_i}{B} &\leq \sum_{i=1}^n\left(\frac{\left(\frac{a_i}{A}\right)^p}{p}+\frac{\left(\frac{b_i}{B}\right)^q}{q}\right)\\ &= \frac{1}{A^pp}\sum_{i=1}^n a_i^p + \frac{1}{B^qq}\sum_{i=1}^n b_i^q\\ &= \frac{A^p}{A^pp}+\frac{B^q}{B^qq}\\ &= \frac{1}{p}+\frac{1}{q}\\ &= 1.\end{aligned}$$

This establishes (25). Next we prove (24). Let $x, y \in (\mathbb{C}^n, \|\cdot\|_p)$. Then

$$
\begin{aligned}
\|x+y\|_p^p &= \left(\sum_{i=1}^n |x_i+y_i|^p\right) \\
&= \sum_{i=1}^n |x_i+y_i||x_i+y_i|^{p-1} \\
&\le \sum_{i=1}^n (|x_i|+|y_i|)|x_i+y_i|^{p-1} \\
&= \sum_{i=1}^n |x_i||x_i+y_i|^{p-1} + \sum_{i=1}^n |y_i||x_i+y_i|^{p-1} \\
&\le \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^n |x_i+y_i|^{(p-1)q}\right)^{\frac{1}{q}} + \left(\sum_{i=1}^n |y_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^n |x_i+y_i|^{(p-1)q}\right)^{\frac{1}{q}} \\
&= \|x\|_p \left(\sum_{i=1}^n |x_i+y_i|^p\right)^{\frac{1}{q}} + \|y\|_p \left(\sum_{i=1}^n |x_i+y_i|^p\right)^{\frac{1}{q}} \\
&= \left(\|x\|_p + \|y\|_p\right) \left(\sum_{i=1}^n |x_i+y_i|^p\right)^{\frac{1}{q}} \\
&= \left(\|x\|_p + \|y\|_p\right) \|x+y\|_p^{\frac{p}{q}} \\
&= \left(\|x\|_p + \|y\|_p\right) \|x+y\|_p^{p-1},
\end{aligned}
$$

which implies $\|x+y\|_p \le \|x\|_p + \|y\|_p$.

Now consider $\ell^p(\mathbb{N})$. We want to show that the ℓ^p norm

$$\|x\|_p := \left(\sum_{i=1}^\infty |x_i|^p\right)^{\frac{1}{p}}$$

is a norm on $\ell^p(\mathbb{N})$. Properties (1) and (2) are easy to prove. We know for each $N \in \mathbb{N}$, we have

$$
\begin{aligned}
\left(\sum_{i=1}^N |x_i+y_i|^p\right)^{\frac{1}{p}} &\le \left(\sum_{i=1}^N |x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^N |y_i|^p\right)^{\frac{1}{p}} \\
&\le \left(\sum_{i=1}^\infty |x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^\infty |y_i|^p\right)^{\frac{1}{p}} \\
&= \|x\|_p + \|y\|_p.
\end{aligned}
$$

Since the left-hand side is monotone increasing and bounded in N, taking the limit $N \to \infty$ gives us $\|x+y\|_p \le \|x\|_p + \|y\|_p$.

13.1 Topology Induced by Norm

Just as in the case of inner-product spaces, we define convergence in normed linear spaces by $x_n \to x$ if $\|x_n - x\| \to 0$. We also define closure, open/closed balls, open/closed sets in exactly the same way. Also the notion of Cauchy sequence is defined in exactly the same way.

Proposition 13.1. *Let $\mathcal{X}$ be a normed linear space. Then*

1. *Every convergent sequence is Cauchy;*
2. *Every Cauchy (and hence every convergent sequence) is bounded;*
3. *If $x_n \to x$ and $y_n \to y$ then $\alpha x_n + \beta y_n \to \alpha x + \beta y$;*
4. *If $\alpha_n \to \alpha$ and $x_n \to x$, then $\alpha_n x_n \to \alpha x$.*

Proof. The proof is the same as in the inner-product case. □

Note that the converse of (1) in the previous proposition is wrong in general.

Definition 13.2. Let $\mathcal{X}$ be a normed linear space. We say $\mathcal{X}$ is a **Banach space** if every Cauchy sequence in $\mathcal{X}$ is convergent.

13.2 Bounded Operators on Normed Linear Spaces

We define the notion of a bounded operator on normed linear spaces in the same way as before.

Definition 13.3. Let $\mathcal{X}$ and $\mathcal{Y}$ be normed linear spaces and let $T\colon \mathcal{X} \to \mathcal{Y}$ be a map. We say T is a **bounded linear operator** if T is linear and if

$$\sup\{\|Tx\|_{\mathcal{Y}} \mid \|x\|_{\mathcal{X}} \leq 1\} < \infty.$$

In this case, we define the **operator norm** of T to be

$$\|T\| := \sup\{\|Tx\|_{\mathcal{Y}} \mid \|x\|_{\mathcal{X}} \leq 1\}.$$

A bounded linear operator $\ell\colon \mathcal{X} \to \mathbb{C}$ is called a **bounded linear functional**.

Proposition 13.2. *Let $\mathcal{X}$ and $\mathcal{Y}$ be normed linear spaces and let $T\colon \mathcal{X} \to \mathcal{Y}$ be a linear map. Then T is bounded if and only if T is continuous at 0 if and only if T is uniformly continuous.*

Proposition 13.3. *If T is bounded then $\|Tx\| \leq \|T\|\|x\|$ for all $x \in \mathcal{X}$.*

13.3 Dual Spaces

Definition 13.4. Given a normed linear space X, the **dual space** of X, denoted by X^*, is the vector space which consists of all bounded linear functionals on X equipped with the operator norm. In other words,

$$X^* = \{\ell\colon X \to \mathbb{C} \mid \ell \text{ is bounded and linear.}\}$$

13.3.1 Hahn-Banach Theorem and its Consequences

Theorem 13.1. *(Hahn-Banach Theorem) Let X be a normed linear space and let Y be a subspace of X. Then every bounded linear functional $\psi\colon Y \to \mathbb{C}$ can be extended to X with the same norm. More precisely, there exists a bounded linear functional $\widetilde{\psi}\colon X \to \mathbb{C}$ such that $\widetilde{\psi}|_Y = \psi$ and moreover $\|\widetilde{\psi}\| = \|\psi\|$.*

Remark 31. Note that $\|\widetilde{\psi}\| \leq \|\psi\|$ is the nontrivial direction.

Proposition 13.4. *For any nonzero vector $x_0 \in X$ there exists $\ell \in X^*$ with $\|\ell\| = 1$ such that $\ell(x_0) = \|x_0\|$.*

Proof. Let $Y = \operatorname{span}(\{x_0\})$. Define $\psi\colon Y \to \mathbb{C}$ by

$$\psi(\lambda x_0) = \lambda\|x_0\|$$

for all $\lambda x_0 \in \operatorname{span}(\{x_0\})$. It is easy to check that ψ is linear and bounded with $\psi(x_0) = \|x_0\|$ and $\|\psi\| = 1$. By the Hahn-Banach Theorem, we can extend ψ to a bounded linear functional $\ell\colon X \to \mathbb{C}$ such that $\|\ell\| = \|\psi\| = 1$ and such that $\ell(x_0) = \psi(x_0) = \|x_0\|$. This completes the proof. □

Proposition 13.5. *For any $x \in X$, we have*

$$\|x\| = \max\{|\ell(x)| \mid \ell \in X^* \textit{ and } \|\ell\| \leq 1\}.$$

Proof. First note that for any $\ell \in X^*$ such that $\|\ell\| \leq 1$ we always have

$$\begin{aligned} |\ell(x)| &\leq \|\ell\|\|x\| \\ &\leq \|x\|. \end{aligned}$$

For the reverse inequality, choose an $\ell \in X^*$ such that $\|\ell\| = 1$ and such that $\ell(x) = \|x\|$ (such a choice exists by Proposition (13.4)). This gives us the reverse direction. □

Proposition 13.6. *X^* is always a Banach space.*

Proof. It suffices to show X^* is complete. Let (ℓ_n) be a Cauchy sequence in X^*. Let $x \in X$ and let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $m, n \geq N$ implies

$$\|\ell_n - \ell_m\| < \frac{\varepsilon}{\|x\|}.$$

Then $n, m \geq N$ implies

$$\begin{aligned} |\ell_n(x) - \ell_m(x)| &= |(\ell_n - \ell_m)(x)| \\ &\leq \|\ell_n - \ell_m\|\|x\| \\ &< \frac{\varepsilon}{\|x\|}\|x\| \\ &= \varepsilon. \end{aligned}$$

In particular, $(\ell_n(x))$ is a Cauchy sequence of complex numbers for all $x \in X$. Since $\mathbb{C}$ is complete, we may define $\ell \colon X \to \mathbb{C}$ by

$$\ell(x) = \lim_{n\to\infty} \ell_n(x)$$

for all $x \in X$.

We claim that $\ell \in X^*$ and $\ell_n \to \ell$ as $n \to \infty$. Linearity of ℓ follows from linearity of ℓ_n and linearity of the limit operator. To see that ℓ is bounded, we consider the inequality

$$|\ell_n(x) - \ell_m(x)| < \varepsilon\|x\|.$$

Then for $\|x\| \leq 1$ and setting $m \to \infty$, we find that $n \geq N$ implies

$$|\ell_n(x) - \ell(x)| \leq \varepsilon.$$

Therefore $\ell_n - \ell$ is bounded. In particular, $\ell = (\ell_n - \ell) + \ell_n$ is bounded since X^* is a vector space. This also implies $\|\ell_n\| \to \|\ell\|$. □

Definition 13.5. Two normed linear spaces $(X_1, \|\cdot\|_1)$ and $(X_2, \|\cdot\|_2)$ are said to be **isometricaly isomorphic** if there exists a bijective linear map $T \colon X_1 \to X_2$ such that $\|Tx\|_2 = \|x\|_1$ for all $x \in X_1$. In this case, we write $X \cong Y$. If T is not necessarily surjective, then we say $(X_1, \|\cdot\|_1)$ is **isometrically embedded** in $(X_2, \|\cdot\|_2)$ and we call such T an (**isometrical**) **embedding**.

Proposition 13.7. *Suppose $1 \leq p < \infty$ and q is the conjugate of p (i.e. $1/p + 1/q = 1$). The dual space of $\ell^p(\mathbb{N})$ is isometrically isomorphic to $\ell^q(\mathbb{N})$.*

Definition 13.6. Let $\mathcal{X}$ be a normed linear space and let A be a subset of $\mathcal{X}$. The **annihilator of** A **in** $\mathcal{X}$ is defined to be the set of all bounded linear functionals $\ell \colon \mathcal{X} \to \mathbb{C}$ that annihilate A, i.e. $\ell(a) = 0$ for all $a \in A$. We denote the annihilator of A by $A^\perp$.

Proposition 13.8. *For any $A \subseteq \mathcal{X}$ we have $A^\perp$ is a closed subspace of $\mathcal{X}^*$.*

Just like in Hilbert spaces, we define

$$d(x, A) = \inf\{\|x - a\| \mid a \in A\}$$

to be the distance from x to A.

Theorem 13.2. *Let $\mathcal{X}$ be a normed linear space and let $\mathcal{Y}$ be a subspace of $\mathcal{X}$. Then*

$$d(x, \mathcal{Y}) = \max\{|\ell(x)| \mid \ell \in \mathcal{Y}^\perp \text{ and } \|\ell\| \leq 1\}.$$

Proof. Let $x \in \mathcal{X}$ and let $\ell \in \mathcal{Y}^\perp$ with $\|\ell\| \leq 1$. Then

$$\begin{aligned} |\ell(x)| &= |\ell(x - y)| \\ &\leq \|\ell\|\|x - y\| \\ &\leq \|x - y\| \end{aligned}$$

for all $y \in \mathcal{Y}$. Therefore

$$\begin{aligned}
d(x, \mathcal{Y}) &\geq \inf_{y \in \mathcal{Y}} \|x - y\| \\
&\geq \inf_{y \in \mathcal{Y}} |\ell(x)| \\
&= |\ell(x)|.
\end{aligned}$$

This implies

$$\sup\{|\ell(x)| \mid \ell \in \mathcal{Y}^{\perp} \text{ and } \|\ell\| \leq 1\} \leq d(x, \mathcal{Y}).$$

Now we want to show that the supremum is actually a maximum. If $x \in \mathcal{Y}$ then it is obvious, so assume $x \notin \mathcal{Y}$. Define $\mathcal{Z} = \operatorname{span}\{x, \mathcal{Y}\}$ and define $\psi\colon \mathcal{Z} \to \mathbb{C}$ by

$$\psi(\lambda x + y) = \lambda d(x, \mathcal{Y})$$

for all $\lambda x + y \in \mathcal{Z}$. Note that any element in $\mathcal{Z}$ can be uniquely expressed as $\lambda x + y$ for some $\lambda \in \mathbb{C}$ and $y \in \mathcal{Y}$ and hence ψ is well-defined. It is easy to show that ψ is a linear functional. Also $\psi(y) = 0$ for all $y \in \mathcal{Y}$ and $\psi(x) = d(x, \mathcal{Y})$. Let $z \in \mathcal{Z}$. If $z \in \mathcal{Y}$, then $|\psi(z)| \leq \|z\|$, otherwise let $z = \lambda x + y$ with $\lambda \neq 0$. Then

$$\begin{aligned}
|\psi(z)| &= |\lambda| d(x, \mathcal{Y}) \\
&= \frac{|\lambda| \|z\|}{\|z\|} d(x, \mathcal{Y}) \\
&= \frac{|\lambda| \|z\|}{\|\lambda x + y\|} d(x, \mathcal{Y}) \\
&= \frac{|\lambda| \|z\|}{\|\lambda(x + \lambda^{-1} y)\|} d(x, \mathcal{Y}) \\
&= \frac{\|z\|}{\|(x + \lambda^{-1} y)\|} d(x, \mathcal{Y}) \\
&\leq \|z\|
\end{aligned}$$

So for any $z \in \mathcal{Z}$, we have $|\psi(z)| \leq \|z\|$. By the Hahn-Banach Theorem there exists $\ell \in \mathcal{X}^*$ such that $\ell(z) = \psi(z)$ for all $z \in \mathcal{Z}$ and $\|\ell\| = \|\psi\| \leq 1$. Then $\ell(y) = \psi(y) = 0$ for all $y \in \mathcal{Y}$. In otherwords, $\ell \in \mathcal{Y}^{\perp}$. Also $\ell(x) = \psi(x) = d(x, \mathcal{Y})$. Therefore the supremum is a maximum. $\square$

13.4 Reflexivity

Any $x \in \mathcal{X}$ defines a natural linear functional $x^{**}\colon \mathcal{X}^* \to \mathbb{C}$ by

$$x^{**}(\ell) = \ell(x)$$

for all $\ell \in \mathcal{X}^*$. Indeed, for linearity let $\alpha_1, \alpha_2 \in \mathbb{C}$ and $\ell_1, \ell_2 \in \mathcal{X}^*$. Then

$$\begin{aligned}
x^{**}(\alpha_1 \ell_1 + \alpha_2 \ell_2) &= (\alpha_1 \ell_1 + \alpha_2 \ell_2)(x) \\
&= \alpha_1 \ell_1(x) + \alpha_2 \ell_2(x) \\
&= \alpha_1 x^{**}(\ell) + \alpha_2 x^{**}(\ell).
\end{aligned}$$

Moreover x^{**} is bounded above by $\|x\|$ since

$$\begin{aligned}
|x^{**}(\ell)| &= |\ell(x)| \\
&\leq \|\ell\| \|x\|
\end{aligned}$$

for all $\ell \in \mathcal{X}^*$. In fact $\|x^{**}\| = \|x\|$.

Proposition 13.9. *Define* $\Phi\colon \mathcal{X} \to \mathcal{X}^{**}$ *by*

$$\Phi(x) = x^{**}$$

for all $x \in \mathcal{X}$. *Then* Φ *is an isometrically embedding.*

Proof. For linearity, let $\alpha, \beta \in \mathbb{C}$ and $x, y \in \mathcal{X}$ and let $\ell \in \mathcal{X}^*$. Then

$$\begin{aligned}(\alpha x + \beta y)^{**}(\ell) &= \ell(\alpha x + \beta y) \\ &= \alpha \ell(x) + \beta \ell(y) \\ &= \alpha x^{**}(\ell) + \beta y^{**}(\ell) \\ &= (\alpha x^{**} + \beta y^{**})(\ell).\end{aligned}$$

Therefore $(\alpha x + \beta y)^{**} = \alpha x^{**} + \beta y^{**}$, and so Φ is linear. Also we proved

$$\begin{aligned}\|\Phi(x)\| &= \|x^{**}\| \\ &= \|x\|.\end{aligned}$$

In other words Φ is an isometry. $\square$

Definition 13.7. A normed linear space $\mathcal{X}$ is said to be **reflexive** if the natural embedding Φ is surjective, i.e. $\mathcal{X} \cong \mathcal{X}^{**}$.

13.4.1 Examples of Reflexive Banach Spaces

Example 13.6. Let $1 < p < \infty$. Then $\ell^p(\mathbb{N})$ is reflexive. On the other hand, $\ell^1(\mathbb{N})$, $\ell^\infty(\mathbb{N})$, and $(C[a,b], \|\cdot\|_{\sup})$ are not reflexive.

Lemma 13.3. *Let $\mathcal{Y}$ be a closed subspace of a normed linear space $\mathcal{X}$. Then there exists $z \in \mathcal{X}$ with $\|z\| = 1$ such that $d(z, \mathcal{Y}) \geq 1/2$.*

Proof. Let $x \notin \mathcal{Y}$. Since $\mathcal{Y}$ is closed we must have $d(x, \mathcal{Y}) > 0$. Choose $y_0 \in \mathcal{Y}$ such that

$$\|x - y_0\| < d(x, \mathcal{Y}) + d(x, \mathcal{Y}).$$

Let $z' = x - y_0$. Then

$$\begin{aligned}\|z' - y\| &= \|x - y_0 - y\| \\ &= \|x - (y + y_0)\| \\ &\geq d(x, \mathcal{Y})\end{aligned}$$

for all $y \in \mathcal{Y}$. Now set $z = z'/\|z'\|$. Then $\|z\| = 1$ and

$$\begin{aligned}\|z - y\| &= \left\| \frac{z'}{\|z'\|} - y \right\| \\ &\frac{\|z' - \|z'\| \cdot y\|}{\|z'\|} \\ &> \frac{d(x, \mathcal{Y})}{\|x - y_0\|} \\ &\geq \frac{d(x, \mathcal{Y})}{2d(x, \mathcal{Y})} \\ &= \frac{1}{2}.\end{aligned}$$

$\square$

Theorem 13.4. *The closed (unit) ball in any infinite-dimensional normed linear space is not compact.*

Proof. Let y_1 with $\|y_1\| = 1$ be arbitrary. Set $\mathcal{Y}_1 = \operatorname{span}\{y_1\}$. Then $\mathcal{Y}_1$ is a closed proper subspace of $\mathcal{X}$. Therefore by the lemma, there exists y_2 with $\|y_2\| = 1$ such that $d(y_2, \mathcal{Y}_1) \geq 1/2$. Choose such y_2. Then clearly

$$\|y_2 - y_1\| \geq \frac{1}{2}.$$

Next set $\mathcal{Y}_2 = \operatorname{span}\{y_1, y_2\}$. Again, $\mathcal{Y}_2$ is a proper closed subspace of $\mathcal{X}$. Choose y_3 with $\|y_3\| = 1$ such that $d(y_3, \mathcal{Y}_2) \geq 1/2$. In particular, we have

$$\|y_3 - y_2\| \geq \frac{1}{2} \quad \text{and} \quad \|y_3 - y_1\| \geq \frac{1}{2}.$$

Continuing in this manner, we construct a sequence (y_n) such that

$$\|y_n - y_m\| \geq \frac{1}{2}$$

for all $n \geq m \geq 1$. Clearly this sequence doesn't have a convergent subsequence. $\square$

Recall that in a Hilbert space $\mathcal{H}$, a sequence (x_n) converges weakly to x if

$$\langle x_n, y \rangle \to \langle x, y \rangle$$

for all $y \in \mathcal{H}$. We do not have inner-products in a normed linear space, however there is still an analogue of weak convergence in normed linear spaces:

Definition 13.8. A sequence (x_n) in a normed linear space $\mathcal{X}$ is said to be **weakly convergent** if there exists $x \in \mathcal{X}$ such that

$$\ell(x_n) \to \ell(x)$$

for all $\ell \in \mathcal{X}^*$. If such x exists it is unique in this case we write $x_n \xrightarrow{w} x$.

Theorem 13.5. *Every weakly convergent sequence in a normed linear space is bounded. Moreover, if $x_n \xrightarrow{w} x$, then*

$$\|x\| \leq \liminf_{n\to\infty} \|x_n\|.$$

Proof. The proof relies on the uniform boundedness principle just as in the Hilbert space case. $\square$

Theorem 13.6. *Every bounded sequence in a (separable) reflexive Banach space has a weakly convergent subsequence.*

Remark 32. The separable assumption can be dropped but the reflexive assumption cannot be dropped.

Proof. Diagonal argument. We need to use reflexivity in the last step in place of the Riesz Representation Theorem. $\square$

Proposition 13.10. *Let $\mathcal{Y}$ be a closed subspace of a normed linear space $\mathcal{X}$. If (x_n) is a sequence in $\mathcal{Y}$ and $x_n \xrightarrow{w} x$, then $x \in \mathcal{Y}$.*

Proof. Similar to the Hilbert space case. We need to use

$$d(x, \mathcal{Y}) = \sup\{|\ell(x)| \mid \ell \in \mathcal{Y}^\perp \text{ and } \|\ell\| \leq 1\}.$$

$\square$

Theorem 13.7. *Let $\mathcal{X}$ be a reflexive Banach space and let $\mathcal{Y}$ be a closed subspace of $\mathcal{X}$. Then for any $x \in \mathcal{X}$ there exists $y_0 \in \mathcal{Y}$ such that*

$$\|x - y_0\| = d(x, \mathcal{Y}).$$

Remark 33. Note that y_0 is not necessarily unique.

Proof. For each $n \in \mathbb{N}$, there exists $y_n \in \mathcal{Y}$ such that

$$d(x, \mathcal{Y}) \leq \|x - y_n\| < d(x, \mathcal{Y}) + \frac{1}{n}.$$

Then

$$\begin{aligned}
\|y_n\| &\leq \|y_n - x\| + \|x\| \\
&< d(x, \mathcal{Y}) + \frac{1}{n} + \|x\| \\
&\leq d(x, \mathcal{Y}) + 1 + \|x\|
\end{aligned}$$

for all $n \in \mathbb{N}$. Therefore (y_n) is a bounded sequence and hence contains a weakly convergent subsequence, say $(y_{\pi(n)})$ with $y_{\pi(n)} \xrightarrow{w} y_0$. Since $(y_{\pi(n)}) \subseteq \mathcal{Y}$, the weak limit y_0 is also in $\mathcal{Y}$. Then $x - y_{\pi(n)} \xrightarrow{w} x - y_0$, and hence

$$\|x - y_0\| \leq \liminf_{n\to\infty} \|x - y_{\pi(n)}\|.$$

Therefore

$$\begin{aligned}
d(x, \mathcal{Y}) &\leq \|x - y_0\| \\
&\leq \liminf_{n\to\infty} \|x - y_{\pi(n)}\| \\
&\leq \liminf_{n\to\infty} \left(d(x, \mathcal{Y}) + \frac{1}{\pi(n)} \right) \\
&= d(x, \mathcal{Y}),
\end{aligned}$$

which implies $\|x - y_0\| = d(x, \mathcal{Y})$. $\square$

Homework Problems and Solutions

14 Homework 1

14.1 Polarization Identity

Proposition 14.1. *(Polarization Identity) For $x, y \in \mathscr{V}$ we have*

$$4\langle x, y\rangle = \|x+y\|^2 + i\|x+iy\|^2 - \|x-y\|^2 - i\|x-iy\|^2$$

Proof. We calculate

$$\begin{aligned}
\|x+y\|^2 &= \langle x+y, x+y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle,
\end{aligned}$$

and

$$\begin{aligned}
i\|x+iy\|^2 &= i\langle x+iy, x+iy\rangle \\
&= i\langle x, x\rangle + i\langle x, iy\rangle + i\langle iy, x\rangle + i\langle iy, iy\rangle \\
&= i\langle x, x\rangle + \langle x, y\rangle - \langle y, x\rangle + i\langle y, y\rangle,
\end{aligned}$$

and

$$\begin{aligned}
-\|x-y\|^2 &= -\langle x-y, x-y\rangle \\
&= -\langle x, x\rangle - \langle x, -y\rangle - \langle -y, x\rangle - \langle -y, -y\rangle \\
&= -\langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle - \langle y, y\rangle,
\end{aligned}$$

and

$$\begin{aligned}
-i\|x-iy\|^2 &= -i\langle x-iy, x-iy\rangle \\
&= -i\langle x, x\rangle - i\langle x, -iy\rangle - i\langle -iy, x\rangle - i\langle -iy, -iy\rangle \\
&= -i\langle x, x\rangle + \langle x, y\rangle - \langle y, x\rangle - i\langle y, y\rangle.
\end{aligned}$$

Adding these together gives us our desired result. $\square$

14.2 Parallelogram Identity

Proposition 14.2. *(Parallelogram Identity) For $x, y \in \mathscr{V}$ we have*

$$\|x-y\|^2 + \|x+y\|^2 = 2\|x\|^2 + 2\|y\|^2$$

Proof. We calculate

$$\begin{aligned}
\|x+y\|^2 &= \langle x+y, x+y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&= \|x\|^2 + 2\mathrm{Re}(\langle x, y\rangle) + \|y\|^2.
\end{aligned}$$

and

$$\begin{aligned}\|x-y\|^2 &= \langle x-y, x-y\rangle \\ &= \langle x,x\rangle + \langle x,-y\rangle + \langle -y,x\rangle + \langle -y,-y\rangle \\ &= \|x\|^2 - 2\mathrm{Re}(\langle x,y\rangle) + \|y\|^2.\end{aligned}$$

Adding these together gives us our desired result. □

The geometric interpretation of Proposition (14.2) in the case where $\mathscr{V} = \mathbb{R}^3$ can be seen below:

14.3 Pythagorean Theorem

Proposition 14.3. *(Pythagorean Theorem) Let x and y be nonzero vectors in $\mathscr{V}$ such that $\langle x, y\rangle = 0$ (we call such vectors* ***orthogonal*** *to one another). Then*

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2.$$

Proof. We have

$$\begin{aligned}\|x+y\|^2 &= \langle x+y, x+y\rangle \\ &= \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle \\ &= \langle x,x\rangle + \langle y,y\rangle \\ &= \|x\|^2 + \|y\|^2.\end{aligned}$$

□

14.4 $x_n \to x$ and $y_n \to y$ implies $\langle x_n, y_n\rangle \to \langle x, y\rangle$

Proposition 14.4. *Let (x_n) and (y_n) be two sequences in $\mathscr{V}$. Then the following statements hold:*

1. *If $x_n \to x$ and $y_n \to y$, then $x_n + y_n \to x + y$.*
2. *If $x_n \to x$ and $y_n \to y$, then $\langle x_n, y_n\rangle \to \langle x, y\rangle$. In particular, $\|x_n\| \to \|x\|$.*

Proof.

1. Let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies $\|x_n - x\| < \varepsilon/2$ and $\|y_n - y\| < \varepsilon/2$. Then $n \geq N$ implies

$$\begin{aligned}\|(x_n + y_n) - (x+y)\| &\leq \|x_n - x\| + \|y_n - y\| \\ &< \varepsilon/2 + \varepsilon/2 \\ &= \varepsilon.\end{aligned}$$

2. Since $y_n \to y$, there exists $M \geq 0$ such that $\|y_n\| \leq M$ for all $n \in \mathbb{N}$. Choose such an M and let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies $\|x_n - x\| < \varepsilon/2M$ and $\|y_n - y\| < \varepsilon/2\|x\|$. Then $n \geq N$ implies

$$\begin{aligned}
|\langle x_n, y_n\rangle - \langle x, y\rangle| &= |\langle x_n, y_n\rangle - \langle x, y_n\rangle + \langle x, y_n\rangle - \langle x, y\rangle| \\
&\leq |\langle x_n, y_n\rangle - \langle x, y_n\rangle| + |\langle x, y_n\rangle - \langle x, y\rangle| \\
&= |\langle x_n - x, y_n\rangle| + |\langle x, y_n - y\rangle| \\
&\leq \|x_n - x\|\|y_n\| + \|x\|\|y_n - y\| \\
&\leq \|x_n - x\|M + \|x\|\|y_n - y\| \\
&< \varepsilon/2 + \varepsilon/2 \\
&= \varepsilon.
\end{aligned}$$

To see that $\|x_n\| \to \|x\|$, we just set $y_n = x_n$. Then

$$\begin{aligned}
\|x_n\| &= \sqrt{\langle x_n, x_n\rangle} \\
&\to \sqrt{\langle x, x\rangle} \\
&= \|x\|,
\end{aligned}$$

where we were allowed to take limits inside the square root function since the square root function is continuous on $\mathbb{R}_{\geq 0}$.

□

14.5 Inner-product on $\mathrm{M}_{m\times n}(\mathbb{R})$

Proposition 14.5. *Let $\langle\cdot,\cdot\rangle\colon M_{m\times n}(\mathbb{R}) \times M_{m\times n}(\mathbb{R}) \to \mathbb{R}$ be given by*

$$\langle A, B\rangle = Tr(B^\top A),$$

for all $A, B \in M_n(\mathbb{C})$. Then the pair $(M_n(\mathbb{C}), \langle\cdot,\cdot\rangle)$ forms an inner-product space.

Proof. Linearity in the first argument follows from distributivity of matrix multiplication and from linearity of the trace function: Let $A, B, C \in \mathrm{M}_{m\times n}(\mathbb{R})$. Then

$$\begin{aligned}
\langle A + B, C\rangle &= \mathrm{Tr}(C^\top(A + B)) \\
&= \mathrm{Tr}(C^\top A + C^\top B) \\
&= \mathrm{Tr}(C^\top A) + \mathrm{Tr}(C^\top B) \\
&= \langle A, C\rangle + \langle B, C\rangle.
\end{aligned}$$

Symmetry of $\langle\cdot,\cdot\rangle$ follows from the fact that $\mathrm{Tr}(A) = \mathrm{Tr}(A^\top)$ for all $A \in \mathrm{M}_{m\times n}(\mathbb{R})$: Let $A, B \in \mathrm{M}_{m\times n}(\mathbb{R})$. Then

$$\begin{aligned}
\langle A, B\rangle &= \mathrm{Tr}(B^\top A) \\
&= \mathrm{Tr}((B^\top A)^\top) \\
&= \mathrm{Tr}(A^\top B) \\
&= \langle B, A\rangle.
\end{aligned}$$

Finally, to see positive-definiteness of $\langle\cdot,\cdot\rangle$, let

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \in \mathrm{M}_{m\times n}(\mathbb{R}).$$

Then

$$\begin{aligned}
\langle A, A\rangle &= \mathrm{Tr}(A^\top A) \\
&= \mathrm{Tr}\begin{pmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \cdots & a_{mn} \end{pmatrix}\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \\
&= \sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij}^2.
\end{aligned}$$

is a sum of its entries squared. This implies positive-definiteness. □

14.6 Inner-product on $\mathbb{C}^n$

Proposition 14.6. *Let $\langle \cdot, \cdot \rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ be given by*

$$\langle x, y\rangle = \langle (x_1, \ldots, x_n), (y_1, \ldots, y_n)\rangle = \sum_{i=1}^{n} x_i \overline{y}_i.$$

for all $x, y \in \mathbb{C}^n$. Then the pair $(\mathbb{C}^n, \langle \cdot, \cdot \rangle)$ forms an inner-product space.

Proof. For linearity in the first argument follows from linearity, let $x, y, z \in \mathbb{C}^n$. Then

$$\begin{aligned}
\langle x + y, z\rangle &= \sum_{i=1}^{n} (x_i + y_i)\overline{z}_i \\
&= \sum_{i=1}^{n} x_i \overline{z}_i + \sum_{i=1}^{n} y_i \overline{z}_i \\
&= \langle x, z\rangle + \langle y, z\rangle.
\end{aligned}$$

For conjugate symmetry of $\langle \cdot, \cdot \rangle$, let $x, y \in \mathbb{C}^n$. Then

$$\begin{aligned}
\langle x, y\rangle &= \sum_{i=1}^{n} x_i \overline{y}_i \\
&= \sum_{i=1}^{n} \overline{\overline{x_i \overline{y}_i}} \\
&= \sum_{i=1}^{n} \overline{y_i \overline{x}_i} \\
&= \overline{\langle y, x\rangle}.
\end{aligned}$$

For positive-definiteness of $\langle \cdot, \cdot \rangle$, let $x \in \mathbb{C}^n$. Then

$$\begin{aligned}
\langle x, x\rangle &= \sum_{i=1}^{n} x_i \overline{x}_i \\
&= \sum_{i=1}^{n} |x_i|^2.
\end{aligned}$$

is a sum of its components absolute squared. This implies positive-definiteness. □

14.7 Cauchy-Schwarz

This follows from an easy application of Cauchy-Schwarz, but here's another method (which turns out to be equivalent to Cauchy-Schwarz). We need the following two lemmas:

Lemma 14.1. *Let a and b be nonnegative real numbers. Then we have*

$$2ab \le a^2 + b^2. \tag{28}$$

Proof. We have

$$\begin{aligned} 0 &\leq (a-b)^2 \\ &= a^2 - 2ab + b^2. \end{aligned}$$

Therefore the inequality (28) follows from adding $2ab$. $\square$

Lemma 14.2. *Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be nonnegative real numbers. Then*

$$\left(\sum_{i=1}^n a_i b_i\right)^2 \leq \left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2\right).$$

Proof. We have

$$\begin{aligned} \left(\sum_{i=1}^n a_i b_i\right)^2 &= \sum_{i=1}^n a_i^2 b_i^2 + \sum_{1 \leq i < j \leq n} 2 a_i b_j a_j b_i \\ &\leq \sum_{i=1}^n a_i^2 b_i^2 + \sum_{1 \leq i < j \leq n} (a_i^2 b_j^2 + a_j^2 b_i^2) \\ &= \left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2\right) \end{aligned}$$

where the inequality in the second line follows from Lemma (14.1) applied to $a_i b_j$ and $a_j b_i$. $\square$

Corollary 9. *Let $x, y \in \mathbb{C}^n$. Then*

$$\sum_{i=1}^n |x_i||y_i| \leq \sqrt{\sum_{i=1}^n |x_i|^2}\sqrt{\sum_{i=1}^n |y_i|^2}.$$

Proof. This follows from by taking squares on both sides and applying Lemma (14.2) since the $|x_i|$ and $|y_i|$ are nonnegative real numbers. $\square$

14.8 Inner-product on $\ell^2(\mathbb{N})$

Proposition 14.7. *Let $\ell^2(\mathbb{N})$ be the set of all sequence (x_n) in $\mathbb{C}$ such that*

$$\sum_{n=1}^\infty |x_n|^2 < \infty$$

and let $\langle \cdot, \cdot \rangle \colon \ell^2(\mathbb{N}) \times \ell^2(\mathbb{N}) \to \mathbb{C}$ be given by

$$\langle (x_n), (y_n) \rangle = \sum_{n=1}^\infty x_n \overline{y}_n.$$

for all $(x_n), (y_n) \in \ell^2(\mathbb{N})$. Then the pair $(\ell^2(\mathbb{N}), \langle \cdot, \cdot \rangle)$ forms an inner-product space.

Proof. We first need to show that $\ell^2(\mathbb{N})$ is indeed a vector space. In fact, we will show that $\ell^2(\mathbb{N})$ is a subspace of $\mathbb{C}^N$, the set of all sequences in $\mathbb{C}$. Let $(x_n), (y_n) \in \ell^2(\mathbb{N})$ and $\lambda \in \mathbb{C}$. Then Lemma (14.1) implies

$$\begin{aligned} \sum_{n=1}^\infty |\lambda x_n + y_n|^2 &\leq \sum_{n=1}^\infty |\lambda x_n|^2 + \sum_{n=1}^\infty |y_n|^2 + \sum_{n=1}^\infty 2|\lambda x_n||y_n| \\ &\leq \lambda^2 \sum_{n=1}^\infty |x_n|^2 + \sum_{n=1}^\infty |y_n|^2 + \lambda^2 \sum_{n=1}^\infty |x_n|^2 + |y_n|^2 \\ &< \infty. \end{aligned}$$

Therefore $(\lambda x_n + y_n) \in \ell^2(\mathbb{N})$, which implies $\ell^2(\mathbb{N})$ is a subspace of $\mathbb{C}^{\mathbb{N}}$.

Next, let us show that the inner product converges, and hence is defined everywhere. Let $(x_n), (y_n) \in \ell^2(\mathbb{N})$. Then it follows from Lemma (14.1) that

$$\begin{aligned}
\sum_{n=1}^{\infty} |x_n \overline{y}_n| &= \sum_{n=1}^{\infty} |x_n||y_n| \\
&\leq \sum_{n=1}^{\infty} \frac{|x_n|^2 + |y_n|^2}{2} \\
&= \frac{1}{2}\sum_{n=1}^{\infty} |x_n|^2 + \frac{1}{2}\sum_{n=1}^{\infty} |y_n|^2 \\
&< \infty.
\end{aligned}$$

Therefore $\sum_{n=1}^{\infty} x_n \overline{y}_n$ is absolutely convergent, which implies it is convergent. (We can't use Cauchy-Schwarz here since we haven't yet shown that $\langle \cdot, \cdot \rangle$ is in fact an inner-product).

Finally, let us shows that $\langle \cdot, \cdot \rangle$ is an inner-product. Linearity in the first argument follows from distrubitivity of multiplication and linearity of taking infinite sums. For conjugate symmetry, let $(x_n), (y_n) \in \ell^2(\mathbb{N})$. Then

$$\begin{aligned}
\langle (x_n), (y_n) \rangle &= \sum_{n=1}^{\infty} x_n \overline{y}_n \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} x_n \overline{y}_n \\
&= \overline{\overline{\lim_{N\to\infty} \sum_{n=1}^{N} x_n \overline{y}_n}} \\
&= \overline{\lim_{N\to\infty} \sum_{n=1}^{N} \overline{x_n \overline{y}_n}} \\
&= \overline{\lim_{N\to\infty} \sum_{n=1}^{N} y_n \overline{x}_n} \\
&= \overline{\sum_{n=1}^{\infty} y_n \overline{x}_n} \\
&= \overline{\langle (y_n), (x_n) \rangle},
\end{aligned}$$

where we were allowed to bring the conjugate inside the limit since the conjugate function is continuous on $\mathbb{C}$. For positive-definiteness, let $(x_n) \in \ell^2(\mathbb{N})$. Then

$$\begin{aligned}
\langle (x_n), (x_n) \rangle &= \sum_{n=1}^{\infty} x_n \overline{x}_n \\
&= \sum_{n=1}^{\infty} |x_n|^2 \\
&\geq 0.
\end{aligned}$$

If $\sum_{n=1}^{\infty} |x_n|^2 = 0$, then clearly we must have $x_n = 0$ for all n. □

14.9 Cauchy-Schwarz Application

Proposition 14.8. *Let $(x_n) \in \ell^2(\mathbb{N})$ such that $\sum_{n=1}^{\infty} |x_n|^2 = 1$. Then*

$$\sum_{n=1}^{\infty} \frac{|x_n|}{2^n} \leq \frac{1}{\sqrt{3}}. \tag{29}$$

where the inequality (29) becomes an equality if and only if $|x_n| = \sqrt{3} \cdot 2^{-n}$ for all n.

Proof. By Cauchy-Schwarz, we have

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{|x_n|}{2^n} &= |\langle (|x_n|), (2^{-n}) \rangle| \\
&\leq \|(|x_n|)\| \|(2^{-n})\| \\
&= \sqrt{\sum_{n=1}^{\infty} |x_n|^2} \sqrt{\sum_{n=1}^{\infty} 2^{-2n}} \\
&1 \cdot \sqrt{\sum_{n=0}^{\infty} \left(\frac{1}{4}\right)^n - 1} \\
&= \sqrt{\frac{1}{1 - 1/4} - 1} \\
&= \sqrt{\frac{4}{3} - 1} \\
&= \frac{1}{\sqrt{3}}.
\end{aligned}$$

where the inequality becomes an equality if and only if $(|x_n|)$ and (2^{-n}) are linearly dependent. This means that there is a $\lambda \in \mathbb{C}$ such that $|x_n| = \lambda 2^{-n}$ for all n. To find this λ, write

$$\begin{aligned}
1 &= \sum_{n=1}^{\infty} |x_n|^2 \\
&= \sum_{n=1}^{\infty} |\lambda 2^{-n}|^2 \\
&= |\lambda|^2 \sum_{n=1}^{\infty} \left(\frac{1}{4}\right)^n \\
&= \frac{|\lambda|^2}{3}.
\end{aligned}$$

Thus, any $\lambda \in \mathbb{C}$ such that $|\lambda| = \sqrt{3}$ works. (Actually, we must have $\lambda = \sqrt{3}$ since $\lambda = |x_n| 2^n$ is positive). $\square$

14.10 Cauchy-Schwarz Application

Proposition 14.9. *Let $f \in C[0,1]$ such that $\int_0^1 |f(x)|^2 dx = 1$. Then*

$$\int_0^1 |f(x)| \sin(\pi x) dx \leq \frac{1}{\sqrt{2}},$$

where the inequality becomes an equality if and only if $|f(x)| = \sqrt{2}\sin(\pi x)$.

Proof. First note that

$$\begin{aligned}
\int_0^1 \sin^2(\pi x) dx &= \int_0^1 \cos^2(\pi x) dx \\
&= \int_0^1 (1 - \sin^2(\pi x)) dx
\end{aligned}$$

implies $\int_0^1 \sin^2(\pi x) dx = 1/2$, where in the first equality above we used integration by parts with $u = \sin(\pi x)$ and $dv = \sin(\pi x) dx$. Therefore, by Cauchy-Schwarz, we have

$$\begin{aligned}
\int_0^1 |f(x)| \sin(\pi x) dx &\leq \sqrt{\int_0^1 |f(x)|^2 dx} \cdot \sqrt{\int_0^1 \sin^2(\pi x) dx} \\
&= 1 \cdot \frac{1}{\sqrt{2}} \\
&= \frac{1}{\sqrt{2}},
\end{aligned}$$

where the inequality becomes an equality if and only if $|f(x)|$ and $\sin(\pi x)$ and linearly dependent. This means that there is a $\lambda \in \mathbb{C}$ such that $|f(x)| = \lambda \sin(\pi x)$ for all x. To find this λ, write

$$\begin{aligned} 1 &= \int_0^1 |f(x)|^2 dx \\ &= \int_0^1 |\lambda \sin(\pi x)|^2 dx \\ &= |\lambda|^2 \int_0^1 \sin^2(\pi x) dx \\ &= \frac{|\lambda|^2}{2}. \end{aligned}$$

Thus, any $\lambda \in \mathbb{C}$ such that $|\lambda| = \sqrt{2}$ works. (Actually, we must have $\lambda = \sqrt{2}$ since $\lambda = |f(x)|/\sin(\pi x)$ is positive). ☐

Remark 34. If we tried to apply Lemma (14.1) at each $x \in [0,1]$, we'd only get the weaker result:

$$\begin{aligned} \int_0^1 |f(x)| \sin(\pi x) dx &\leq \frac{1}{2} \left(\int_0^1 |f(x)|^2 dx + \int_0^1 \sin^2(\pi x) dx \right) \\ &= \frac{1}{2} + \frac{1}{4} \\ &= \frac{3}{4}. \end{aligned}$$

15 Homework 2

Throughout this homework, let $\mathscr{V}$ be an inner-product space over $\mathbb{C}$. If $x \in \mathscr{V}$ and $r > 0$, then we define

$$B_r(x) := \{y \in \mathscr{V} \mid \|y - x\| < r\}$$

to be the **open ball centered at** x **and of radius** r. We also define

$$B_r[x] := \{y \in \mathscr{V} \mid \|y - x\| \leq r\}$$

to be the **closed ball centered at** x **and of radius** r.

15.1 Translating Open Balls

Proposition 15.1. *Let $a \in \mathscr{V}$ and $r > 0$. Then*

$$B_r(a) = a + rB_1(0).$$

Proof. We prove this in two steps.

Step 1: We show $B_r(a) = a + B_r(0)$: Let $x \in B_r(a)$, so $\|x - a\| < r$. This implies $x - a \in B_r(0)$. Thus

$$\begin{aligned} x &= a + (x - a) \\ &\in a + B_r(0). \end{aligned}$$

Therefore $B_r(a) \subseteq a + B_r(0)$.

Conversely, let $a + y \in a + B_r(0)$ where $y \in B_r(0)$, so $\|y\| < r$. This implies $\|(a + y) - a\| < r$. In other words, $a + y \in B_r(a)$. Therefore $a + B_r(0) \subseteq B_r(a)$.

Step 2: We show $B_r(0) = rB_1(0)$: Let $x \in B_r(0)$, so $\|x\| < r$. Then since $r > 0$, we have

$$\begin{aligned} 1 &> (1/r)\|x\| \\ &= \|x/r\|. \end{aligned}$$

In other words, $x/r \in B_1(0)$. Thus

$$\begin{aligned} x &= r(x/r) \\ &\in rB_1(0). \end{aligned}$$

Therefore $B_r(0) \subseteq rB_1(0)$.

Conversely, let $ry \in rB_1(0)$ where $y \in B_1(0)$, so $\|y\| < 1$. Then since $r > 0$, we have

$$\begin{aligned} \|ry\| &= r\|y\| \\ &< 1. \end{aligned}$$

In other words, $ry \in B_1(0)$. Therefore $rB_1(0) \subseteq B_r(0)$. □

15.2 Closure of Open Balls

Lemma 15.1. *Let $x \in \mathscr{V}$ and for each $n \in \mathbb{N}$ let $x_n \in \mathscr{V}$ such that $\|x_n - x\| < 1/n$. Then $x_n \to x$.*

Proof. Let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $1/N < \varepsilon$. Then $n \geq N$ implies

$$\begin{aligned} \|x_n - x\| &< 1/n \\ &\leq 1/N \\ &< \varepsilon. \end{aligned}$$

□

Proposition 15.2. *Let $a \in \mathscr{V}$ and $r > 0$. Then*

$$\overline{B_r(a)} = B_r[a].$$

Proof. Let $x \in \overline{B_r(a)}$. Choose a sequence (x_n) of elements in $B_r(a)$ such that $x_n \to x$. Let $\varepsilon > 0$ and choose $N \in \mathbb{N}$ such that $\|x_N - x\| < \varepsilon$. Then

$$\begin{aligned} \|x - a\| &= \|x - x_N + x_N - a\| \\ &\leq \|x - x_N\| + \|x_N - a\| \\ &< \varepsilon + r. \end{aligned}$$

Thus $\|x - a\| < r + \varepsilon$ for all $\varepsilon > 0$. This implies $\|x - a\| \leq r$, or in other words, $x \in B_r[a]$. Thus $\overline{B_r(a)} \subseteq B_r[a]$.

Conversely, let $x \in B_r[a]$ and let $n \in \mathbb{N}$. We first observe that for each $t \in (0,1)$, we have

$$\begin{aligned} \|(x + t(a - x)) - a\| &= \|(1-t)x - (1-t)a\| \\ &= (1-t)\|x - a\| \\ &< r. \end{aligned}$$

Thus $x + t(a - x) \in B_r(a)$ for all $t \in (0,1)$. Now let $n \in \mathbb{N}$. Choose $t_n \in (0,1)$ such that $t_n < \|x - a\|/n$. Then

$$\begin{aligned} \|(x + t_n(a - x)) - x\| &= \|t_n(x - a)\| \\ &= t_n\|x - a\| \\ &< 1/n. \end{aligned}$$

Thus $(x + t_n(a - x))$ is a sequence of elements of elements in $B_r(a)$ such that $x + t_n(a - x) \to x$ (by Lemma (15.1)), hence $x \in \overline{B_r(a)}$. Thus $B_r[a] \subseteq \overline{B_r(a)}$. □

15.3 Closed Set Properties

Lemma 15.2. *Let $A \subseteq \mathscr{V}$.*

Proposition 15.3. *Let $A \subseteq \mathscr{V}$ and let $C_1, C_2 \subseteq \mathscr{V}$ such that C_1 and C_2 are closed. Then*

1. *$\overline{A}$ is a closed set.*
2. *$\overline{A}$ is the smallest closed set that contains A, i.e., for any closed set B such that $A \subseteq B$ we have $\overline{A} \subseteq B$. In particular, $\overline{A} = A$ if and only if A is closed.*
3. *The union of C_1 and C_2 is closed.*
4. *The intersection of C_1 and C_2 is closed.*
5. *An infinite union of closed sets may not be closed.*

Proof.

1. We will show that $\overline{A}$ is closed by showing that $\mathscr{V}\backslash\overline{A}$ is open. To show that $\mathscr{V}\backslash\overline{A}$ is open, it suffices to show that for each $x \in \mathscr{V}\backslash\overline{A}$ there exists an open neighborhood of x which is contained in $\mathscr{V}\backslash\overline{A}$. Assume (for a contradiction) that $\mathscr{V}\backslash\overline{A}$ is not open. Choose $x \in \mathscr{V}\backslash\overline{A}$ such that every open neighborhood of x meets $\overline{A}$. In particular, for each $n \in \mathbb{N}$, there exists $x_n \in B_{1/n}(x) \cap \overline{A}$. Choose such x_n for all $n \in \mathbb{N}$. Then by Lemma (15.1), we must have $x_n \to x$, and hence $x \in \overline{\overline{A}} = \overline{A}$. This is a contradiction.
2. Let B be any closed set which contains A. Suppose $x \in \overline{A}$. Choose a sequence (x_n) of elements in A such that $x_n \to x$. Assume (for a contradiction) that $x \in \mathscr{V}\backslash B$. Choose $\varepsilon > 0$ such that $B_\varepsilon(x) \cap B = \emptyset$ (we can do this since $\mathscr{V}\backslash B$ is open). But then the sequence (x_n) of elements in B cannot converge to x since $x_n \notin B_\varepsilon(x)$ for all $n \in \mathbb{N}$. This is a contradiction. For the last statement. If A is closed, then since $\overline{A}$ is the smallest closed set containing A, we must have $A = \overline{A}$. And if $A = \overline{A}$, then since $\overline{A}$ is the smallest closed set containing A, the set A itself must be closed.
3. Combining 2 with an identity we proved in class, we have
$$\begin{aligned} C_1 \cup C_2 &= \overline{C}_1 \cup \overline{C}_2 \\ &= \overline{C_1 \cup C_2}. \end{aligned}$$
Therefore $C_1 \cup C_2$ is closed.
4. Combining 2 with a couple identities that we proved in class, we have
$$\begin{aligned} \overline{C_1 \cap C_2} &\supseteq C_1 \cap C_2 \\ &= \overline{C}_1 \cap \overline{C}_2 \\ &\supseteq \overline{C_1 \cap C_2}. \end{aligned}$$
Therefore $C_1 \cap C_2$ is closed.
5. Consider $\mathscr{V} = \mathbb{R}$ and $C_n = [0, 1 - 1/n]$ for all $n \in \mathbb{N}$. Then $\bigcup_{n=1}^{\infty} C_n = [0, 1)$, which is not closed in $\mathbb{R}$.

□

15.4 Distance Between Point and Set Properties

Proposition 15.4. *Let $E \subseteq \mathscr{V}$ and let $x, y \in \mathscr{V}$. Then*

1. *$d(x, E) = 0$ if and only if $x \in \overline{E}$;*
2. *$|d(x, E) - d(y, E)| \leq \|x - y\|$.*

Proof.

1. First suppose that $d(x, E) = 0$. For each $n \in \mathbb{N}$, choose $x_n \in E$ such that $\|x_n - x\| < 1/n$ (if we couldn't find such an x_n, then 0 would not be the infinum). Now we apply Lemma (15.1) to find that (x_n) is a sequence of elements in E such that $x_n \to x$. Therefore $x \in \overline{E}$. Conversely, suppose that $x \in \overline{E}$. Choose a sequence (x_n) of elements in E such that $x_n \to x$. Then we have
$$0 \leq d(x, E) < \|x_n - x\|$$
for all $n \in \mathbb{N}$. This implies $d(x, E) = 0$.

2. Without loss of generality, we may assume that $d(x,E) \geq d(y,E)$. Thus we are trying to show that $d(x,E) \leq \|x-y\| + d(y,E)$. Choose $y_n \in E$ such that $\|y_n - y\| < d(y,E) + 1/n$ for all $n \in \mathbb{N}$. Then

$$\begin{aligned} d(x,E) &\leq \|x-y_n\| \\ &= \|x-y+y-y_n\| \\ &\leq \|x-y\| + \|y-y_n\| \\ &< \|x-y\| + d(y,E) + 1/n. \end{aligned}$$

Taking $n \to \infty$ gives us our desired result.

□

15.5 Distance is a Norm

Proposition 15.5. *Let $\mathscr{H}$ be a Hilbert space of $\mathbb{C}$, let $\mathscr{K}$ be a closed subspace of $\mathscr{H}$, let $x, y \in \mathscr{H}$, and let $\lambda \in \mathbb{C}$. Then*

1. $d(\lambda x, \mathscr{K}) = |\lambda| d(x, \mathscr{K})$;
2. $d(x+y, \mathscr{K}) \leq d(x, \mathscr{K}) + d(y, \mathscr{K})$.

Proof.

1. Choose a sequence (y_n) of elements in $\mathcal{A}$ such that

$$\|x-y_n\| < d(x,\mathcal{A}) - 1/n$$

for all $n \in \mathbb{N}$. Then

$$\begin{aligned} \|\lambda x - \lambda y_n\| &= |\lambda| \|x-y_n\| \\ &< |\lambda| d(x,\mathcal{A}) - |\lambda|/n \end{aligned}$$

$$|\lambda| d(x,\mathcal{A}) \leq |\lambda| \|x-y_n\| < d(x,\mathcal{A}) - 1/n$$

2. Let a be the unique element in $\mathscr{K}$ such that $d(x,\mathscr{K}) = \|x-a\|$. Then $\lambda a \in \mathscr{K}$, and so

$$\begin{aligned} |\lambda| d(x,\mathscr{K}) &= |\lambda| \|x-a\| \\ &= \|\lambda x - \lambda a\| \\ &\geq d(\lambda x, \mathscr{K}). \end{aligned}$$

Conversely, let b be the unique element in $\mathscr{K}$ such that $d(\lambda x, \mathscr{K}) = \|x-b\|$. Then $b/\lambda \in \mathscr{K}$, and so

$$\begin{aligned} d(\lambda x, \mathscr{K}) &= \|\lambda x - b\| \\ &= |\lambda| \|x - b/\lambda\| \\ &\geq |\lambda| d(x, \mathscr{K}). \end{aligned}$$

3. Let a be the unique element in $\mathscr{K}$ such that $d(x,\mathscr{K}) = \|x-a\|$ and let b be the unique element in $\mathscr{K}$ such that $d(y,\mathscr{K}) = \|y-b\|$. Then $a+b \in \mathscr{K}$, and so

$$\begin{aligned} d(x+y,\mathscr{K}) &\leq \|x+y-(a+b)\| \\ &= \|x-a\| + \|y-b\| \\ &= d(x,\mathscr{K}) + d(y,\mathscr{K}). \end{aligned}$$

□

16 Homework 3

16.1 Orthogonal Projection $P_{\mathcal{K}}$ Properties

Proposition 16.1. *Let $\mathcal{K}$ be a closed subspace of a Hilbert space $\mathcal{H}$ and let $x \in \mathcal{H}$. Then*

1. *$P_{\mathcal{K}}x = x$ if and only if $x \in \mathcal{K}$.*
2. *$\|P_{\mathcal{K}}x\| = \|x\|$ if and only if $x \in \mathcal{K}$.*
3. *$\langle P_{\mathcal{K}}x, x\rangle = \|P_{\mathcal{K}}x\|^2$.*

Proof.

1. If $P_{\mathcal{K}}x = x$, then it is clear that $x \in \mathcal{K}$ since $P_{\mathcal{K}}x \in \mathcal{K}$. For the reverse direction, suppose $x \in \mathcal{K}$. Then

$$\begin{aligned} 0 &= \|x - x\| \\ &\geq d(x, \mathcal{K}) \\ &= \|x - P_{\mathcal{K}}x\| \\ &\geq 0 \end{aligned}$$

implies $\|x - x\| = d(x, \mathcal{K}) = \|x - P_{\mathcal{K}}x\|$, and so by uniqueness of $P_{\mathcal{K}}x$, we must have $x = P_{\mathcal{K}}x$.

2. If $x \in \mathcal{K}$, then it is clear that $\|P_{\mathcal{K}}x\| = \|x\|$ since $x = P_{\mathcal{K}}x$ by 1. For the reverse direction, suppose $\|P_{\mathcal{K}}x\| = \|x\|$. Since $\langle x - P_{\mathcal{K}}x, P_{\mathcal{K}}x\rangle = 0$, the Pythagorean Theorem[3] implies

$$\begin{aligned} \|x\|^2 &= \|x - P_{\mathcal{K}}x + P_{\mathcal{K}}x\|^2 \\ &= \|x - P_{\mathcal{K}}x\|^2 + \|P_{\mathcal{K}}x\|^2 \\ &= \|x - P_{\mathcal{K}}x\|^2 + \|x\|^2. \end{aligned}$$

Thus $\|x - P_{\mathcal{K}}x\|^2 = 0$, which implies $x = P_{\mathcal{K}}x$ since the metric is positive definite.

3. We have

$$\begin{aligned} 0 &= \langle x - P_{\mathcal{K}}x, P_{\mathcal{K}}x\rangle \\ &= \langle x, P_{\mathcal{K}}x\rangle - \langle P_{\mathcal{K}}x, P_{\mathcal{K}}x\rangle \\ &= \langle x, P_{\mathcal{K}}x\rangle - \|P_{\mathcal{K}}x\|^2, \end{aligned}$$

which implies $\langle x, P_{\mathcal{K}}x\rangle = \|P_{\mathcal{K}}x\|^2$. Since $\|P_{\mathcal{K}}x\|^2$ is a real number, this implies $\langle P_{\mathcal{K}}x, x\rangle = \|P_{\mathcal{K}}x\|^2$.

□

16.2 $\mathcal{K}_1 \subseteq \mathcal{K}_2$ if and only if $\langle P_{\mathcal{K}_1}x, x\rangle \leq \langle P_{\mathcal{K}_2}x, x\rangle$ for all $x \in \mathcal{H}$

Proposition 16.2. *Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be closed subspaces of a Hilbert space $\mathcal{H}$. Then $\mathcal{K}_1 \subseteq \mathcal{K}_2$ if and only if $\langle P_{\mathcal{K}_1}x, x\rangle \leq \langle P_{\mathcal{K}_2}x, x\rangle$ for all $x \in \mathcal{H}$.*

Proof. By Proposition (16.1), we can replace the condition $\langle P_{\mathcal{K}_1}x, x\rangle \leq \langle P_{\mathcal{K}_2}x, x\rangle$ for all $x \in \mathcal{H}$ with $\|P_{\mathcal{K}_1}x\|^2 \leq \|P_{\mathcal{K}_2}x\|^2$ for all $x \in \mathcal{H}$. Suppose $\mathcal{K}_1 \subseteq \mathcal{K}_2$. Then

$$\begin{aligned} \|x - P_{\mathcal{K}_2}x\| &= \mathrm{d}(x, \mathcal{K}_2) \\ &= \inf\{\|x - y\| \mid y \in \mathcal{K}_2\} \\ &\leq \inf\{\|x - y\| \mid y \in \mathcal{K}_1\} \\ &= \mathrm{d}(x, \mathcal{K}_1) \\ &= \|x - P_{\mathcal{K}_1}x\|. \end{aligned}$$

[3] Theorem (16.1) in the Appendix.

Therefore by the Pythagorean Theorem, we have

$$\begin{aligned}\|P_{\mathcal{K}_1}x\|^2 &= \|x\|^2 - \|x - P_{\mathcal{K}_1}x\|^2 \\ &\leq \|x\|^2 - \|x - P_{\mathcal{K}_2}x\|^2 \\ &= \|P_{\mathcal{K}_2}x\|^2.\end{aligned}$$

Conversely, suppose $\|P_{\mathcal{K}_1}x\|^2 \leq \|P_{\mathcal{K}_2}x\|^2$ for all $x \in \mathcal{H}$. Equivalently, by the Pythagorean Theorem, we have

$$\begin{aligned}\|x - P_{\mathcal{K}_1}x\|^2 &= \|x\|^2 - \|P_{\mathcal{K}_1}x\|^2 \\ &\leq \|x\|^2 - \|P_{\mathcal{K}_2}x\|^2 \\ &= \|x - P_{\mathcal{K}_2}x\|^2\end{aligned}$$

for all $x \in \mathcal{K}_1$. Now let $x \in \mathcal{K}_1$. Then $x = P_{\mathcal{K}_1}x$ by Proposition (16.1). Thus

$$\begin{aligned}0 &= \|x - x\|^2 \\ &= \|x - P_{\mathcal{K}_1}x\|^2 \\ &\geq \|x - P_{\mathcal{K}_2}x\|^2,\end{aligned}$$

which implies $x = P_{\mathcal{K}_2}x$ since the metric is positive definite. Applying Proposition (16.1) again, we see that $x \in \mathcal{K}_2$, and hence $\mathcal{K}_1 \subseteq \mathcal{K}_2$. □

16.3 $\|P_{\mathcal{K}^\perp}x\| = d(x, \mathcal{K})$ for all $x \in \mathcal{H}$

Proposition 16.3. *Let $\mathcal{K}$ be a closed subspace of a Hilbert space $\mathcal{H}$. Then $\|P_{\mathcal{K}^\perp}x\| = d(x, \mathcal{K})$ for all $x \in \mathcal{H}$.*

Proof. From a theorem[4] we proved in class, we know that x can be uniquely decomposed as

$$x = P_{\mathcal{K}}x + (x - P_{\mathcal{K}}x), \tag{30}$$

for unique $P_{\mathcal{K}}x \in \mathcal{K}$ and unique $x - P_{\mathcal{K}}x \in \mathcal{K}^\perp$. Since $\mathcal{K}^\perp$ is another closed subspace[5] of $\mathcal{H}$, we can uniquely decompose x as

$$x = P_{\mathcal{K}^\perp}x + (x - P_{\mathcal{K}^\perp}x) \tag{31}$$

for unique $P_{\mathcal{K}^\perp}x \in \mathcal{K}^\perp$ and unique $x - P_{\mathcal{K}^\perp}x \in (\mathcal{K}^\perp)^\perp = \mathcal{K}$. It follows from uniqueness of (30) and (31) that

$$P_{\mathcal{K}^\perp}x = x - P_{\mathcal{K}}x \quad \text{and} \quad P_{\mathcal{K}}x = x - P_{\mathcal{K}^\perp}x$$

In particular, we have

$$\begin{aligned}d(x, \mathcal{K}) &= \|x - P_{\mathcal{K}}x\| \\ &= \|P_{\mathcal{K}^\perp}x\|.\end{aligned}$$

□

16.4 span(E) Properties

Proposition 16.4. *Let $\mathcal{V}$ be an inner-product space and let $E \subseteq V$. Define*

$$\mathit{Span}(E) := \left\{ \sum_{i=1}^{n} \lambda_i v_i \mid n \in \mathbb{N},\ \lambda_i \in \mathbb{C}, \text{ and } v_i \in E \text{ for } 1 \leq i \leq n \right\}$$

Then

1. *Span(E) is a subspace of $\mathcal{V}$.*
2. *Span(E) is the smallest subspace containing E.*

[4] Theorem (16.3) in the Appendix.
[5] This was also shown in class and is given in Theorem (16.3) in the Appendix

Proof.

1. Let $\lambda \in \mathbb{C}$ and let $v, w \in \mathrm{Span}(E)$ where

$$v = \sum_{i=1}^{m} \lambda_i v_i \text{ and } w = \sum_{j=1}^{n} \mu_j w_j$$

where $\lambda_i, \mu_j \in \mathbb{C}$ and $v_i, w_j \in E$ for all $1 \leq i \leq m$ and $1 \leq j \leq n$. Then

$$\begin{aligned} \lambda v + w &= \sum_{i=1}^{m} \lambda \lambda_i v_i + \sum_{j=1}^{n} \lambda \mu_j w_j \\ &= \sum_{i=1}^{m} \kappa_k u_k \in \mathrm{Span}(E), \end{aligned}$$

where $\kappa_k = \lambda \lambda_k$ and $u_k = v_k$ for $1 \leq k \leq m$ and $\kappa_k = \lambda \mu_{k-m}$ and $u_k = w_{k-m}$ for $m < k \leq m+n$. Therefore $\mathrm{Span}(E)$ is a subspace of $\mathcal{V}$.

2. Let $\mathcal{U}$ be any subspace of $\mathcal{V}$ which contains E. Suppose that $v \in \mathrm{Span}(E)$, where

$$v = \sum_{i=1}^{n} \lambda_i v_i \tag{32}$$

where $\lambda_i \in \mathbb{C}$ and $v_i \in E$ for all $1 \leq i \leq n$. As $\mathcal{U}$ is a subspace of $\mathcal{V}$, it must be closed under taking finite linear combinations of elements in $\mathcal{U}$. Since for each $1 \leq i \leq n$, we have $\lambda_i \in \mathbb{C}$ and $v_i \in E \subseteq \mathcal{U}$, it is clear that from (32) that $v \in \mathcal{U}$. Thus $\mathrm{Span}(E) \subseteq \mathcal{U}$.

□

16.5 $\overline{\mathrm{Span}}(E)$ Properties

Proposition 16.5. *Let $\mathcal{V}$ be an inner-product space and let $E \subseteq V$. Define the closed span of E, denoted $\overline{Span}(E)$, as the closure of $Span(E)$. Then*

1. *$\overline{Span}(E)$ is a closed subspace of $\mathcal{V}$.*
2. *$\overline{Span}(E)$ is the smallest closed subspace containing E.*

Proof.

1. By Proposition (16.4), we know that $\mathrm{Span}(E)$ is a subspace. By a theorem[6] which we proved in class, the closure of a subspace is a closed subspace. Therefore $\overline{\mathrm{Span}}(E)$ is a closed subspace of $\mathcal{V}$.

2. Let $\mathcal{U}$ be any closed subspace of $\mathcal{V}$ which contains E. By Proposition (16.4), we know that $\mathrm{Span}(E) \subseteq \mathcal{U}$. Therefore

$$\begin{aligned} \overline{\mathrm{Span}}(E) &= \overline{\mathrm{Span}(E)} \\ &\subseteq \overline{\mathcal{U}} \\ &= \mathcal{U}, \end{aligned}$$

where $\mathcal{U} = \overline{\mathcal{U}}$ since $\mathcal{U}$ is closed (this was proved in the second homework).

□

[6]Theorem (16.2) in the Appendix.

16.6 $(E^{\perp})^{\perp} = \overline{\text{Span}}(E)$

Proposition 16.6. *Let $\mathcal{H}$ be a Hilbert space and let $E \subseteq \mathcal{H}$. Then*

$$(E^{\perp})^{\perp} = \overline{Span}(E).$$

Proof. First note that $E \subseteq (E^{\perp})^{\perp}$. Indeed, if $x \in E$, then $\langle x, y \rangle = 0$ for all $y \in E^{\perp}$, hence $x \in (E^{\perp})^{\perp}$. Also, from a theorem[7] we proved in class, we know that $(E^{\perp})^{\perp}$ is a closed subspace. Thus $(E^{\perp})^{\perp}$ is a closed subspace which contains E, which implies $(E^{\perp})^{\perp} \supseteq \overline{\text{Span}}(E)$ by Proposition (16.5).

Converesely, since taking orthononal complements is inclusion-reversing[8], $E \subseteq \overline{\text{Span}}(E)$ implies $E^{\perp} \supseteq \overline{\text{Span}}(E)^{\perp}$ which implies $(E^{\perp})^{\perp} \subseteq (\overline{\text{Span}}(E)^{\perp})^{\perp} = \overline{\text{Span}}(E)$, where the last equality follows from a theorem[9] in class. □

Appendix

Theorem 16.1. *(Pythagorean Theorem) Let x and y be nonzero vectors in $\mathcal{V}$ such that $\langle x, y \rangle = 0$ (we call such vectors **orthogonal** to one another). Then*

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2.$$

Proof. We have

$$\begin{aligned}
\|x + y\|^2 &= \langle x + y, x + y \rangle \\
&= \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle \\
&= \langle x, x \rangle + \langle y, y \rangle \\
&= \|x\|^2 + \|y\|^2
\end{aligned}$$

□

Theorem 16.2. *Let $\mathcal{U}$ be a subspace of $\mathcal{V}$. Then $\overline{\mathcal{U}}$ is a subspace of $\mathcal{V}$.*

Proof. Let $x, y \in \overline{\mathcal{U}}$ and $\lambda \in \mathbb{C}$. Let (x_n) and (y_n) be two sequences of elements in $\mathcal{U}$ such that $x_n \to x$ and $y_n \to y$. Then $(\lambda x_n + y_n)$ is a sequence of elements in $\mathcal{U}$ such that $\lambda x_n + y_n \to \lambda x + y$. Therefore $\lambda x + y \in \overline{\mathcal{U}}$, which implies $\overline{\mathcal{U}}$ is a subspace of $\mathcal{V}$. □

Theorem 16.3. *Let $\mathcal{H}$ be a Hilbert space and let $\mathcal{K} \subseteq \mathcal{L} \subseteq \mathcal{H}$. Then*

1. *we have $\mathcal{K}^{\perp} \supseteq \mathcal{L}^{\perp}$.*

2. *$\mathcal{K}^{\perp}$ is a closed subspace of $\mathcal{H}$.*

3. *If $\mathcal{K}$ is a closed subspace of $\mathcal{H}$, then every $x \in \mathcal{H}$ can be decomposed in a* unique *way as a sum of a vector in $\mathcal{K}$ and a vector in $\mathcal{K}^{\perp}$. In other words, we have $\mathcal{H} = \mathcal{K} \oplus \mathcal{K}^{\perp}$.*

4. *If $\mathcal{K}$ is a closed subspace of $\mathcal{H}$, then $(\mathcal{K}^{\perp})^{\perp} = \mathcal{K}$.*

Proof.

1. We have

$$\begin{aligned}
x \in \mathcal{L}^{\perp} &\implies \langle x, y \rangle = 0 \text{ for all } y \in \mathcal{L} \\
&\implies \langle x, y \rangle = 0 \text{ for all } y \in \mathcal{K} \\
&\implies x \in \mathcal{K}^{\perp}.
\end{aligned}$$

Thus $\mathcal{K}^{\perp} \supseteq \mathcal{L}^{\perp}$.

[7] Theorem (16.3) in the Appendix
[8] Theorem (16.3) in the Appendix
[9] Theorem (16.3) in the Appendix

2. First we show that $\mathcal{K}^\perp$ is a subspace of $\mathcal{V}$. Let $x, z \in \mathcal{K}^\perp$ and $\lambda \in \mathbb{C}$. Then

$$\begin{aligned}\langle x + \lambda z, y \rangle &= \langle x, y \rangle + \lambda \langle z, y \rangle \\ &= 0\end{aligned}$$

for all $y \in \mathcal{K}$. This implies $\mathcal{K}^\perp$ is a subspace of $\mathcal{V}$. Now we will show that $\mathcal{K}^\perp$ is closed. Let (x_n) be a sequence of points in $\mathcal{K}^\perp$ such that $x_n \to x$ for some $x \in \mathcal{H}$. Then since $\langle x_n, y \rangle = 0$ for all $n \in \mathbb{N}$ and $y \in \mathcal{K}$, we have

$$\begin{aligned}\langle x, y \rangle &= \lim_{n\to\infty} \langle x_n, y \rangle \\ &= \lim_{n\to\infty} 0 \\ &= 0.\end{aligned}$$

for all $y \in \mathcal{K}$. Therefore $x \in \mathcal{K}^\perp$, which implies $\mathcal{K}^\perp$ is closed.

3. Let $x \in \mathcal{H}$. Then $x = \mathrm{P}_\mathcal{K} x + x - \mathrm{P}_\mathcal{K} x$ where $\mathrm{P}_\mathcal{K} x \in \mathcal{K}$ and $x - \mathrm{P}_\mathcal{K} x \in \mathcal{K}^\perp$. This establishes existence. For uniqueness, first note that $\mathcal{K} \cap \mathcal{K}^\perp = \{0\}$. Indeed, if $y \in \mathcal{K} \cap \mathcal{K}^\perp$, then we must have $\langle y, y \rangle = 0$, which implies $y = 0$. Now suppose that $x = y + z$ is another decomposition of x where $y \in \mathcal{K}$ and $z \in \mathcal{K}^\perp$. Then we have

$$(\mathrm{P}_\mathcal{K} x) + (x - \mathrm{P}_\mathcal{K} x) = x = y + z$$

implies $\mathrm{P}_\mathcal{K} x - y = (x - \mathrm{P}_\mathcal{K} x) - z$ which implies $\mathrm{P}_\mathcal{K} x - y \in \mathcal{K} \cap \mathcal{K}^\perp = \{0\}$ and $(x - \mathrm{P}_\mathcal{K} x) - z \in \mathcal{K} \cap \mathcal{K}^\perp = \{0\}$.

4. Let $x \in \mathcal{K}$. Then $\langle x, y \rangle = 0$ for all $y \in \mathcal{K}^\perp$. Thus $x \in (\mathcal{K}^\perp)^\perp$, and so $\mathcal{K} \subseteq (\mathcal{K}^\perp)^\perp$. Conversely, let $x \in (\mathcal{K}^\perp)^\perp$. Then $\langle x, y \rangle = 0$ for all $y \in \mathcal{K}^\perp$. In particular, we have

$$\begin{aligned}\|x - \mathrm{P}_\mathcal{K} x\|^2 &= \langle x - \mathrm{P}_\mathcal{K} x, x - \mathrm{P}_\mathcal{K} x \rangle \\ &= \langle x, x - \mathrm{P}_\mathcal{K} x \rangle - \langle \mathrm{P}_\mathcal{K} x, x - \mathrm{P}_\mathcal{K} x \rangle \\ &= 0 - 0 \\ &= 0,\end{aligned}$$

which implies $x = \mathrm{P}_\mathcal{K} x$. This implies $x \in \mathcal{K}$, and hence $(\mathcal{K}^\perp)^\perp \subseteq \mathcal{K}$.

□

17 Homework 4

17.1 Equivalent Definitions of Norm of Operator

Proposition 17.1. *Let $\mathcal{H}$ be a Hilbert space and let $T \colon \mathcal{H} \to \mathcal{H}$ be a bounded operator. Then*

1. $\|T\| = \sup\{\|Tx\| \mid \|x\| = 1\}$;
2. $\|T\| = \sup\left\{\frac{\|Tx\|}{\|x\|} \mid x \in \mathcal{H}\backslash\{0\}\right\}$.

Proof.

1. First note that

$$\begin{aligned}\sup\{\|Tx\| \mid \|x\| = 1\} &\leq \sup\{\|Tx\| \mid \|x\| \leq 1\} \\ &= \|T\|.\end{aligned}$$

We prove the reverse inequality by contradiction. Assume that $\|T\| > \sup\{\|Tx\| \mid \|x\| = 1\}$. Choose $\varepsilon > 0$ such that

$$\|T\| - \varepsilon > \sup\{\|Tx\| \mid \|x\| = 1\} \tag{33}$$

Next, choose $x \in \mathcal{H}$ such that $\|x\| \leq 1$ and $\|Tx\| \geq \|T\| - \varepsilon$. Then since $\|x\| \leq 1$ and $\left\| \frac{x}{\|x\|} \right\| = 1$, we have

$$\begin{aligned}
\|T\| &\geq \left\| T\left(\frac{x}{\|x\|}\right) \right\| \\
&= \frac{\|Tx\|}{\|x\|} \\
&\geq \|Tx\| \\
&> \|T\| - \varepsilon,
\end{aligned}$$

and this contradicts (33).

2. We have

$$\begin{aligned}
\sup\left\{ \frac{\|Tx\|}{\|x\|} \mid x \in \mathcal{H}\backslash\{0\} \right\} &= \sup\left\{ \left\| T\left(\frac{x}{\|x\|}\right) \right\| \mid x \in \mathcal{H}\backslash\{0\} \right\} \\
&= \sup\{\|Ty\| \mid \|y\| = 1\} \\
&= \|T\|,
\end{aligned}$$

where the last equality follows from 1. □

17.2 Multiplication by $k \in C[a,b]$ is Bounded Operator

Proposition 17.2. *Let $k \in C[a,b]$. Then the operator $T\colon C[a,b] \to C[a,b]$ defined by*

$$Tf = kf$$

for all $f \in C[a,b]$ is bounded. It's norm will be explicitly computed in the proof below.

Proof. We first show it is linear. Let $f, g \in C[a,b]$ and let $\lambda, \mu \in \mathbb{C}$. Then we have

$$\begin{aligned}
T(\lambda f + \mu g) &= k(\lambda f + \mu g) \\
&= \lambda kf + \mu kg \\
&= \lambda T(f) + \mu T(g).
\end{aligned}$$

Thus, T is linear.

Next we show it is bounded. If $k = 0$, then $\|T\| = 0$, so assume $k \neq 0$. Since k is continuous on the compact interval $[a,b]$, there exists $c \in [a,b]$ such that $|k(x)| \leq |k(c)|$ for all $x \in [a,b]$. Choose such a $c \in [a,b]$ and let $f \in C[a,b]$ such that $\|f\| \leq 1$. Then

$$\begin{aligned}
\|Tf\| &= \|kf\| \\
&= \sqrt{\int_a^b |k(x)|^2 |f(x)|^2 dx} \\
&\leq |k(c)| \sqrt{\int_a^b |f(x)|^2 dx} \\
&\leq |k(c)|.
\end{aligned}$$

implies $\|T\| \leq |k(c)|$, and hence T is bounded.

To find the norm of T, let $\varepsilon > 0$ such that $\varepsilon < |k(c)|$. Without loss of generality, assume that $c < b$ (if $c = b$, then we swap the role of b with a in the argument which follows). Choose $c' \in (c,b)$ such that $|k(x)| \geq |k(c)| - \varepsilon$ for all $x \in (c,c')$ (such a c' must exist since k is continuous) and choose f to be a nonzero continuous function in $C[a,b]$ which vanishes outside the interval (c,c'). Then

$$|k(x)||f(x)| \geq (|k(c)| - \varepsilon)|f(x)|$$

for all $x \in (a,b)$. In particular, this implies

$$\begin{aligned}
\|Tf\| &= \|kf\| \\
&= \sqrt{\int_a^b |k(x)f(x)|^2 dx} \\
&\geq \sqrt{\int_a^b (|k(c)| - \varepsilon)|f(x)|^2 dx} \\
&= (|k(c)| - \varepsilon)\sqrt{\int_a^b |f(x)|^2 dx} \\
&= (|k(c)| - \varepsilon)\|f\|.
\end{aligned}$$

Therefore $\|T(f/\|f\|)\| \geq |k(c)| - \varepsilon$, and this implies

$$\|T\| \geq |k(c)| - \varepsilon \tag{34}$$

Since (34) holds for all $\varepsilon > 0$, we must have $\|T\| \geq |k(c)|$. Thus $\|T\| = |k(c)|$. □

17.3 Gram_Schmidt Properties

Proposition 17.3. *Let $\{x_n \mid n \in \mathbb{N}\}$ be a linearly independent set of vectors in a Hilbert space $\mathcal{H}$. Consider the so called Gram-Schmidt process: set $e_1 = \frac{1}{\|x_1\|}x_1$. Proceed inductively. If $e_1, e_2, \ldots, e_{n-1}$ are computed, compute e_n in two steps by*

$$f_n := x_n - \sum_{k=1}^{n-1} \langle x_n, e_k \rangle e_k, \text{ and then set } e_n := \frac{1}{\|f_n\|} f_n.$$

Then

1. *for every $N \in \mathbb{N}$ we have span$\{x_1, x_2, \ldots, x_N\}$ = span$\{e_1, e_2, \ldots, e_N\}$;*
2. *the set $\{e_n \mid n \in \mathbb{N}\}$ is an orthonormal set in $\mathcal{H}$;*
3. *if $\overline{\text{span}}\{x_n \mid n \in \mathbb{N}\} = \mathcal{H}$, then $\{e_n \mid n \in \mathbb{N}\}$ is an orthonormal basis for $\mathcal{H}$.*

Proof.
1. Let $N \in \mathbb{N}$. Then for each $1 \leq n \leq N$, we have

$$x_n = \sum_{k=1}^{n-1} \langle x_n, e_k \rangle e_k + \|f_n\| e_n.$$

This implies $\text{span}\{x_1, x_2, \ldots, x_N\} \subseteq \text{span}\{e_1, e_2, \ldots, e_N\}$. We show the reverse inclusion by induction on n such that $1 \leq n \leq N$. The base case $n = 1$ being $\text{span}\{x_1\} \supseteq \text{span}\{e_1\}$, which holds since $e_1 = \frac{1}{\|x_1\|}x_1$. Now suppose for some n such that $1 \leq n < N$ we have

$$\text{span}\{x_1, x_2, \ldots, x_k\} \supseteq \text{span}\{e_1, e_2, \ldots, e_k\} \tag{35}$$

for all $1 \leq k \leq n$. Then

$$e_{n+1} = \frac{1}{\|f_n\|}x_n - \sum_{k=1}^{n} \frac{1}{\|f_n\|}\langle x_n, e_k \rangle e_k \in \text{span}\{x_1, x_2, \ldots, x_n\}.$$

where we used the induction step (35) on the e_k's ($1 \leq k \leq n$). Therefore

$$\text{span}\{x_1, x_2, \ldots, x_k\} \supseteq \text{span}\{e_1, e_2, \ldots, e_k\}$$

for all $1 \leq k \leq n+1$, and this proves our claim.

2. By construction, we have $\langle e_n, e_n \rangle = 1$ for all $n \in \mathbb{N}$. Thus, it remains to show that $\langle e_m, e_n \rangle = 0$ whenever $m \neq n$. We prove by induction on $n \geq 2$ that $\langle e_n, e_m \rangle = 0$ for all $m < n$. Proving this also give us $\langle e_m, e_n \rangle = 0$ for all $m < n$, since

$$\begin{aligned}\langle e_m, e_n \rangle &= \overline{\langle e_n, e_m \rangle} \\ &= \overline{0} \\ &= 0.\end{aligned}$$

The base case is

$$\begin{aligned}\langle e_2, e_1 \rangle &= \frac{1}{\|x_1\|\|f_2\|} \left\langle \left(x_2 - \frac{\langle x_2, x_1 \rangle}{\langle x_1, x_1 \rangle} x_1 \right), x_1 \right\rangle \\ &= \frac{1}{\|x_1\|\|f_2\|} \left(\langle x_2, x_1 \rangle - \langle x_2, x_1 \rangle \right) \\ &= 0\end{aligned}$$

Now suppose that $n > 2$ and that $\langle e_n, e_m \rangle = 0$ for all $m < n$. Then

$$\begin{aligned}\langle e_{n+1}, e_m \rangle &= \frac{1}{\|f_{n+1}\|} \langle x_{n+1} - \sum_{k=1}^{n} \langle x_{n+1}, e_k \rangle e_k, e_m \rangle \\ &= \frac{1}{\|f_{n+1}\|} \left(\langle x_{n+1}, e_m \rangle - \sum_{k=1}^{n} \langle x_{n+1}, e_k \rangle \langle e_k, e_m \rangle \right) \\ &= \frac{1}{\|f_{n+1}\|} \left(\langle x_{n+1}, e_m \rangle - \langle x_{n+1}, e_m \rangle \langle e_m, e_m \rangle \right) \\ &= \frac{1}{\|f_{n+1}\|} \left(\langle x_{n+1}, e_m \rangle - \langle x_{n+1}, e_m \rangle \right) \\ &= 0,\end{aligned}$$

for all $m < n+1$, where we used the induction hypothesis to get from the second line to the third line. This proves the induction step, which finishes the proof of part 2 of the proposition.

3. By 2, we know that $\{e_n \mid n \in \mathbb{N}\}$ is an orthonormal set. Thus, it suffices to show that $\{e_n \mid n \in \mathbb{N}\}$ is complete. To do this, we use the criterion that the set $\{e_n \mid n \in \mathbb{N}\}$ is complete if and only if the only $x \in \mathcal{H}$ such that $\langle x, e_n \rangle = 0$ for all $n \in \mathbb{N}$ is $x = 0$.

Let $x \in \mathcal{H}$ and suppose $\langle x, e_n \rangle = 0$ for all $n \in \mathbb{N}$. Then

$$\begin{aligned}\langle x, x_n \rangle &= \left\langle x, \sum_{k=1}^{n-1} \langle x_n, e_k \rangle e_k + \|f_n\| e_n \right\rangle \\ &= \sum_{k=1}^{n-1} \langle x_n, e_k \rangle \langle x, e_k \rangle + \|f_n\| \langle x, e_n \rangle \\ &= 0\end{aligned}$$

for all $n \in \mathbb{N}$. Since $\{x_n \mid n \in \mathbb{N}\}$ is complete, this implies $x = 0$. Therefore $\{e_n \mid n \in \mathbb{N}\}$ is complete. □

17.4 Gram-Schmidt Example Worked out on Legendre Polynomials

Example 17.1. The first three Legendre polynomials are

$$P_1(x) = 1, \quad P_2(x) = x, \quad , P_3(x) = \frac{1}{2}(3x^2 - 1).$$

We apply Gram-Schmidt process to the polynomials $1, x, x^2$ in the space $C[-1, 1]$ to get scalar multiples of the Legendre polynomials above. First we set $f_1(x) = 1$ and then calculate

$$\begin{aligned}\|f_1(x)\| &= \sqrt{\int_{-1}^{1} dx} \\ &= \sqrt{2}.\end{aligned}$$

Thus we set $e_1(x) = 1/\sqrt{2}$. Next we calculate

$$\begin{aligned} f_1(x) &= x - \left\langle x, \frac{1}{\sqrt{2}} \right\rangle \frac{1}{\sqrt{2}} \\ &= x - \frac{1}{2}\int_{-1}^{1} x dx \\ &= x. \end{aligned}$$

Next we calculate

$$\begin{aligned} \|f_1(x)\| &= \sqrt{\int_{-1}^{1} x^2 dx} \\ &= \sqrt{\frac{2}{3}}. \end{aligned}$$

Thus we set $e_2(x) = \sqrt{3/2}x$. Next we calculate

$$\begin{aligned} f_2(x) &= x^2 - \left\langle x^2, \sqrt{\frac{3}{2}}x \right\rangle \sqrt{\frac{3}{2}}x - \left\langle x^2, \sqrt{\frac{1}{2}} \right\rangle \sqrt{\frac{1}{2}} \\ &= x^2 - \frac{3}{2}x\int_{-1}^{1} x^3 dx - \frac{1}{2}\int_{-1}^{1} x^2 dx \\ &= x^2 - \frac{1}{3}. \end{aligned}$$

Then we finally calculate

$$\begin{aligned} \|f_2(x)\| &= \sqrt{\int_{-1}^{1} \left(x^2 - \frac{1}{3}\right)^2 dx} \\ &= \sqrt{\int_{-1}^{1} \left(x^4 - \frac{2}{3}x^2 + \frac{1}{9}\right) dx} \\ &= \sqrt{\int_{-1}^{1} x^4 dx - \frac{2}{3}\int_{-1}^{1} x^2 dx + \frac{1}{9}\int_{-1}^{1} dx} \\ &= \sqrt{\frac{2}{5} - \frac{4}{9} + \frac{2}{9}} \\ &= \sqrt{\frac{8}{45}}. \end{aligned}$$

Thus we set $e_3(x) = \sqrt{45/8}(x^2 - 1/3)$. Now observe that

$$\begin{aligned} P_1(x) &= \sqrt{2}e_1(x) \\ P_2(x) &= \sqrt{\frac{2}{3}}e_2(x) \\ P_3(x) &= \sqrt{\frac{2}{5}}e_3(x) \end{aligned}$$

17.5 Minimizing Integral Example

For this problem, we needed to establish some basic results which we proved in the Appendix.

Proposition 17.4. *The expression*

$$\int_{-1}^{1} |x^3 - a - bx - cx^2|^2 \mathrm{d}x. \tag{36}$$

is minimized in $a, b, c \in \mathbb{C}$ *if and only if* $a = 0$, $b = 3/5$, *and* $c = 0$.

Proof. Let

$$\mathcal{H} = \{p(x) \in \mathbb{C}[x] \mid \deg(p(x)) \leq 3\} \quad \text{and} \quad \mathcal{K} = \{p(x) \in \mathbb{C}[x] \mid \deg(p(x)) \leq 2\}.$$

Then $\mathcal{H}$ and $\mathcal{K}$ are subspaces of $C[-1,1]$, Proposition (17.7) implies they are inner-product spaces with the inner-product inherited from $C[-1,1]$. Since $\mathcal{H}$ is finite dimensional, Proposition (17.8) implies $\mathcal{H}$ is a separable Hilbert space. Since $\mathcal{K}$ is a finite dimensional subspace of $\mathcal{H}$, Proposition (17.9) implies $\mathcal{K}$ is closed in $\mathcal{H}$. Let $\{e_1, e_2, e_3\}$ be the orthonormal basis computed in problem 4. A proposition proved in class implies

$$\begin{aligned}
\mathrm{P}_{\mathcal{K}}(x^3) &= \langle x^3, e_1\rangle e_1 + \langle x^3, e_2\rangle e_2 + \langle x^3, e_3\rangle e_3 \\
&= \frac{1}{2}\int_{-1}^{1} x^3 dx + \frac{3}{2}x\int_{-1}^{1} x^4 dx + \frac{45}{8}\left(x^2 - \frac{1}{3}\right)\int_{-1}^{1} x^3\left(x^2 - \frac{1}{3}\right) \mathrm{d}x \\
&= \frac{3}{5}x.
\end{aligned}$$

where we used the fact that $x^3(x^2 - 1/3)$ is an odd function to get $\int_{-1}^{1} x^3(x^2-1/3)dx = 0$. Therefore

$$\begin{aligned}
\int_{-1}^{1}\left|x^3 - \frac{3}{5}x\right|^2 \mathrm{d}x &= \|x^3 - \mathrm{P}_{\mathcal{K}}(x^3)\|^2 \\
&= \inf\left\{\|x^3 - (a + bx + cx^2)\|^2 \mid a + bx + cx^2 \in \mathcal{K}\right\} \\
&= \inf\left\{\int_{-1}^{1} |x^3 - a - bx - cx^2|^2 dx \mid a, b, c \in \mathbb{C}\right\}.
\end{aligned}$$

By uniqueness of $\mathrm{P}_{\mathcal{K}}x^3$, (36) is minimized in $a, b, c \in \mathbb{C}$ if and only if $a = 0$, $b = 3/5$, and $c = 0$. □

17.6 $\ell^2(\mathbb{N})$ is a Hilbert Space

Proposition 17.5. *$\ell^2(\mathbb{N})$ is a Hilbert space.*

Proof. Let $(a^n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\ell^2(\mathbb{N})$.

Step 1: We show that for each $k \in \mathbb{N}$, the sequence of kth coordinates $(a_k^n)_{n\in\mathbb{N}}$ is a Cauchy sequence of complex numbers, and hence must converge (as $\mathbb{C}$ is complete). Let $k \in \mathbb{N}$ and let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $m, n \geq N$ implies $\|a^n - a^m\| < \varepsilon^2$. Then $n, m \geq N$ implies

$$\begin{aligned}
|a_k^n - a_k^m|^2 &\leq \sum_{i=1}^{\infty} |a_i^n - a_i^m|^2 \\
&= \|a^n - a^m\| \\
&< \varepsilon^2,
\end{aligned}$$

which implies $|a_k^n - a_k^m| < \varepsilon$. Therefore $(a_k^n)_{n\in\mathbb{N}}$ is a Cauchy sequence of complex numbers. In particular, the sequence $(a_k^n)_{n\in\mathbb{N}}$ converges to some element, say $a_k^n \to a_k$.

Step 2: We show that the sequence $(a_k)_{k\in\mathbb{N}}$ defined in step 1 is square summable. Since (a^n) is a Cauchy sequence of elements in $\ell^2(\mathbb{N})$, there exists an $M > 0$ such that $\|a^n\| < M$ for all $n \in \mathbb{N}$ (see Lemma ((24.1) for a proof of this). Choose such an $M > 0$. Let $\varepsilon > 0$ and let $K \in \mathbb{N}$. Choose $N \in \mathbb{N}$ such that

$$|a_k|^2 < |a_k^N|^2 + \varepsilon/K$$

for all $1 \leq k \leq K$. Then

$$\begin{aligned}
\sum_{k=1}^{K} |a_k|^2 &< \sum_{k=1}^{K} |a_k^N|^2 + \varepsilon \\
&\leq \|a^N\| + \varepsilon \\
&\leq M + \varepsilon.
\end{aligned}$$

Taking the limit $K \to \infty$, we see that

$$\begin{aligned}
\|a\| &= \sum_{k=1}^{\infty} |a_k|^2 \\
&\leq M + \varepsilon \\
&\leq 0.
\end{aligned}$$

In particular, a is square summable.

Step 3: Let a be the sequence $(a_k)_{k\in\mathbb{N}}$ defined in step 1. We show that $a^n \to a$ in the ℓ^2 norm. Let $\varepsilon > 0$ and let $K \in \mathbb{N}$. Choose $N \in \mathbb{N}$ such that $m, n \geq N$ implies $\|a^n - a^m\|^2 < \varepsilon/2$. Then

$$\begin{aligned}
\sum_{k=1}^{K} |a_k^n - a_k^m|^2 &\leq \sum_{k=1}^{\infty} |a_k^n - a_k^m|^2 \\
&= \|a^n - a^m\|^2 \\
&< \varepsilon/2
\end{aligned}$$

for all $n, m \geq N$. Since $a_k^m \to a_k$ as $m \to \infty$ implies

$$\sum_{k=1}^{K} |a_k^n - a_k^m|^2 \to \sum_{k=1}^{K} |a_k^n - a_k|^2$$

as $m \to \infty$, we see that after taking the limit $m \to \infty$, we have

$$\sum_{k=1}^{K} |a_k^n - a_k|^2 \leq \varepsilon/2. \tag{37}$$

for all $n \geq N$. Taking the limit $K \to \infty$ in (37) gives us

$$\|a^n - a\|^2 < \varepsilon$$

for all $n \geq N$. It follows that $a^n \to a$. □

17.7 $C[a,b]$ is not a Hilbert Space

Proposition 17.6. *$C[a,b]$ is not a Hilbert space.*

Proof. For each $n \in \mathbb{N}$, define $f_n \in C[a,b]$ by

$$f_n(x) = \begin{cases} 0 & x \in [a, c - \frac{1}{n}] \\ nx + 1 - nc & x \in [c - \frac{1}{n}, c] \\ 1 & x \in [c, b], \end{cases}$$

where $c = \frac{a+b}{2}$. We will show that the sequence (f_n) is a Cauchy sequence which is not convergent.

Step 1: We first show that the sequence (f_n) is a Cauchy sequence. Let $\varepsilon > 0$ and let $m, n \in \mathbb{N}$ such that $n \geq m$. Then

$$\begin{aligned}
\|f_n - f_m\|^2 &= \int_{c-\frac{1}{m}}^{c-\frac{1}{n}} |mx + 1 - mc|^2 \, dx + \int_{c-\frac{1}{n}}^{c} |nx + 1 - nc - (mx + 1 - mc)|^2 \, dx \\
&= \int_{c-\frac{1}{m}}^{c-\frac{1}{n}} |m(x-c) + 1|^2 \, dx + (n-m)^2 \int_{c-\frac{1}{n}}^{c} |x - c|^2 \, dx \\
&\leq \left(\frac{1}{m} - \frac{1}{n}\right) \left|1 - \frac{m}{n}\right|^2 + \frac{(n-m)^2}{n^3} \\
&\leq \frac{1}{m} - \frac{1}{n} + \frac{(n-m)^2}{n^3}.
\end{aligned}$$

Choose $N \in \mathbb{N}$ such that $n \geq m \geq N$ implies

$$\frac{1}{m} - \frac{1}{n} + \frac{(n-m)^2}{n^3} < \varepsilon^2.$$

Then $n \geq m \geq N$ implies $\|f_n - f_m\| < \varepsilon$. Therefore (f_n) is a Cauchy sequence.

Step 2: We show that the sequence (f_n) is not convergent. Assume for a contradiction that $f_n \to f$ where $f \in C[a,b]$. Then

$$\begin{aligned}\|f_n - f\|^2 &= \int_a^{c-\frac{1}{n}} |f(x)|^2 dx + \int_{c-\frac{1}{n}}^{c} |f(x) - f_n(x)|^2 dx + \int_c^b |f(x)-1|^2 dx \\ &\leq (c - a - \frac{1}{n}) \sup_{x \in [a, c-\frac{1}{n}]} |f(x)|^2 + \int_{c-\frac{1}{n}}^{c} |f(x) - f_n(x)|^2 dx + (b-c) \sup_{x \in [c, c-\frac{1}{n}]} |f(x)-1|^2 dx.\end{aligned}$$

Since $\|f_n - f\| \to 0$ as $n \to \infty$, we see that (after taking the limit $n \to \infty$) we must have

$$f(x) = \begin{cases} 0 & \text{if } x \in [a,c] \\ 1 & \text{if } x \in [c,b] \end{cases}$$

but this is not a continuous function. Thus we obtain a contradiction. □

Appendix

Proposition 17.7. *Let $(V, \langle \cdot, \cdot \rangle)$ be an inner-product space and let W be a subspace of V. Then $(W, \langle \cdot, \cdot \rangle|_{W \times W}\rangle$ is an inner-product space, where $\langle \cdot, \cdot \rangle|_{W \times W} \colon W \times W \to \mathbb{C}$ is the restriction of $\langle \cdot, \cdot \rangle$ to $W \times W$.*

Proof. All of the required properties for $\langle \cdot, \cdot \rangle|_{W \times W}$ to be an inner-product are *inherited* by $\langle \cdot, \cdot \rangle$ since W is a subset of V. For instance, let $x, y, z \in V$ and let $\lambda \in \mathbb{C}$. Then

$$\begin{aligned}\langle x + \lambda y, z \rangle|_{W \times W} &= \langle x + \lambda y, z \rangle \\ &= \langle x, z \rangle + \lambda \langle y, z \rangle \\ &= \langle x, z \rangle|_{W \times W} + \lambda \langle y, z \rangle|_{W \times W}\end{aligned}$$

gives us linearity in the first argument. The other properties follow similarly. □

Remark 35. As long as context is clear, then we denote $\langle \cdot, \cdot \rangle|_{W \times W}$ simply by $\langle \cdot, \cdot \rangle$.

Proposition 17.8. *Let $(V, \langle \cdot, \cdot \rangle)$ be an n-dimensional inner-product space. Then $(V, \langle \cdot, \cdot \rangle)$ is unitarily equivalent to $(\mathbb{C}^n, \langle \cdot, \cdot \rangle_e\rangle$, where $\langle \cdot, \cdot \rangle_e$ is the standard Euclidean inner-product on $\mathbb{C}^n$. In particular, $(V, \langle \cdot, \cdot \rangle)$ is a separable Hilbert space.*

Proof. Let $\{v_1, \dots, v_n\}$ be a basis for V. By applying the Gram-Schmidt process to $\{v_1, \dots, v_n\}$, we can get an orthonormal basis, say $\{u_1, \dots, u_n\}$, of V. Let $\varphi \colon V \to \mathbb{C}^n$ be the unique linear isomorphism such that

$$\varphi(u_i) = e_i$$

where e_i is the standard ith coordinate vector in $\mathbb{C}^n$ for all $1 \leq i \leq n$. Then φ is a unitary equivalence. Indeed, it is an isomorphism since it restricts to a bijection on basis sets. Moreover we have

$$\langle u_i, u_j \rangle = \langle \varphi(u_i), \varphi(u_j) \rangle_{\mathrm{e}} = \langle e_i, e_j \rangle_{\mathrm{e}}$$

for all $1 \leq i, j \leq n$. This implies

$$\langle x, y \rangle = \langle \varphi(x), \varphi(y) \rangle_{\mathrm{e}}$$

for all $x, y \in V$. □

Proposition 17.9. *Let $\mathcal{V}$ be an inner-product space over $\mathbb{C}$ and let $\mathcal{W}$ be a finite dimensional subspace of $\mathcal{V}$. Then $\mathcal{W}$ is a closed.*

Proof. Let $\{w_1, \dots, w_k\}$ be an orthonormal basis for $\mathcal{W}$ and let (x_n) be a sequence of vectors in $\mathcal{W}$ such that $x_n \to x$ where $x \in \mathcal{V}$. For each $n \in \mathbb{N}$, express x_n in terms of the basis $\{w_1, \dots, w_k\}$ say as

$$x_n = \lambda_{1n} w_1 + \cdots + \lambda_{kn} w_k,$$

where $\lambda_{1n}, \dots, \lambda_{kn} \in \mathbb{C}$. Since $x_n \to x$ as $n \to \infty$, the sequence (x_n) is a Cauchy sequence. This implies the sequence $(\lambda_{jn})_{n \in \mathbb{N}}$ of complex numbers is a Cauchy sequence, for each $1 \leq j \leq k$. Indeed, letting $\varepsilon > 0$, choose $N \in \mathbb{N}$ such that $n, m \geq N$ implies $\|x_n - x_m\| < \varepsilon$. Then $n, m \geq N$ implies

$$\begin{aligned} |\lambda_{jn} - \lambda_{jm}| &\leq |\lambda_{1n} - \lambda_{1m}| + \cdots + |\lambda_{kn} - \lambda_{km}| \\ &= \|(\lambda_{1n} - \lambda_{1m}) w_1 + \cdots + (\lambda_{kn} - \lambda_{km}) w_k\| \\ &= \|x_n - x_m\| \\ &< \varepsilon \end{aligned}$$

for each $1 \leq j \leq k$. Now since $\mathbb{C}$ is complete, we must have $\lambda_{jn} \to \lambda_j$ as $n \to \infty$ for some $\lambda_j \in \mathbb{C}$ for all $1 \leq j \leq n$. In particular, we have

$$\begin{aligned} x &= \lim_{n \to \infty} x_n \\ &= \lim_{n \to \infty} (\lambda_{1n} w_1 + \cdots + \lambda_{kn} w_k) \\ &= \lim_{n \to \infty} (\lambda_{1n} w_1) + \cdots + \lim_{n \to \infty} (\lambda_{kn} w_k) \\ &= \lambda_1 w_1 + \cdots + \lambda_k w_k, \end{aligned}$$

and this implies $x \in \mathcal{W}$, which implies $\mathcal{W}$ is closed. □

Lemma 17.1. *Let (x_n) be a Cauchy sequence in $\mathcal{V}$. Then (x_n) is bounded.*

Proof. Let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $m, n \geq N$ implies $\|x_n - x_m\| < \varepsilon$. Thus, fixing $m \in \mathbb{N}$, we see that $n \geq N$ implies

$$\|x_n\| < \|x_m\| + \varepsilon.$$

Now we let

$$M = \max\{\|x_1\|, \|x_2\|, \dots, \|x_m\| + \varepsilon\}.$$

Then M is a bound for (x_n). □

Proposition 17.10. *Let (x_n) and (y_n) be Cauchy sequences of vectors in $\mathcal{V}$. Then $(\langle x_n, y_n \rangle)$ is a Cauchy sequence of complex numbers.*

Proof. Let $\varepsilon > 0$. Choose M_x and M_y such that $\|x_n\| < M_x$ and $\|y_n\| < M_y$ for all $n \in \mathbb{N}$. We can do this by Lemma (24.1). Next, choose $N \in \mathbb{N}$ such that $n, m \geq N$ implies $\|x_n - x_m\| < \frac{\varepsilon}{2M_y}$ and $\|y_n - y_m\| < \frac{\varepsilon}{2M_x}$. Then $n, m \geq M$ implies

$$\begin{aligned} |\langle x_n, y_n \rangle - \langle x_m, y_m \rangle| &= |\langle x_n, y_n \rangle - \langle x_m, y_n \rangle + \langle x_m, y_n \rangle - \langle x_m, y_m \rangle| \\ &= |\langle x_n - x_m, y_n \rangle + \langle x_m, y_n - y_m \rangle| \\ &\leq |\langle x_n - x_m, y_n \rangle| + |\langle x_m, y_n - y_m \rangle| \\ &\leq \|x_n - x_m\| \|y_n\| + \|x_m\| \|y_n - y_m\| \\ &\leq \|x_n - x_m\| M_y + M_x \|y_n - y_m\| \\ &< \varepsilon. \end{aligned}$$

This implies $(\langle x_n, y_n \rangle)$ is a Cauchy sequence of complex numbers in $\mathbb{C}$. The latter statement in the proposition follows from the fact that $\mathbb{C}$ is complete. □

Homework 2, Problem 5

Proposition 17.11. *Let $\mathcal{V}$ be an inner-product space, let $\mathcal{A}$ be a subspace of $\mathcal{V}$, let $x \in \mathcal{V}$, and let $\lambda \in \mathbb{C}$. Then*

$$d(\lambda x, \mathcal{A}) = |\lambda| d(x, \mathcal{A}).$$

Proof. Choose a sequence (y_n) of elements in $\mathcal{A}$ such that

$$\|x - y_n\| < d(x, \mathcal{A}) + \frac{1}{|\lambda n|}$$

for all $n \in \mathbb{N}$. Then since $\mathcal{A}$ is a subspace, we have

$$\begin{aligned} d(\lambda x, \mathcal{A}) &\leq \|\lambda x - \lambda y_n\| \\ &= |\lambda| \|x - y_n\| \\ &< |\lambda| \left(d(x, \mathcal{A}) + \frac{1}{|\lambda n|} \right) \\ &= |\lambda| d(x, \mathcal{A}) + \frac{1}{n} \end{aligned}$$

for all $n \in \mathbb{N}$. In particular, this implies $d(\lambda x, \mathcal{A}) \leq |\lambda| d(x, \mathcal{A})$.

Conversely, choose a sequence (z_n) of elements in $\mathcal{A}$ such that

$$\|\lambda x - z_n\| < d(\lambda x, \mathcal{A}) + \frac{1}{n}$$

Then since $\mathcal{A}$ is a subspace, we have

$$\begin{aligned} |\lambda| d(x, \mathcal{A}) &\leq |\lambda| \|x - z_n / |\lambda|\| \\ &= \|\lambda x - z_n\| \\ &< d(\lambda x, \mathcal{A}) + \frac{1}{n} \end{aligned}$$

for all $n \in \mathbb{N}$. In particular, this implies $|\lambda| d(x, \mathcal{A}) \leq d(\lambda x, \mathcal{A})$. □

Proposition 17.12. *Let $\mathcal{V}$ be an inner-product space, let $\mathcal{A}$ be a subspace of $\mathcal{V}$, and let $x, y \in \mathcal{V}$. Then*

$$d(x + y, \mathcal{A}) \leq d(x, \mathcal{A}) + d(y, \mathcal{A}).$$

Proof. Choose a sequences (w_n) and (z_n) of elements in $\mathcal{A}$ such that

$$\|x - w_n\| < d(x, \mathcal{A}) + \frac{1}{2n} \quad \text{and} \quad \|y - z_n\| < d(y, \mathcal{A}) + \frac{1}{2n}$$

for all $n \in \mathbb{N}$. Then since $\mathcal{A}$ is a subspace, we have

$$\begin{aligned} d(x + y, \mathcal{A}) &\leq \|(x + y) - (w_n + z_n)\| \\ &\leq \|x - w_n\| + \|y - z_n\| \\ &< d(x, \mathcal{A}) + \frac{1}{2n} + d(y, \mathcal{A}) + \frac{1}{2n} \\ &= d(x, \mathcal{A}) + d(y, \mathcal{A}) + \frac{1}{n} \end{aligned}$$

for all $n \in \mathbb{N}$. In particular, this implies $d(x + y, \mathcal{A}) \leq d(x, \mathcal{A}) + d(y, \mathcal{A})$. □

18 Homework 5

Throughout this homework, let $\mathcal{H}$ be a Hilbert space.

18.1 Map Sending Bounded Operator to its Adjoint Operator is Conjugate-Linear

Proposition 18.1. *Let $T\colon \mathcal{H} \to \mathcal{H}$ and $S\colon \mathcal{H} \to \mathcal{H}$ be two bounded operators. Then*

$$(\alpha T + \beta S)^* = \overline{\alpha} T^* + \overline{\beta} S^* \tag{38}$$

for all $\alpha, \beta, \in \mathbb{C}$.

Proof. Let $\alpha, \beta, \in \mathbb{C}$ and let $y \in \mathcal{H}$. Then for all $x \in \mathcal{H}$, we have

$$\begin{aligned}
\langle x, (\alpha T + \beta S)^* y \rangle &= \langle (\alpha T + \beta S) x, y \rangle \\
&= \alpha \langle Tx, y \rangle + \beta \langle Sx, y \rangle \\
&= \alpha \langle x, T^* y \rangle + \beta \langle x, S^* y \rangle \\
&= \langle x, (\overline{\alpha} T^* + \overline{\beta} S^*) y \rangle
\end{aligned}$$

In particular, this implies $(\alpha T + \beta S)^* y = (\overline{\alpha} T^* + \overline{\beta} S^*) y$ for all $y \in \mathcal{H}$ (by positive-definiteness of the inner-product) which implies (38). □

18.2 Composite of Bounded Operators is Bounded

Proposition 18.2. *Let $T\colon \mathcal{H} \to \mathcal{H}$ and $S\colon \mathcal{H} \to \mathcal{H}$ be two bounded operators. Then*

1. *TS is bounded and $\|TS\| \leq \|T\| \|S\|$;*
2. *$(TS)^* = S^* T^*$.*

Proof.

1. Let $x \in \mathcal{H}$ such that $\|x\| = 1$. Then

$$\begin{aligned}
\|TSx\| &\leq \|T\| \|Sx\| \\
&\leq \|T\| \|S\| \|x\| \\
&= \|T\| \|S\|.
\end{aligned}$$

Thus TS is bounded and $\|TS\| \leq \|T\| \|S\|$.

2. Let $y \in \mathcal{H}$. Then for all $x \in \mathcal{H}$, we have

$$\begin{aligned}
\langle x, (TS)^* y \rangle &= \langle TSx, y \rangle \\
&= \langle Sx, T^* y \rangle \\
&= \langle x, S^* T^* y \rangle.
\end{aligned}$$

In particular, this implies $(TS)^* y = S^* T^* y$ for all $y \in \mathcal{H}$, which implies $(TS)^* = S^* T^*$. □

18.3 The Adjoint Operator of $T_{u,v}$ is $T_{v,u}$

Proposition 18.3. *Let $u, v \in \mathcal{H}$ be fixed vectors.*

1. *The operator $T_{u,v}\colon \mathcal{H} \to \mathcal{H}$ defined by*
$$T_{u,v} x = \langle x, u \rangle v$$
for all $x \in \mathcal{H}$ is bounded. Moreover, we have $\|T_{u,v}\| = \|u\| \|v\|$.
2. *The adjoint of $T_{u,v}$ is given by $T_{v,u}$, that is,*
$$(T_{u,v})^* y = \langle y, v \rangle u$$
for all $y \in \mathcal{H}$.

Proof.

1. Let $x \in \mathcal{H}$. Then

$$\begin{aligned} \|T_{u,v}x\| &= \|\langle x, u\rangle v\| \\ &= |\langle x, u\rangle| \|v\| \\ &\leq \|x\|\|u\|\|v\|, \end{aligned}$$

where we used Cauchy-Schwarz to get from the second to the third line. This implies $\|T_{u,v}\| \leq \|u\|\|v\|$. We have equality at the Cauchy-Schwarz step if and only if $x = \lambda u$ for some $\lambda \in \mathbb{C}$. In particular, setting $x = u/\|u\|$ gives us $\|T_{u,v}\| = \|u\|\|v\|$.

2. Let $y \in \mathcal{H}$. Then

$$\begin{aligned} \langle x, (T_{u,v})^* y\rangle &= \langle T_{u,v}x, y\rangle \\ &= \langle \langle x, u\rangle v, y\rangle \\ &= \langle x, u\rangle \langle v, y\rangle \\ &= \langle x, \overline{\langle v, y\rangle} u\rangle \\ &= \langle x, \langle y, v\rangle u\rangle \end{aligned}$$

for all $x \in \mathcal{H}$. This implies $(T_{u,v})^* y = \langle y, v\rangle u$ for all $y \in \mathcal{H}$. □

18.4 Computing Adjoint of Operator from $\ell^2(\mathbb{N})$ to Itself

Corollary 10. *Let $T \colon \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$ be operator defined by*

$$T(x)_n = \sum_{m=1}^{\infty} \frac{x_m}{2^n 3^m},$$

for all $x = (x_m) \in \ell^2(\mathbb{N})$, where $T(x)_n$ denotes the n-th coordinate of $T(x) \in \ell^2(\mathbb{N})$. Then T is bounded with

$$\|T\| = \sqrt{\frac{1}{24}}.$$

The adjoint of T is given by

$$T^*(y)_n = \sum_{m=1}^{\infty} \frac{y_m}{2^m 3^n},$$

for all $y \in \ell^2(\mathbb{N})$.

Proof. Set $u = (1/3^m)$ and $v = (1/2^n)$. Then

$$\begin{aligned} T(x)_n &= \sum_{m=1}^{\infty} \frac{x_m}{2^n 3^m} \\ &= \langle x, u\rangle \frac{1}{2^n} \\ &= \langle x, u\rangle v_n \end{aligned}$$

for all $x \in \mathcal{H}$. Thus $Tx = \langle x, u\rangle v$ for all $x \in \mathcal{H}$. Therefore we can apply Proposition (18.3) and obtain

$$\begin{aligned} \|T\| &= \|u\|\|v\| \\ &= \sqrt{\sum_{n=1}^{\infty} 9^{-n}} \sqrt{\sum_{n=1}^{\infty} 4^{-n}} \\ &= \sqrt{\left(\frac{1}{1-\frac{1}{9}} - 1\right)\left(\frac{1}{1-\frac{1}{4}} - 1\right)} \\ &= \sqrt{\frac{1}{24}}. \end{aligned}$$

The adjoint of T is given by

$$\begin{aligned} T^*(y)_n &= \langle y, v \rangle u_n \\ &= \sum_{m=1}^{\infty} \frac{y_m}{2^m 3^n} \end{aligned}$$

for all $y \in \mathcal{H}$. □

18.5 Properties of T^*T

Proposition 18.4. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a bounded operator. Then*

1. $\|T^*T\| = \|T\|^2$;
2. *Ker*$(T^*T) =$ *Ker*(T).

Proof.

1. First note that Proposition (18.2) implies $\|T^*T\| \leq \|T^*\|\|T\| = \|T\|^2$. For the reverse inequality, let $x \in \mathcal{H}$ such that $\|x\| = 1$. Then

$$\begin{aligned} \|Tx\|^2 &= \langle Tx, Tx \rangle \\ &= \langle x, T^*Tx \rangle \\ &\leq \|x\|\|T^*Tx\| \\ &= \|T^*Tx\|, \end{aligned}$$

where we used Cauchy-Schwarz to get from the second line to the third line. In particular, this implies

$$\begin{aligned} \|T\|^2 &= \sup\{\|Tx\|^2 \mid \|x\| \leq 1\} \\ &\leq \sup\{\|T^*Tx\| \mid \|x\| \leq 1\} \\ &= \|T^*T\|, \end{aligned}$$

where the first line is justifed in the Appendix.

2. Let $x \in \mathrm{Ker}(T)$. Then

$$\begin{aligned} T^*Tx &= T^*(Tx) \\ &= T^*(0) \\ &= 0 \end{aligned}$$

implies $x \in \mathrm{Ker}(T^*T)$. Thus $\mathrm{Ker}(T) \subseteq \mathrm{Ker}(T^*T)$.
For the reverse inclusion, let $x \in \mathrm{Ker}(T^*T)$. Then

$$\begin{aligned} \langle Tx, Tx \rangle &= \langle x, T^*Tx \rangle \\ &= \langle x, 0 \rangle \\ &= 0 \end{aligned}$$

implies $Tx = 0$ (by positive-definiteness of inner-product) which implies $x \in \mathrm{Ker}(T)$. Therefore $\mathrm{Ker}(T) \supseteq \mathrm{Ker}(T^*T)$. □

18.6 $\ker(T^*) = (\operatorname{im} T)^\perp$ and $(\ker T)^\perp = \overline{\operatorname{im}(T^*)}$

Proposition 18.5. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a bounded operator. Then*

1. $\ker(T^*) = (\operatorname{im} T)^\perp$;
2. $(\ker T)^\perp = \overline{\operatorname{im}(T^*)}$.

Proof.

1. Let $x \in \ker(T^*)$. Then

$$\begin{aligned}\langle Ty, x\rangle &= \langle y, T^*x\rangle \\ &= \langle y, 0\rangle \\ &= 0\end{aligned}$$

for all $Ty \in \operatorname{im} T$. This implies $x \in (\operatorname{im} T)^\perp$ and so $\ker(T^*) \subseteq (\operatorname{im} T)^\perp$. For the reverse inclusion, let $x \in (\operatorname{im} T)^\perp$. Then

$$\begin{aligned}0 &= \langle x, TT^*x\rangle \\ &= \langle T^*x, T^*x\rangle\end{aligned}$$

implies $T^*x = 0$ (by positive-definiteness of inner-product) which implies $x \in \ker(T^*)$ and so $\ker(T^*) \supseteq (\operatorname{im} T)^\perp$.

2. Let $T^*y \in \operatorname{im}(T^*)$. Then for all $x \in \ker T$, we have

$$\begin{aligned}\langle x, T^*y\rangle &= \langle Tx, y\rangle \\ &= \langle 0, y\rangle \\ &= 0.\end{aligned}$$

In particular, this implies $\operatorname{im}(T^*) \subseteq (\ker T)^\perp$ which implies $\overline{\operatorname{im}(T^*)} \subseteq (\ker T)^\perp$ (as $(\ker T)^\perp$ is a closed subspace which contains $\operatorname{im}(T^*)$). For the reverse inclusion, we have

$$\begin{aligned}(\ker T)^\perp &= \ker((T^*)^*)^\perp \\ &= (\operatorname{im}(T^*)^\perp)^\perp \\ &\subseteq ((\overline{\operatorname{im}(T^*)})^\perp)^\perp \\ &= \overline{\operatorname{im}(T^*)},\end{aligned}$$

where we used part 1 of this proposition to get from the first line to the second line. □

18.7 T is an Isometry if and only if $T^*T = 1_{\mathcal{H}}$

Definition 18.1. An **isometry** between normed vector spaces $\mathcal{V}_1$ and $\mathcal{V}_2$ is an operator $T\colon \mathcal{V}_1 \to \mathcal{V}_2$ such that

$$\|Tx - Ty\| = \|x - y\|$$

for all $x, y \in \mathcal{V}$.

Proposition 18.6. *Let $\mathcal{V}_1$ and $\mathcal{V}_2$ be inner-product spaces and let $T\colon \mathcal{V}_1 \to \mathcal{V}_2$ be an operator. Then T is an isometry (where $\mathcal{V}_1$ and $\mathcal{V}_2$ are viewed as the induced normed vector spaces with respect to their inner-products) if and only if*

$$\langle x, y\rangle = \langle Tx, Ty\rangle \tag{39}$$

for all $x, y \in \mathcal{V}_1$.

Proof. Suppose (39) holds for all $x, y \in \mathcal{V}_1$. Then

$$\begin{aligned}\|Tx - Ty\| &= \sqrt{\langle Tx - Ty, Tx - Ty\rangle} \\ &= \sqrt{\langle Tx, Tx\rangle - \langle Tx, Ty\rangle - \langle Ty, Tx\rangle + \langle Ty, Ty\rangle} \\ &= \sqrt{\langle x, x\rangle - \langle x, y\rangle - \langle y, x\rangle + \langle y, y\rangle} \\ &= \sqrt{\langle x - y, x - y\rangle} \\ &= \|x - y\|.\end{aligned}$$

for all $x, y \in \mathcal{V}_1$. Thus T is an isometry.

Conversely, suppose T is an isometry and let $x, y \in \mathcal{V}_1$. Then

$$\begin{aligned}
\|x\|^2 - 2\mathrm{Re}(\langle x, y\rangle) + \|y\|^2 &= \langle x - y, x - y\rangle \\
&= \langle Tx - Ty, Tx - Ty\rangle \\
&= \|Tx\|^2 - 2\mathrm{Re}(\langle Tx, Ty\rangle) + \|Ty\|^2 \\
&= \|x\|^2 - 2\mathrm{Re}(\langle Tx, Ty\rangle) + \|y\|^2
\end{aligned}$$

implies $\mathrm{Re}(\langle x, y\rangle) = \mathrm{Re}(\langle Tx, Ty\rangle)$ for all $x, y \in \mathcal{V}_1$. Note that this also implies

$$\begin{aligned}
\mathrm{Im}(\langle x, y\rangle) &= -\mathrm{Re}(i\langle x, y\rangle) \\
&= -\mathrm{Re}(\langle ix, y\rangle) \\
&= -\mathrm{Re}(\langle T(ix), Ty\rangle) \\
&= -\mathrm{Re}(i\langle Tx, Ty\rangle) \\
&= \mathrm{Im}(\langle Tx, Ty\rangle)
\end{aligned}$$

for all $x, y \in \mathcal{V}_1$. Thus we have (39) for all $x, y \in \mathcal{V}_1$. □

Proposition 18.7. *Let $T \colon \mathcal{H} \to \mathcal{H}$ be a bounded operator. Then*

1. *T is an isometry if and only if $T^*T = 1_{\mathcal{H}}$.*
2. *There exists isometries T such that $TT^* \neq 1_{\mathcal{H}}$.*

Proof.

1. Suppose T is an isometry. Then for all $y \in \mathcal{H}$, we have

$$\begin{aligned}
\langle x, 1_{\mathcal{H}} y\rangle &= \langle x, y\rangle \\
&= \langle Tx, Ty\rangle \\
&= \langle x, T^*Ty\rangle
\end{aligned}$$

for all $x \in \mathcal{H}$. In particular, this implies $T^*Ty = 1_{\mathcal{H}} y$ for all $y \in \mathcal{H}$, which implies $T^*T = 1_{\mathcal{H}}$.

Conversely, suppose $T^*T = 1_{\mathcal{H}}$. Then

$$\begin{aligned}
\langle Tx, Ty\rangle &= \langle x, T^*Ty\rangle \\
&= \langle x, 1_{\mathcal{H}} y\rangle \\
&= \langle x, y\rangle
\end{aligned}$$

for all $x, y \in \mathcal{H}$. This implies T is an isometry.

2. Consider the shift operator $S \colon \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$, given by

$$S(x_n) = (x_{n-1})$$

for all $(x_n) \in \ell^2(\mathbb{N})$, where $x_0 = 0$. In class, it was shown that

$$S^*(x_n) = (x_{n+1})$$

for all $(x_n) \in \ell^2(\mathbb{N})$. Thus, whenever $x_1 \neq 0$, we have

$$\begin{aligned}
SS^*(x_n) &= SS^*(x_1, x_2, \dots) \\
&= S(x_2, x_3, \dots) \\
&= (0, x_2, x_3, \dots) \\
&\neq (x_n).
\end{aligned}$$

On the other hand, S is an isometry. Indeed, let $(x_n), (y_n) \in \ell^2(\mathbb{N})$. Then

$$\begin{aligned}
\langle S(x_n), S(y_n) \rangle &= \langle (x_{n-1}), (y_{n-1}) \rangle \\
&= \sum_{n=1}^{\infty} x_{n-1} \overline{y}_{n-1} \\
&= \sum_{m=0}^{\infty} x_m \overline{y}_m \\
&= x_0 y_0 + \sum_{m=1}^{\infty} x_m \overline{y}_m \\
&= \sum_{m=1}^{\infty} x_m \overline{y}_m \\
&= \langle (x_n), (y_n) \rangle.
\end{aligned}$$

□

Appendix

Proposition 18.8. *Let $T \colon \mathcal{U} \to \mathcal{V}$ be a bounded linear operator. Then*

$$\|T\|^2 = \sup\{\|Tx\|^2 \mid \|x\| \leq 1\}$$

Proof. For any $x \in \mathcal{U}$ such that $\|x\| \leq 1$, we have $\|Tx\|^2 \leq \|T\|^2$. Thus

$$\|T\|^2 \geq \sup\{\|Tx\|^2 \mid \|x\| \leq 1\}. \tag{40}$$

To show the reverse inequality, we assume (for a contradiction) that (40) is a strictly inequality. Choose $\delta > 0$ such that

$$\|T\|^2 - \delta > \sup\{\|Tx\|^2 \mid \|x\| \leq 1\}.$$

Now let $\varepsilon = \delta/2\|T\|$, and choose $x \in \mathcal{U}$ such that $\|x\| \leq 1$ and such that

$$\|T\| - \varepsilon < \|Tx\|.$$

Then

$$\begin{aligned}
\|Tx\|^2 &> (\|T\| - \varepsilon)^2 \\
&= \|T\|^2 - 2\varepsilon\|T\| + \varepsilon^2 \\
&\geq \|T\|^2 - 2\varepsilon\|T\| \\
&= \|T\|^2 - \delta
\end{aligned}$$

gives us a contradiction. □

19 Homework 6

Throughout this homework, let $\mathcal{H}$ be a Hilbert space.

19.1 Decomposition of T as $T = A + iB$ where A and B are Self-Adjoint

Proposition 19.1. *Let $T \colon \mathcal{H} \to \mathcal{H}$ be a bounded operator. There exists unique self-adjoint operators $A \colon \mathcal{H} \to \mathcal{H}$ and $B \colon \mathcal{H} \to \mathcal{H}$ such that $T = A + iB$.*

Proof. Define

$$A = \frac{1}{2}(T + T^*) \quad \text{and} \quad B = \frac{-i}{2}(T - T^*).$$

Then

$$\begin{aligned}A + iB &= \frac{1}{2}(T + T^*) + \frac{1}{2}(T - T^*)\\ &= \left(\frac{1}{2} + \frac{1}{2}\right) T + \left(\frac{1}{2} - \frac{1}{2}\right) T^*\\ &= T\end{aligned}$$

Furthermore, A and B are self-adjoint. Indeed,

$$\begin{aligned}A^* &= \left(\frac{1}{2}(T + T^*)\right)^*\\ &= \frac{1}{2}(T^* + T^{**})\\ &= \frac{1}{2}(T^* + T)\\ &= \frac{1}{2}(T + T^*)\\ &= A,\end{aligned}$$

and similarly

$$\begin{aligned}B^* &= \left(\frac{-i}{2}(T - T^*)\right)^*\\ &= \frac{i}{2}(T^* - T^{**})\\ &= \frac{i}{2}(T^* - T)\\ &= \frac{-i}{2}(T - T^*)\\ &= B.\end{aligned}$$

This establishes existence.

For uniqueness, suppose that $A' \colon \mathcal{H} \to \mathcal{H}$ and $B' \colon \mathcal{H} \to \mathcal{H}$ are two other self-adjoint operators such that $T = A' + iB'$. Then since

$$\begin{aligned}T^* &= (A + iB)^*\\ &= A^* - iB^*\\ &= A - iB,\end{aligned}$$

and since

$$\begin{aligned}T^* &= (A' + iB')^*\\ &= A'^* - iB'^*\\ &= A' - iB',\end{aligned}$$

we have

$$\begin{aligned}A + iB &= A' + iB'\\ A - iB &= A' - iB'.\end{aligned}$$

Adding these together gives us $2A = 2A'$, and hence $A = A'$. Similarly, subtracting these gives us $2iB = 2iB'$, and hence $B = B'$. $\square$

19.2 S^*S and Orthogonal Projection are Positive Operators

Definition 19.1. A self-adjoint operator $T \colon \mathcal{H} \to \mathcal{H}$ is said to be **positive** if $\langle Tx, x\rangle \geq 0$ for all $x \in \mathcal{H}$. We say T is **strictly positive** if $\langle Tx, x\rangle > 0$ for all $x \in \mathcal{H}\backslash\{0\}$.

Remark 36. Equivalently, $T: \mathcal{H} \to \mathcal{H}$ is positive if and only if $\langle x, Tx \rangle \geq 0$ for all $x \in \mathcal{H}$. Indeed, if $\langle Tx, x \rangle \geq 0$ for all $x \in \mathcal{H}$, then $\langle Tx, x \rangle$ is real for all $x \in \mathcal{H}$, and so

$$\begin{aligned} \langle x, Tx \rangle &= \overline{\langle Tx, x \rangle} \\ &= \langle Tx, x \rangle \\ &\geq 0. \end{aligned}$$

Similarly, $\langle x, Tx \rangle \geq 0$ for all $x \in \mathcal{H}$ implies $\langle Tx, x \rangle \geq 0$ for all $x \in \mathcal{H}$.

Problem 2.a

Proposition 19.2. *Let $S: \mathcal{H} \to \mathcal{H}$ be a bounded operator. Then S^*S is positive.*

Proof. Let $x \in \mathcal{H}$. Then

$$\begin{aligned} \langle S^*Sx, x \rangle &= \langle Sx, Sx \rangle \\ &\geq 0 \end{aligned}$$

by positive-definiteness of the inner-product. It follows that S^*S is positive. □

Remark 37. I think we do not need S to be bounded here, but we only defined the adjoint of a bounded operator in class.

Problem 2.b

Proposition 19.3. *Let $\mathcal{K}$ be a closed subspace of $\mathcal{H}$. Then the orthogonal projection $P_{\mathcal{K}}: \mathcal{H} \to \mathcal{H}$ is positive.*

Proof. Let $x \in \mathcal{H}$. Then

$$\begin{aligned} 0 &= \langle x - P_{\mathcal{K}}x, P_{\mathcal{K}}x \rangle \\ &= \langle x, P_{\mathcal{K}}x \rangle - \langle P_{\mathcal{K}}x, P_{\mathcal{K}}x \rangle \\ &= \langle x, P_{\mathcal{K}}x \rangle - \|P_{\mathcal{K}}x\|^2. \end{aligned}$$

It follows that $\langle x, P_{\mathcal{K}}x \rangle = \|P_{\mathcal{K}}x\|^2 \geq 0$ which implies $P_{\mathcal{K}}$ is positive by Remark (36). □

19.3 Another Version of Polarization Identity and $\langle Tx, x \rangle = 0$ for all $x \in \mathcal{H}$ implies $T = 0$

Problem 3.a

Proposition 19.4. *(Another Version of Polarization Identity) Let $T: \mathcal{H} \to \mathcal{H}$ be any operator. Then*

$$4\langle Tx, y \rangle = \sum_{i=0}^{3} \langle T(x + i^k y), x + i^k y \rangle \tag{41}$$

Proof. We have

$$\begin{aligned} \langle T(x+y), x+y \rangle &= \langle Tx + Ty, x + y \rangle \\ &= \langle Tx, x \rangle + \langle Tx, y \rangle + \langle Ty, x \rangle + \langle Ty, y \rangle \end{aligned}$$

and

$$\begin{aligned} i\langle T(x+iy), x+iy \rangle &= i\langle Tx + iTy, x + iy \rangle \\ &= i\langle Tx, x \rangle + i\langle Tx, iy \rangle + i\langle iTy, x \rangle + i\langle iTy, iy \rangle \\ &= i\langle Tx, x \rangle + \langle Tx, y \rangle - \langle Ty, x \rangle + i\langle Ty, y \rangle \end{aligned}$$

and

$$\begin{aligned} -\langle T(x-y), x-y \rangle &= -\langle Tx - Ty, x - y \rangle \\ &= -\langle Tx, x \rangle - \langle Tx, -y \rangle - \langle -Ty, x \rangle - \langle -Ty, -y \rangle \\ &= -\langle Tx, x \rangle + \langle Tx, y \rangle + \langle Ty, x \rangle - \langle Ty, y \rangle \end{aligned}$$

and

$$\begin{aligned}-i\langle T(x-iy),x-iy\rangle &= -i\langle Tx-iTy,x-iy\rangle \\ &= -i\langle Tx,x\rangle - i\langle Tx,-iy\rangle - i\langle -iTy,x\rangle - i\langle -iTy,-iy\rangle \\ &= -i\langle Tx,x\rangle + \langle Tx,y\rangle - \langle Ty,x\rangle - i\langle Ty,y\rangle.\end{aligned}$$

Adding these together gives us our desired result. □

Problem 3.b

Proposition 19.5. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be any operator such that $\langle Tx,x\rangle = 0$ for all $x \in \mathcal{H}$. Then $T = 0$.*

Proof. Let $x \in \mathcal{H}$. Then it follows from the polarization identity proved above that

$$\begin{aligned}4\langle Tx,y\rangle &= \sum_{i=0}^{3}\langle T(x+i^k y),x+i^k y\rangle \\ &= \sum_{i=0}^{3} 0 \\ &= 0\end{aligned}$$

for all $y \in \mathcal{H}$. It follows that $\langle Tx,y\rangle = 0$ for all $y \in \mathcal{H}$. This implies $Tx = 0$ by positive-definiteness of the inner-product. Since x was arbitrary, this implies $T = 0$. □

19.4 Twisting Inner-Product by Strictly Positive Self-Adjoint Operator

Problem 4.a

Proposition 19.6. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a strictly positive self-adjoint operator. Define a map $\langle \cdot,\cdot\rangle_T\colon \mathcal{H}\times\mathcal{H} \to \mathbb{C}$ by*

$$\langle x,y\rangle_T = \langle Tx,y\rangle$$

for all $x,y \in \mathcal{H}$. Then $\langle \cdot,\cdot\rangle_T$ is an inner-product.

Proof. We first check that $\langle \cdot,\cdot\rangle_T$ is linear in the first argument. Let $x,y,z \in \mathcal{H}$. Then

$$\begin{aligned}\langle x+z,y\rangle_T &= \langle T(x+z),y\rangle \\ &= \langle Tx+Tz,y\rangle \\ &= \langle Tx,y\rangle + \langle Tz,y\rangle \\ &= \langle x,y\rangle_T + \langle z,y\rangle_T.\end{aligned}$$

Next we check that $\langle \cdot,\cdot\rangle_T$ is conjugate-symmetric. Let $x,y \in \mathcal{H}$. Then since T is self-adjoint, we have

$$\begin{aligned}\langle x,y\rangle_T &= \langle Tx,y\rangle \\ &= \overline{\langle y,Tx\rangle} \\ &= \overline{\langle Ty,x\rangle} \\ &= \overline{\langle y,x\rangle}_T.\end{aligned}$$

Next we check that $\langle \cdot,\cdot\rangle_T$ is positive-definite. Let $x \in \mathcal{H}$. Then since T is strictly positive, we have

$$\begin{aligned}\langle x,x\rangle_T &= \langle Tx,x\rangle \\ &> 0,\end{aligned}$$

where $\langle x,x\rangle_T = 0$ if and only if $x = 0$. □

Problem 4.b

Proposition 19.7. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a strictly positive self-adjoint operator. Then*

$$|\langle Tx, y\rangle|^2 \leq \langle Tx, x\rangle\langle Ty, y\rangle \tag{42}$$

for all $x, y \in \mathcal{H}$.

Proof. We have

$$\begin{aligned}|\langle Tx, y\rangle|^2 &= |\langle x, y\rangle_T|^2 \\ &=\leq \|x\|_T^2\|y\|_T^2 \\ &= \langle x, x\rangle_T\langle y, y\rangle_T \\ &= \langle Tx, x\rangle\langle Ty, y\rangle,\end{aligned}$$

where we applied Cauchy-Schwarz for the $\langle\cdot,\cdot\rangle_T$ inner-product. □

Problem 4.c

Proposition 19.8. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a strictly positive self-adjoint operator. Then*

$$\|Tx\|^2 \leq \|T\|\langle Tx, x\rangle \tag{43}$$

for all $x \in \mathcal{H}$.

Proof. Let $x \in \mathcal{H}$. Then

$$\begin{aligned}\|Tx\|^4 &= \langle Tx, Tx\rangle^2 \\ &\leq \langle Tx, x\rangle\langle T^2x, Tx\rangle \\ &\leq \langle Tx, x\rangle\|T^2x\|\|Tx\| \\ &\leq \langle Tx, x\rangle\|T\|\|Tx\|\|Tx\| \\ &= \langle Tx, x\rangle\|T\|\|Tx\|^2,\end{aligned}$$

where we used (42) to get from the first line to the second line. Now dividing both sides by $\|Tx\|^2$[10], we obtain $\|Tx\|^2 \leq \langle Tx, x\rangle\|T\|$. □

19.5 T is Self-Adjoint if and only if $\langle Tx, x\rangle \in \mathbb{R}$ for all $x \in \mathcal{H}$

Proposition 19.9. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a bounded operator. Then T is self-adjoint if and only if $\langle Tx, x\rangle \in \mathbb{R}$ for all $x \in \mathcal{H}$.*

Proof. Suppose that T is self-adjoint. Let $x \in \mathcal{H}$. Then

$$\begin{aligned}\langle Tx, x\rangle &= \langle x, Tx\rangle \\ &= \overline{\langle Tx, x\rangle}\end{aligned}$$

implies $\langle Tx, x\rangle \in \mathbb{R}$.

Conversely, suppose that $\langle Tx, x\rangle \in \mathbb{R}$ for all $x \in \mathcal{H}$. Then

$$\begin{aligned}\langle (T - T^*)x, x\rangle &= \langle Tx - T^*x, x\rangle \\ &= \langle Tx, x\rangle - \langle T^*x, x\rangle \\ &= \overline{\langle x, Tx\rangle} - \langle x, Tx\rangle \\ &= \langle x, Tx\rangle - \langle x, Tx\rangle \\ &= 0\end{aligned}$$

for all $x \in \mathcal{H}$. Therefore by Proposition (19.5), we see that $T - T^* = 0$, i.e. $T = T^*$. □

[10] If $Tx = 0$, then we clearly have (43), thus we assume $Tx \neq 0$.

19.6 If T is Self-Adjoint Operator, then $\|T^n\| = \|T\|^n$

Problem 6.a

Proposition 19.10. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a self-adjoint operator. Then*

$$\|T^n\|^2 \leq \|T^{n+1}\|\|T^{n-1}\| \tag{44}$$

for all $n \in \mathbb{N}$.

Proof. Let $n \in \mathbb{N}$ and let $x \in \mathcal{H}$ such that $\|x\| \leq 1$. Then

$$\begin{aligned} \|T^n x\|^2 &= \langle T^n x, T^n x\rangle \\ &= \langle T^{n+1} x, T^{n-1} x\rangle \\ &\leq \|T^{n+1}x\|\|T^{n-1}x\| \\ &\leq \|T^{n+1}\|\|x\|\|T^{n-1}\|\|x\| \\ &\leq \|T^{n+1}\|\|T^{n-1}\|, \end{aligned}$$

which implies (44). □

Problem 6.b

Proposition 19.11. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a self-adjoint operator. Then*

$$\|T^n\| = \|T\|^n \tag{45}$$

for all $n \in \mathbb{N}$.

Proof. We prove (45) by induction on $n \geq 0$. The base case $n = 0$ and the case $n = 1$ are trivial. Assume that (45) holds for some $n \geq 1$. Then by (44), we have

$$\begin{aligned} \|T^{n+1}\| &\geq \|T^{n-1}\|^{-1}\|T^n\|^2 \\ &= \|T\|^{1-n}\|T\|^{2n} \\ &= \|T\|^{n+1}, \end{aligned}$$

where we used the induction step to get from the first line to the second line.

For the reverse inequality, let $x \in \mathcal{H}$ such that $\|x\| \leq 1$. Then

$$\begin{aligned} \|T^{n+1}x\| &\leq \|T^n x\|\|Tx\| \\ &\leq \|T^n\|\|x\|\|Tx\| \\ &\leq \|T^n\|\|Tx\| \\ &\leq \|T^n\|\|T\| \\ &= \|T\|^n\|T\| \\ &= \|T\|^{n+1}, \end{aligned}$$

where we used the induction step to get from the fourth line to the fifth line. It follows that $\|T^{n+1}\| \leq \|T\|^{n+1}$. □

20 Homework 7

Throughout this homework, let $\mathcal{H}$ be a Hilbert space.

20.1 Weak Convergence Properties

Problem 1.a

Proposition 20.1. *Let (x_n) and (y_n) be two sequences in $\mathcal{H}$ such that $x_n \xrightarrow{w} x$ and $y_n \xrightarrow{w} y$ and let $\alpha, \beta \in \mathbb{C}$. Then*

$$\alpha x_n + \beta y_n \xrightarrow{w} \alpha x + \beta y.$$

Proof. Let $z \in \mathcal{H}$. Then

$$\begin{aligned}\lim_{n\to\infty}(\langle \alpha x_n + \beta y_n, z\rangle) &= \lim_{n\to\infty}(\alpha\langle x_n, z\rangle + \beta\langle y_n, z\rangle)\\ &= \alpha \lim_{n\to\infty}(\langle x_n, z\rangle) + \beta \lim_{n\to\infty}(\langle y_n, z\rangle)\\ &= \alpha\langle x, z\rangle + \beta\langle y, z\rangle\\ &= \langle \alpha x + \beta y, z\rangle.\end{aligned}$$

Therefore $\alpha x_n + \beta y_n \xrightarrow{w} \alpha x + \beta y$. □

Problem 1.b

Proposition 20.2. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a bounded operator and let (x_n) be a sequence in $\mathcal{H}$ such that $x_n \xrightarrow{w} x$. Then*

$$Tx_n \xrightarrow{w} Tx.$$

Proof. Let $z \in \mathcal{H}$. Then

$$\begin{aligned}\lim_{n\to\infty}(\langle Tx_n, z\rangle) &= \lim_{n\to\infty}(\langle x_n, T^* z\rangle)\\ &= \langle x, T^* z\rangle\\ &= \langle Tx, z\rangle.\end{aligned}$$

Therefore $Tx_n \xrightarrow{w} Tx$. □

Remark 38. Note that we may not have $Tx_n \to Tx$. Indeed, suppose $\mathcal{H}$ is separable. Let (e_n) be an orthonormal sequence in $\mathcal{H}$ and let $T\colon \mathcal{H} \to \mathcal{H}$ be the identity map. Then $e_n \xrightarrow{w} 0$ but $Te_n = e_n \not\to 0$. In fact (Te_n) doesn't even converge.

20.2 Weak Convergence and Orthonormal Basis

Problem 2.a

Proposition 20.3. *Let $\mathcal{Y}$ be a dense subset of $\mathcal{H}$. Let (x_n) be a bounded sequence of elements in $\mathcal{H}$ and suppose $\langle x_n, y\rangle \to \langle x, y\rangle$ for all $y \in \mathcal{Y}$. Then $x_n \xrightarrow{w} x$.*

Proof. Let $z \in \mathcal{H}$. Choose $M \geq 0$ such that $\|x_n\| \leq M$ for all $n \in \mathbb{N}$. Choose $y \in \mathcal{Y}$ such that

$$\|z - y\| < \frac{\varepsilon}{3\max\{\|x\|, M\}}$$

(we can do this since $\mathcal{Y}$ is dense in $\mathcal{H}$). Choose $N \in \mathbb{N}$ such that $n \geq N$ implies

$$|\langle x_n, y\rangle - \langle x, y\rangle| < \frac{\varepsilon}{3}.$$

Then $n \geq N$ implies

$$\begin{aligned}|\langle x_n, z\rangle - \langle x, z\rangle| &= |\langle x_n, z\rangle - \langle x_n, y\rangle + \langle x_n, y\rangle - \langle x, y\rangle + \langle x, y\rangle - \langle x, z\rangle|\\ &\leq |\langle x_n, z - y\rangle| + |\langle x_n, y\rangle - \langle x, y\rangle| + |\langle x, y - z\rangle|\\ &\leq M\|z - y\| + |\langle x_n, y\rangle - \langle x, y\rangle| + \|x\|\|y - z\|\\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3}\\ &= \varepsilon.\end{aligned}$$

Therefore $x_n \xrightarrow{w} x$. □

Problem 2.b

Lemma 20.1. *Let $\mathcal{H}$ be a separable Hilbert space, let (e_m) be an othonormal basis in $\mathcal{H}$, let (x_n) be a bounded sequence in $\mathcal{H}$, and let $x \in \mathcal{H}$. Then $x_n \xrightarrow{w} x$ if and only if $\langle x_n, e_m \rangle \to \langle x, e_m \rangle$ for all $m \in \mathbb{N}$.*

Proof. Suppose $x_n \xrightarrow{w} x$. Then $\langle x_n, y \rangle \to \langle x, y \rangle$ for all $y \in \mathcal{H}$, so certainly $\langle x_n, e_m \rangle \to \langle x, e_m \rangle$ for all $m \in \mathbb{N}$. Conversely, suppose $\langle x_n, e_m \rangle \to \langle x, e_m \rangle$ for all $m \in \mathbb{N}$. We first show that $\langle x_n, y \rangle \to \langle x, y \rangle$ for all $y \in \mathcal{Y}$, where $\mathcal{Y} = \operatorname{span}\{e_m \mid m \in \mathbb{N}\}$, so let $y \in \mathcal{Y}$. Then

$$y = \lambda_1 e_{m_1} + \cdots + \lambda_r e_{m_r}$$

for some (unique) $\lambda_1, \ldots, \lambda_r \in \mathbb{C}$ and $m_1, \ldots, m_r \in \mathbb{N}$. Then

$$\begin{aligned}
\lim_{n\to\infty} \langle x_n, y \rangle &= \lim_{n\to\infty} \langle x_n, \lambda_1 e_{m_1} + \cdots + \lambda_r e_{m_r} \rangle \\
&= \overline{\lambda}_1 \lim_{n\to\infty} \langle x_n, e_{m_1} \rangle + \cdots + \overline{\lambda}_r \lim_{n\to\infty} \langle x_n, e_{m_r} \rangle \\
&= \overline{\lambda}_1 \langle x, e_{m_1} \rangle + \cdots + \overline{\lambda}_r \langle x, e_{m_r} \rangle \\
&= \langle x, \lambda_1 e_{m_1} + \cdots + \lambda_r e_{m_r} \rangle \\
&= \langle x, y \rangle.
\end{aligned}$$

Therefore $\langle x_n, y \rangle \to \langle x, y \rangle$ for all $y \in \mathcal{Y}$. Now since $\overline{\mathcal{Y}} = \mathcal{H}$, we may use Proposition (20.3) to conclude that $\langle x_n, z \rangle \to \langle x, z \rangle$ for all $z \in \mathcal{H}$. In other words, we have $x_n \xrightarrow{w} x$. $\square$

Corollary 11. *Let (x^n) be a sequence in $\ell^2(\mathbb{N})$ that converges coordinate-wise to $x = (x_m) \in \ell^2(\mathbb{N})$. Then $x^n \xrightarrow{w} x$.*

Proof. Saying (x^n) converges coordinate-wise to $x = (x_m)$ is equivalent to saying

$$\begin{aligned}
\lim_{n\to\infty} (\langle x^n, e^m \rangle) &= x_m \\
&= \langle x, e^m \rangle
\end{aligned}$$

for all $m \in \mathbb{N}$. Thus Lemma (20.1) implies $x^n \xrightarrow{w} x$. $\square$

20.3 If $\mathcal{K}$ is a Closed Subspace of $\mathcal{H}$ and (x_n) is a Sequence in $\mathcal{K}$ such that $x_n \xrightarrow{w} x$, then $x \in \mathcal{K}$

Proposition 20.4. *Let $\mathcal{K}$ be a closed subspace of $\mathcal{H}$. If (x_n) is a sequence of elements in $\mathcal{K}$ and $x_n \xrightarrow{w} x$, then $x \in \mathcal{K}$.*

Proof. Let $y \in \mathcal{K}^\perp$. Then

$$\begin{aligned}
0 &= \lim_{n\to\infty} (0) \\
&= \lim_{n\to\infty} (\langle x_n, y \rangle) \\
&= \langle x, y \rangle.
\end{aligned}$$

This implies $x \in (\mathcal{K}^\perp)^\perp = \mathcal{K}$. $\square$

Remark 39. The same proof shows that if A is a subset of $\mathcal{H}$ and (x_n) is a sequence of elements in A and $x_n \xrightarrow{w} x$, then $x \in \overline{\operatorname{span}}(A)$.

20.4 T Bounded and S Compact Implies TS and ST Compact

Proposition 20.5. *Let $T \colon \mathcal{H} \to \mathcal{H}$ be a bounded operator and let $S \colon \mathcal{H} \to \mathcal{H}$ be a compact operator. Then $TS \colon \mathcal{H} \to \mathcal{H}$ and $ST \colon \mathcal{H} \to \mathcal{H}$ are both compact operators.*

Proof. We first show ST is compact. Let (x_n) be a sequence in $\mathcal{H}$ such that $x_n \xrightarrow{w} x$. Since T is a bounded operator, we have $Tx_n \xrightarrow{w} Tx$ by Proposition (20.2). Since S is compact, we have $S(Tx_n) \to S(Tx)$. Thus, ST is compact. Now we show TS is compact. Let (x_n) be a sequence in $\mathcal{H}$ such that $x_n \xrightarrow{w} x$. Since S is compact, we have $Sx_n \to Sx$. Since T is bounded (and in particular continuous) we have $T(Sx_n) \to T(Sx)$. Thus, TS is compact. $\square$

20.5 Equivalent Definition of Compact Operator

Lemma 20.2. *Let (x_n) be a sequence in $\mathcal{H}$ such that $x_n \xrightarrow{w} x$ and let $(x_{\pi(n)})$ be a subsequence of (x_n). Then $x_{\pi(n)} \xrightarrow{w} x$.*

Remark 40. Here we view π as a strictly increasing function from $\mathbb{N}$ to $\mathbb{N}$ whose range consists of the indices in the subsequence.

Proof. Let $y \in \mathcal{H}$ and let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies

$$|\langle x_n, y\rangle - \langle x, y\rangle| < \varepsilon.$$

Then $\pi(n) \geq N$ implies

$$|\langle x_{\pi(n)}, y\rangle - \langle x, y\rangle| < \varepsilon$$

□

Proposition 20.6. *Let $T \colon \mathcal{H} \to \mathcal{H}$ be a bounded operator with the property that for any bounded sequence (x_n) in $\mathcal{H}$, the sequence (Tx_n) has a convergent subsequence. Then T is compact.*

Proof. Let (x_n) be a sequence in $\mathcal{H}$ such that $x_n \xrightarrow{w} x$. Assume for a contradiction that $Tx_n \not\to Tx$. Choose $\varepsilon > 0$ and choose a subsequence $(Tx_{\pi(n)})$ of (Tx_n) such that $\|Tx_{\pi(n)} - Tx\| > \varepsilon$ for all $n \in \mathbb{N}$. By the (baby version) of the uniform boundedness principle, the sequence (x_n) is bounded, and hence the subsequence $(x_{\pi(n)})$ must be bounded too. Thus the sequence $(Tx_{\pi(n)})$ has a convergent subsequence (by the hypothesis on T). Choose a convergent subsequence of $(Tx_{\pi(n)})$, say $(Tx_{\rho(n)})$. Since $(x_{\rho(n)})$ is a subsequence of (x_n), we must have $x_{\rho(n)} \xrightarrow{w} x$ by Lemma (20.2), and since T is a bounded operator, we must have $Tx_{\rho(n)} \xrightarrow{w} Tx$ by Proposition (20.2). Since $(Tx_{\rho(n)})$ is a convergent sequence and since $Tx_{\rho(n)} \xrightarrow{w} Tx$, we must in fact have $Tx_{\rho(n)} \to Tx$. But since $(Tx_{\rho(n)})$ is a subsequence of $(Tx_{\pi(n)})$, we have $\|Tx_{\rho(n)} - Tx\| > \varepsilon$ for all $n \in \mathbb{N}$, and so $Tx_{\rho(n)} \not\to Tx$. This is a contradiction. □

20.6 Finite-Dimensional Spaces, Bounded Sequences, and Compact Operators

Let $\mathcal{K}$ be a Hilbert space.

Problem 6.a

Proposition 20.7. *If every bounded sequence in $\mathcal{K}$ has a convergent subsequence, then $\mathcal{K}$ must be finite-dimensional.*

Proof. We prove the contrapositive statement: if $\mathcal{K}$ is infinite-dimensional, then there exists a bounded sequence which has no convergent subsequence. Assume $\mathcal{K}$ is infinite-dimensional. Let (e_n) be an orthonormal sequence in $\mathcal{K}$. Then the sequence (e_n) cannot have a convergent subsequence. Indeed, the distance squared between any two elements e_n, e_m (with $m \neq n$) in the sequence is

$$\begin{aligned}\|e_n - e_m\|^2 &= \|e_n\|^2 + \|e_m\|^2 \\ &= 1 + 1 \\ &= 2,\end{aligned}$$

by the Pythagorean Theorem. Thus any subsequence of (e_n) will fail to be Cauchy (and in particular will fail to converge). □

Problem 6.b

Proposition 20.8. *Let $T \colon \mathcal{H} \to \mathcal{H}$ be compact. Then $\ker(1_{\mathcal{H}} - T)$ is finite-dimensional.*

Proof. Let (x_n) be a bounded sequence in $\ker(1_{\mathcal{H}} - T)$. Choose a subsequence $(x_{\pi(n)})$ of (x_n) such that $x_{\pi(n)} \xrightarrow{w} x$ for some $x \in \mathcal{H}$ (such a subsequence exists by a theorem proved in class). Since T is compact, we have

$$\begin{aligned}x_{\pi(n)} &= Tx_{\pi(n)} \\ &\to Tx \\ &= x.\end{aligned}$$

Thus $(x_{\pi(n)})$ is a convergent subsequence of (x_n). It follows from Proposition (20.7) that $\ker(1_{\mathcal{H}} - T)$ is finite-dimensional. □

Problem 6.c

Lemma 20.3. *Let $S\colon \mathcal{H} \to \mathcal{H}$ be a bounded operator. Assume im(S) is closed. Then im(S^*) is closed.*

we have $\mathcal{H} = \ker(S^*) \oplus \mathrm{im}(S)$

Proof. It suffices to show that $\ker(S)^\perp \subseteq \mathrm{im}(S^*)$. Let $x \in \ker(S)^\perp$, so $\langle y, x\rangle = 0$ for all $y \in \ker(S)$. We claim that $x = S^*z$ for some $S^*z \in \mathrm{im}(S^*)$. Indeed we have

$$\begin{aligned}\langle y, x\rangle &= \\ &= \langle Sy, z\rangle \\ &= \langle y, S^*z\rangle\end{aligned}$$

for all $y \in \mathcal{H}$.
Let $y \in \overline{\mathrm{im}}(S^*)$. Observe that $\langle y, x\rangle = 0$ for all $x \in \ker(S)$. Chooose a sequence (S^*x_n) in $\mathrm{im}(S^*)$ such that such that $S^*x_n \to y$. Since (S^*x_n) is convergent, the sequence (Sx_n) is convergent too. Indeed, let $\varepsilon > 0$.
For all $Sx \in \mathrm{im}(S)$, we have

$$\begin{aligned}\langle S^*x_n, Sx\rangle &\to \langle y, Sx\rangle \\ &= \langle S^*y, x\rangle.\end{aligned}$$

Similarly, for all $x \in \ker(S^*)$, we have

$$\begin{aligned}\langle S^*x_n, Sx\rangle &\to \langle y, x\rangle \\ &= \langle S^*y, x\rangle.\end{aligned}$$

□

Proposition 20.9. *Let $S\colon \mathcal{H} \to \mathcal{H}$ be a bounded operator whose image im(S) is closed and infinite-dimensional. Then S cannot be compact.*

Proof. Let (y_n) be a bounded sequence in $\mathrm{im}(S)$ and let $\mathcal{K} := \ker(S)$. Choose $N > 0$ such that $\|y_n\| \leq N$ for all $n \in \mathbb{N}$. Choose a sequence (x_n) in $\mathcal{H}$ such that $Sx_n = y_n$ for all $n \in \mathbb{N}$. By replacing x_n with $x_n - \mathrm{P}_{\mathcal{K}}x_n$ if necessary, we may assume that $\langle x_n, z\rangle = 0$ for all $n \in \mathbb{N}$ and for all $z \in \mathcal{K}$. We claim that for each $z \in \mathcal{H}$, the sequence $(|\langle x_n, z\rangle|)$ is bounded. Indeed, since $\mathcal{H} = \ker(S) \oplus \overline{\mathrm{im}}(S^*)$, it suffices to show that for each $z \in \overline{\mathrm{im}}(S^*)$, the sequence $(|\langle x_n, z\rangle|)$. Let $z \in \overline{\mathrm{im}}(S^*)$ and choose a sequence (S^*z_m) in $\mathrm{im}(S^*)$ such that $S^*z_m \to z$. Then

$$\begin{aligned}|\langle x_n, S^*z_m\rangle| &= |\langle Sx_n, z_m\rangle| \\ &= |\langle y_n, z_m\rangle| \\ &\leq \|y_n\|\|z_m\| \\ &\leq N\|z_m\|.\end{aligned}$$

and choose a sequence (S^*z_m) in $\mathrm{im}(S^*)$ such that $S^*z_m \to z$. Then

$$\begin{aligned}|\langle x_n, S^*x\rangle| &= |\langle Sx_n, x\rangle| \\ &= |\langle y_n, x\rangle| \\ &\leq \|y_n\|\|x\| \\ &\leq N\|x\|.\end{aligned}$$

Thus for each $z \in \mathcal{H}$, the sequence $(|\langle x_n, z\rangle|)$ is bounded. It follows from the Uniform Boundedness Principle (stated in Theorem (12.3) in the Appendix) that the sequence (x_n) is bounded. Since S is compact and (x_n) is bounded, the sequence $(Sx_n) = (y_n)$ must have a convergent subsequence. This contradicts the fact that $\mathcal{K}$ is infinite-dimensional. □

Proof using Open Mapping Theorem:

Proof. Assume (for a contradiction) that S is compact. Denote $\mathcal{K} := \mathrm{im}(S)$ and let (y_n) be a bounded sequence in $\mathcal{K}$. Let $M > 0$ and choose $N > 0$ such that if $y \in \mathcal{K}$ and $\|y\| < N$, then there exists an $x \in \mathcal{H}$ such that $Sx = y$ and $\|x\| < M$ (this follows from the Open Mapping Theorem). By scaling the sequence (y_n) if necessary, we may assume that $\|y_n\| < N$ for all $n \in \mathbb{N}$. Thus there exists $x_n \in \mathcal{H}$ such that $Sx_n = y$ and $\|x_n\| < M$ for all $n \in \mathbb{N}$. Thus (x_n) is a bounded sequence in $\mathcal{H}$. Since S is compact and (x_n) is bounded, the sequence $(Sx_n) = (y_n)$ must have a convergent subsequence. This contradicts the fact that $\mathcal{K}$ is infinite-dimensional. □

21 Homework 8

Throughout this homework, let $\mathcal{H}$ be a separable Hilbert space. If $x \in \mathcal{H}$ and $r > 0$, then we write

$$B_r(x) := \{y \in \mathcal{H} \mid \|y - x\| < r\}$$

for the open ball centered at x and of radius r. We also write

$$B_r[x] := \{y \in \mathcal{H} \mid \|y - x\| \leq r\}$$

for the closed ball centered at x and of radius r.

21.1 Equivalent Definition of Compact Operator

Proposition 21.1. *Let $T \colon \mathcal{H} \to \mathcal{H}$ be a bounded linear operator. Then T is compact if and only if $\overline{T(B_1[0])}$ is a compact space.*

Proof. Suppose T is compact. To show that $\overline{T(B_1[0])}$ is compact, it suffices to show that $T(B_1[0])$ is precompact, by Proposition (21.9) (stated and proved in the Appendix). Let (Tx_n) be a sequence in $T(B_1[0])$. Then (x_n) is a bounded sequence in $B_1[0]$. Since T is compact, it follows that (Tx_n) has a convergent subsequence (by homework 7 probem 5). It follows that $T(B_1[0])$ is precompact.

Conversely, suppose $\overline{T(B_1[0])}$ is compact. Then $T(B_1[0])$ is precompact by Proposition (21.9). Let (x_n) be a bounded sequence in $\mathcal{H}$. Choose $M > 0$ such that $\|x_n\| < M$ for all $n \in \mathbb{N}$. Then $(T(x_n/M))$ is a sequence in the precompact space $T(B_1[0])$, and hence must have a convergent subsequence, say $(T(x_{\pi(n)}/M))$. This impies $(T(x_{\pi(n)}))$ is a convergent subsequence $(T(x_n))$. Thus, T is compact (again by homework 7 probem 5). □

21.2 Sequence of Compact Operators (T_n) Which Converge in Operator Norm to T Implies T is Compact

Proposition 21.2. *Let $(T_n \colon \mathcal{H} \to \mathcal{H})$ be a sequence of compact operators that converges in the operator norm to an operator $T \colon \mathcal{H} \to \mathcal{H}$. Then T is compact.*

Proof. Let (x_k) be a weakly convegent sequence. We claim that (Tx_k) is Cauchy. Indeed, let $\varepsilon > 0$. Since (x_k) is weakly convergent, it must be bounded. Choose $M > 0$ such that $\|x_k\| \leq M$ for all $k \in \mathbb{N}$. Choose $N \in \mathbb{N}$ such that $\|T - T_N\| < \varepsilon/3M$. Since the sequence $(T_N x_k)_{k\in\mathbb{N}}$ is Cauchy, there exists $K \in \mathbb{N}$ such that $j, k \geq K$ implies $\|T_N x_k - T_N x_j\| < \varepsilon/3$. Choose such a $K \in \mathbb{N}$. Then $j, k \geq K$ implies

$$\begin{aligned}
\|Tx_k - Tx_j\| &= \|Tx_k - T_N x_k + T_N x_k - T_N x_j + T_N x_j - Tx_j\| \\
&\leq \|Tx_k - T_N x_k\| + \|T_N x_k - T_N x_j\| + \|T_N x_j - Tx_j\| \\
&\leq \|T - T_N\|\|x_k\| + \|T_N x_k - T_N x_j\| + \|T_N - T\|\|x_j\| \\
&< \frac{\varepsilon}{3M} \cdot M + \frac{\varepsilon}{3} + \frac{\varepsilon}{3M} \cdot M \\
&= \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \\
&= \varepsilon.
\end{aligned}$$

Thus (Tx_k) is a Cauchy sequence. It follows that T is compact. □

21.3 Hilbert-Schmidt Operator

Proposition 21.3. *Let $T \colon \mathcal{H} \to \mathcal{H}$ be a bounded operator and let (e_n) and (f_m) be any two orthonormal bases for $\mathcal{H}$. Then*

$$\sum_{n=1}^{\infty} \|Te_n\|^2 = \sum_{m=1}^{\infty} \|T^* f_m\|^2.$$

Proof. Since $\mathcal{H}$ is a separable Hilbert space, we have

$$\|x\|^2 = \sum_{n=1}^{\infty} |\langle x, e_n \rangle|^2 \quad \text{and} \quad \|x\|^2 = \sum_{m=1}^{\infty} |\langle x, f_m \rangle|^2$$

for every $x \in \mathcal{H}$. Therefore

$$\begin{aligned}
\sum_{n=1}^{\infty} \|Te_n\|^2 &= \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} |\langle Te_n, f_m \rangle|^2 \\
&= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} |\langle Te_n, f_m \rangle|^2 \\
&= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} |\langle e_n, T^* f_m \rangle|^2 \\
&= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} |\langle T^* f_m, e_n \rangle|^2 \\
&= \sum_{m=1}^{\infty} \|T^* f_m\|^2,
\end{aligned}$$

where we are justified in changing the order of the infinite sums by Lemma (21.1) (stated and proved in the Appendix). By swapping the roles of T with T^* in the proof above, we see that the quantity $\sum_{n=1}^{\infty} \|Te_n\|^2$ doesn't depend on the choice of the orthonormal basis (e_n). □

21.4 Hilbert-Schmidt Operator

Definition 21.1. An operator $T\colon \mathcal{H} \to \mathcal{H}$ is said to be a **Hilbert-Schmidt** operator if if

$$\sum_{n=1}^{\infty} \|Te_n\|^2 < \infty$$

for some or equivalently any orthonormal basis (e_n) of $\mathcal{H}$. In this case, the Hilbert-Schmidt norm of T is defined by

$$\|T\|_{\text{HS}} := \sqrt{\sum_{n=1}^{\infty} \|Te_n\|^2}.$$

Problem 4.a

Proposition 21.4. *Let (e_n) be an orthonormal basis of $\mathcal{H}$. For each $k \in \mathbb{N}$ define a projection operator $P_k\colon \mathcal{H} \to \mathcal{H}$ onto span$\{e_1, e_2, \dots, e_k\}$ by*

$$P_k(x) = \sum_{n=1}^{k} \langle x, e_n \rangle e_n$$

for all $x \in \mathcal{H}$. If $T\colon \mathcal{H} \to \mathcal{H}$ is a Hilbert-Schmidt operator, then $\|T - P_k T\| \to 0$ as $k \to \infty$.

Proof. Let $\varepsilon > 0$ and let $x \in B_1[0]$. Since the sum $\sum_{n=1}^{\infty} \|T^* e_n\|^2$ converges, there exists $K \in \mathbb{N}$ such that

$$\sum_{n=K}^{\infty} \|T^* e_n\|^2 < \varepsilon.$$

Choose such $K \in \mathbb{N}$. Then $k \geq K$ implies

$$\begin{aligned}
\|Tx - \mathrm{P}_k Tx\|^2 &= \left\| \sum_{n=1}^{\infty} \langle Tx, e_n \rangle e_n - \sum_{n=1}^{k} \langle Tx, e_n \rangle e_n \right\|^2 \\
&= \left\| \sum_{n=k+1}^{\infty} \langle Tx, e_n \rangle e_n \right\|^2 \\
&= \left\| \sum_{n=k+1}^{\infty} \langle x, T^* e_n \rangle e_n \right\|^2 \\
&= \sum_{n=k+1}^{\infty} |\langle x, T^* e_n \rangle|^2 \\
&\leq \sum_{n=k+1}^{\infty} \|T^* e_n\|^2 \\
&\leq \sum_{n=K}^{\infty} \|T^* e_n\|^2 \\
&< \varepsilon.
\end{aligned}$$

This implies $\|T - \mathrm{P}_k T\| \to 0$ as $k \to \infty$ by Remark (41) (stated in the Appendix). □

Problem 4.b

Proposition 21.5. *Every Hilbert-Schmidt operator is compact.*

Proof. Let $T \colon \mathcal{H} \to \mathcal{H}$ be a Hilbert-Schmidt operator. To show that T is compact, it suffices to show that $\mathrm{P}_k T$ is compact for all $k \in \mathbb{N}$ since Proposition (21.4) implies $\|\mathrm{P}_k T - T\| \to 0$ as $k \to \infty$ and Proposition (21.2) would then imply T is compact.

Let $k \in \mathbb{N}$ and let (x_n) be a weakly convergent sequence in $\mathcal{H}$, say $x_n \xrightarrow{w} x$. We claim that $\mathrm{P}_k x_n \to \mathrm{P}_k x$ as $n \to \infty$. Indeed, let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies

$$|\langle x_n, e_m \rangle - \langle x, e_m \rangle| < \frac{\varepsilon}{k}$$

for all $m = 1, \ldots, k$. Then $n \geq N$ implies

$$\begin{aligned}
\|\mathrm{P}_k x_n - \mathrm{P}_k x\| &= \left\| \sum_{m=1}^{k} \langle x_n, e_m \rangle e_m - \sum_{m=1}^{k} \langle x, e_m \rangle e_m \right\| \\
&= \left\| \sum_{m=1}^{k} (\langle x_n, e_m \rangle - \langle x, e_m \rangle) e_m \right\| \\
&\leq \sum_{m=1}^{k} |\langle x_n, e_m \rangle - \langle x, e_m \rangle| \\
&< \sum_{m=1}^{k} \frac{\varepsilon}{k} \\
&= \varepsilon.
\end{aligned}$$

□

Problem 4.c

Proposition 21.6. *Let $T \colon \mathcal{H} \to \mathcal{H}$ be a Hilbert-Schmidt operator. Then $\|T\| \leq \|T\|_{HS}$.*

Proof. Let $x \in B_1[0]$. Then

$$\begin{aligned}
\|Tx\|^2 &= \sum_{n=1}^{\infty} |\langle Tx, e_n\rangle|^2 \\
&= \sum_{n=1}^{\infty} |\langle x, T^* e_n\rangle|^2 \\
&\le \sum_{n=1}^{\infty} \|T^* e_n\|^2 \\
&= \|T\|_{\text{HS}}^2.
\end{aligned}$$

In particular this implies

$$\begin{aligned}
\|T\|^2 &= \sup\{\|Tx\|^2 \mid x \in B_1[0]\} \\
&\le \|T\|_{\text{HS}}^2,
\end{aligned}$$

where the first line is justified in the Appendix. Thus $\|T\| = \|T\|_{\text{HS}}$. □

21.5 If T is Compact Self-Adjoint and $T^m = 0$, then $T = 0$

Proposition 21.7. *Let $T \colon \mathcal{H} \to \mathcal{H}$ be a compact self-adjoint operator. Suppose $T^m = 0$ for some $m \in \mathbb{N}$. Then we must have $T = 0$.*

Proof. If $T^m = 0$ for some $m \in \mathbb{N}$, then 0 is the only eigenvalue for T. Indeed, suppose λ is an eigenvalue of T. Choose an eigenvector of λ, say x. Then

$$\begin{aligned}
0 &= T^m x \\
&= \lambda^m x,
\end{aligned}$$

which implies $\lambda^m = 0$, and hence $\lambda = 0$. Now choose an orthonormal basis (e_n) consisting of eigenvectors of T (the existence of such basis is guaranteed by the spectral theorem for compact self-adjoint operators). Then for all $x \in \mathcal{H}$, we have

$$\begin{aligned}
Tx &= \sum_{n=1}^{\infty} \lambda_n \langle x, e_n\rangle e_n \\
&= \sum_{n=1}^{\infty} 0 \cdot \langle x, e_n\rangle e_n \\
&= 0.
\end{aligned}$$

□

21.6 Every Compact Self-Adjoint Operator is Limit of Operators with Finite-Dimensional Range

Proposition 21.8. *Let $\mathcal{H}$ be a separable Hilbert space and let $T \colon \mathcal{H} \to \mathcal{H}$ be a compact self-adjoint operator. Then there exists a sequence T_m of operators with finite dimensional range such that $\|T - T_m\| \to 0$ and $m \to \infty$.*

Proof. Choose an orthonormal basis (e_n) consisting of eigenvectors of T and let (λ_n) be the corresponding sequence of eigenvalues. By reindexing if necessary, we may assume that $|\lambda_n| \ge |\lambda_{n+1}|$ for all $n \in \mathbb{N}$. For each $m \in \mathbb{N}$, we define $T_m \colon \mathcal{H} \to \mathcal{H}$ by

$$T_m x = \sum_{n=1}^{m} \lambda_n \langle x, e_n\rangle e_n$$

for all $x \in \mathcal{H}$. Observe that $\operatorname{im}(T_m) = \operatorname{span}(\{e_1, \dots, e_m\})$ is finite dimensional. We claim that $\|T - T_m\| \to 0$ and $m \to \infty$. Indeed, let $\varepsilon > 0$ and let Λ denote the set of all eigenvalues of T. If Λ is finite, then the claim is clear by the spectral theorem for compact self-adjoint operators, so assume Λ is infinite. Then 0 must be an accumulation point of Λ. In particular, $|\lambda_m| \to 0$ as $m \to \infty$. Choose $M \in \mathbb{N}$ such that $m \geq M$ implies $|\lambda_m| < \varepsilon$. Then $m \geq M$ implies

$$\begin{aligned}
\|Tx - T_m x\|^2 &= \left\| \sum_{n=1}^{\infty} \lambda_n \langle x, e_n \rangle e_n - \sum_{n=1}^{m} \lambda_n \langle x, e_n \rangle e_n \right\|^2 \\
&= \left\| \sum_{n=m+1}^{\infty} \lambda_n \langle x, e_n \rangle e_n \right\|^2 \\
&= \sum_{n=m+1}^{\infty} |\lambda_n \langle x, e_n \rangle|^2 \\
&\leq |\lambda_M|^2 \sum_{n=m+1}^{\infty} |\langle x, e_n \rangle|^2 \\
&\leq |\lambda_M|^2 \|x\|^2 \\
&< \varepsilon^2.
\end{aligned}$$

for all $x \in B_1[0]$. This implies $\|T - T_m\| \to 0$ and $m \to \infty$. □

Appendix

Problem 1

Definition 21.2. A subspace $A \subseteq \mathcal{H}$ is said to be **precompact** if every sequence in A has a convergent subsequence.

Proposition 21.9. *Let A be a subspace of $\mathcal{H}$. Then A is precompact if and only if $\overline{A}$ is compact.*

Proof. Suppose A is precompact. Let (a_n) be a sequence in $\overline{A}$. For each $n \in \mathbb{N}$ choose $b_n \in A$ such that

$$\|a_n - b_n\| < \frac{1}{n}.$$

Choose a convergent subsequence of (b_n), say $(b_{\pi(n)})$ (we can do this since A is precompact). We claim that the sequence $(a_{\pi(n)})$ is Cauchy, and hence convergent subsequence of (a_n). Indeed, let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $\pi(n) \geq \pi(m) \geq N$ implies

$$\|b_{\pi(n)} - b_{\pi(m)}\| < \frac{\varepsilon}{3} \quad \text{and} \quad \frac{1}{\pi(m)} < \frac{\varepsilon}{3}.$$

Then $\pi(n) \geq \pi(m) \geq N$ implies

$$\begin{aligned}
\|a_{\pi(n)} - a_{\pi(m)}\| &= \|a_{\pi(n)} - b_{\pi(n)} + b_{\pi(n)} - b_{\pi(m)} + b_{\pi(m)} - a_{\pi(m)}\| \\
&\leq \|a_{\pi(n)} - b_{\pi(n)}\| + \|b_{\pi(n)} - b_{\pi(m)}\| + \|b_{\pi(m)} - a_{\pi(m)}\| \\
&< \frac{1}{\pi(n)} + \frac{\varepsilon}{3} + \frac{1}{\pi(m)} \\
&\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \\
&= \varepsilon.
\end{aligned}$$

Finally, since $(a_{\pi(n)})$ is Cauchy and since $\mathcal{H}$ is a Hilbert space, we must have $a_{\pi(n)} \to a$ for some $a \in \overline{A}$. Therefore $\overline{A}$ is compact.

Conversely, suppose $\overline{A}$ is compact. Let (a_n) be a sequence in A. Then (a_n) is a sequence in $\overline{A}$. Since $\overline{A}$ is compact, the sequence (a_n) has a convergent subsequence. Therefore A is precompact. □

Convergence in Operator Norm

Remark 41. Let $\mathcal{V}$ be an inner-product space and let $(T_n \colon \mathcal{V} \to \mathcal{V})$ be a sequence of bounded linear operators. If we want to show $\|T_n - T\| \to 0$ as $n \to \infty$, then it suffices to show that for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$\|T_n x - Tx\| < \varepsilon$$

for all $x \in B_1[0]$. Indeed, assuming this is true, choose $M \in \mathbb{N}$ such that $n \geq M$ implies

$$\|T_n x - Tx\| < \varepsilon/2$$

for all $x \in B_1[0]$. Then $n \geq M$ implies

$$\begin{aligned}\|T_n - T\| &= \sup\{\|T_n x - Tx\| \mid x \in B_1[0]\} \\ &\leq \varepsilon/2 \\ &< \varepsilon.\end{aligned}$$

Problem 3

Lemma 21.1. *Let f be a nonnegative function defined on $\mathbb{N} \times \mathbb{N}$. Then*

$$\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} f(m,n) = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} f(m,n).$$

Proof. Let $M \in \mathbb{N}$. Then

$$\begin{aligned}\sum_{m=1}^{M}\sum_{n=1}^{\infty} f(m,n) &= \sum_{m=1}^{M}\lim_{N\to\infty}\sum_{n=1}^{N} f(m,n) \\ &= \lim_{N\to\infty}\sum_{n=1}^{N}\sum_{m=1}^{M} f(m,n) \\ &= \sum_{n=1}^{\infty}\sum_{m=1}^{M} f(m,n) \\ &\leq \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} f(m,n).\end{aligned}$$

Taking the limit as $M \to \infty$ gives us

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} f(m,n) \leq \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} f(m,n).$$

A similar argument gives us

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} f(m,n) \geq \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} f(m,n).$$

□

Problem 4.c

Proposition 21.10. *Let $T \colon \mathcal{U} \to \mathcal{V}$ be a bounded linear operator. Then*

$$\|T\|^2 = \sup\{\|Tx\|^2 \mid \|x\| \leq 1\}$$

Proof. For any $x \in \mathcal{U}$ such that $\|x\| \leq 1$, we have $\|Tx\|^2 \leq \|T\|^2$. Thus

$$\|T\|^2 \geq \sup\{\|Tx\|^2 \mid \|x\| \leq 1\}. \tag{46}$$

To show the reverse inequality, we assume (for a contradiction) that (46) is a strict inequality. Choose $\delta > 0$ such that

$$\|T\|^2 - \delta > \sup\{\|Tx\|^2 \mid \|x\| \leq 1\}.$$

Now let $\varepsilon = \delta/2\|T\|$, and choose $x \in \mathcal{U}$ such that $\|x\| \leq 1$ and such that

$$\|T\| - \varepsilon < \|Tx\|.$$

Then

$$\begin{aligned}\|Tx\|^2 &> (\|T\| - \varepsilon)^2 \\ &= \|T\|^2 - 2\varepsilon\|T\| + \varepsilon^2 \\ &\geq \|T\|^2 - 2\varepsilon\|T\| \\ &= \|T\|^2 - \delta\end{aligned}$$

gives us a contradiction. □

22 Homework 9

Throughout this homework, let $\mathcal{H}$ be a separable Hilbert space.

22.1 Singular Values and Eigenvalues of Compact Positive Self-Adjoint Operator Coincide

Proposition 22.1. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a compact positive self-adjoint operator. Then $T = |T|$, and consequently the eigenvalues of T coincide with the singular values of T.*

Proof. Choose an orthonormal eigenbasis (e_n) of T with $Te_n = \lambda_n e_n$ for all $n \in \mathbb{N}$ (this exists since T is compact and self-adjoint). Then (e_n) is an orthonormal basis consisting of eigenvectors of $T^2 = T^*T$ with $T^2 e_n = \lambda_n^2 e_n$ for all $n \in \mathbb{N}$. Then since $\lambda_n \geq 0$ for all $n \in \mathbb{N}$ (since T is positive and self-adjoint), we have

$$\begin{aligned}|T|x &= \sum_{n=1}^{\infty} s_n \langle x, e_n \rangle e_n \\ &= \sum_{n=1}^{\infty} \sqrt{\lambda_n^2} \langle x, e_n \rangle e_n \\ &= \sum_{n=1}^{\infty} \lambda_n \langle x, e_n \rangle e_n \\ &= Tx\end{aligned}$$

for all $x \in \mathcal{H}$. It follows that $T = |T|$, and consequently $s_n = \lambda_n$ for all $n \in \mathbb{N}$. □

22.2 Compact Operator that is not Hilbert-Schmidt

Proposition 22.2. *Let (e_n) be an orthonormal basis for $\mathcal{H}$. Define $T\colon \mathcal{H} \to \mathcal{H}$ by*

$$T(x) = \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} \langle x, e_n \rangle e_n.$$

for all $x \in \mathcal{H}$. Then $T\colon \mathcal{H} \to \mathcal{H}$ is compact but not Hilbert-Schmidt.

Remark 42. For this problem, I decided to prove this in an arbitrary separable Hilbert space than just $\ell^2(\mathbb{N})$.

Proof. We first show T is compact. For each $k \in \mathbb{N}$, define $T_k\colon \mathcal{H} \to \mathcal{H}$ by

$$T_k(x) = \sum_{n=1}^{k} \frac{1}{\sqrt{n}} \langle x, e_n \rangle e_n$$

for all $x \in \mathcal{H}$. First note that for each $k \in \mathbb{N}$, the operator T_k is bounded and has finite rank, and hence must be compact. Moreover, we have $\|T - T_k\| \to 0$ as $k \to \infty$. Indeed, let $\varepsilon > 0$ and let $x \in B_1[0]$ (so $\sum_{n=1}^{\infty} |\langle x, e_n \rangle|^2 \leq 1$).

Choose $K \in \mathbb{N}$ such that $1/K < \varepsilon$. Then $k \geq K$ implies

$$\begin{aligned}
\|Tx - T_k x\|^2 &= \left\| \sum_{n=k+1}^{\infty} \frac{1}{\sqrt{n}} \langle x, e_n \rangle e_n \right\|^2 \\
&= \sum_{n=k+1}^{\infty} \left| \frac{\langle x, e_n \rangle}{\sqrt{n}} \right|^2 \\
&= \sum_{n=k+1}^{\infty} \frac{|\langle x, e_n \rangle|^2}{n} \\
&\leq \frac{1}{K} \sum_{n=k+1}^{\infty} |\langle x, e_n \rangle|^2 \\
&\leq \frac{1}{K} \\
&< \varepsilon.
\end{aligned}$$

This implies $\|T - T_k\| \to 0$ as $k \to \infty$. Thus (T_k) is a sequence of compact operators such that $\|T - T_k\| \to 0$ as $k \to \infty$. Therefore T is compact.

To see that T is not Hilbert-Schmidt, observe that

$$\begin{aligned}
\sum_{n=1}^{\infty} \|Te_n\|^2 &= \sum_{n=1}^{\infty} \|\frac{1}{\sqrt{n}} e_n\|^2 \\
&= \sum_{n=1}^{\infty} \frac{1}{n}
\end{aligned}$$

is the harmonic series which does not converge. □

22.3 Eigenvalue for Self-Adjoint Operator is less than or equal to the Operator Norm

Proposition 22.3. *Let $T \colon \mathcal{H} \to \mathcal{H}$ be a self-adjoint operator and let λ be an eigenvalue of T. Then $|\lambda| \leq \|T\|$.*

Proof. Choose an eigenvector x corresponding to the eigenvalue λ. By scaling if necessary, we may assume $\|x\| = 1$. Then

$$\begin{aligned}
\|T\| &= \sup\{|\langle Ty, y \rangle| \mid \|y\| \leq 1\} \\
&\geq |\langle Tx, x \rangle| \\
&= |\langle \lambda x, x \rangle| \\
&= |\lambda|.
\end{aligned}$$

□

If T is Compact, then $\||T|\| = \|T\|$

Lemma 22.1. *Let $T \colon \mathcal{H} \to \mathcal{H}$ be a compact operator. Then $\||T|\| = \|T\|$.*

Proof. Combining problem 5 on HW5 and problem 6.b on HW6, we have

$$\begin{aligned}
\||T|\|^2 &= \||T|^2\| \\
&= \|T^* T\| \\
&= \|T\|^2.
\end{aligned}$$

It follows that $\||T|\| = \|T\|$ since the norm of an operator is nonnegative. □

Proposition 22.4. *Let $T \colon \mathcal{H} \to \mathcal{H}$ be a compact operator and let s be a singular value of T. Then we have $0 \leq s \leq \|T\|$.*

Proof. Clearly we have $s \geq 0$ by definition. Combining Lemma (22.1) and Proposition (22.3) gives us

$$\begin{aligned}
|s| &\leq \||T|\| \\
&= \|T\|.
\end{aligned}$$

□

22.4 The Square of Hilbert-Schmidt Norm of Compact Operator equals sum of Singular Values Squared

Proposition 22.5. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a compact operator. Let (s_n) be the sequence of singular values of T. Then $\|T\|_{HS} = \sqrt{\sum_{n=1}^\infty s_n^2}$.*

Proof. Let (x_n) be an orthonormal basis for T^*T. Then

$$\begin{aligned}
\|T\|_{\mathrm{HS}} &= \sqrt{\sum_{n=1}^\infty \|Tx_n\|^2} \\
&= \sqrt{\sum_{n=1}^\infty \langle Tx_n, Tx_n\rangle} \\
&= \sqrt{\sum_{n=1}^\infty \langle T^*Tx_n, x_n\rangle} \\
&= \sqrt{\sum_{n=1}^\infty \langle s_n^2 x_n, x_n\rangle} \\
&= \sqrt{\sum_{n=1}^\infty s_n^2}.
\end{aligned}$$

□

22.5 Compact Self-Adjoint Operator which is not Zero

Proposition 22.6. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a compact self-adjoint operator. Then $T^2 + T + 1$ cannot be the zero operator.*

Proof. Choose an orthonormal eigenbasis (e_n) of T with $Te_n = \lambda_n e_n$ for all $n \in \mathbb{N}$. Assume for a contradiction that $T^2 + T + 1 = 0$. Then

$$\begin{aligned}
0 &= (T^2 + T + 1)e_n \\
&= \sum_{n=1}^\infty (\lambda_n^2 + \lambda_n + 1)\langle e_n, e_n\rangle e_n \\
&= (\lambda_n^2 + \lambda_n + 1)e_n,
\end{aligned}$$

which implies $\lambda_n^2 + \lambda_n + 1 = 0$ for all $n \in \mathbb{N}$. Therefore $\lambda_n = \pm e^{2\pi i/3}$ for all $n \in \mathbb{N}$, but this contradicts the fact that the λ_n must be real. □

22.6 Compact Operator is Limit of Operators with Finite-Dimensional Range

Proposition 22.7. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a compact operator. Then there exists a sequence $T_n\colon \mathcal{H} \to \mathcal{H}$ of operators with finite dimensional range such that $\|T_n - T\| \to 0$ as $n \to \infty$.*

Proof. Let $T = U|T|$ be the polar decomposition of T. Choose a sequence (S_n) of bounded operators with finite dimensional range such that $\|S_n - |T|\| \to 0$ as $n \to \infty$ (such a sequence exists by problem 6 HW8). Then for each $n \in \mathbb{N}$, the operator $T_n := US_n$ has finite dimensional range since S_n has finite dimensional range. Moreover we have $\|T - T_n\| \to 0$ as $n \to \infty$. Indeed, let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies $\||T| - S_n\| < \frac{\varepsilon}{\|U\|}$. Then $n \geq N$ implies

$$\begin{aligned}
\|T - T_n\| &= \|U|T| - US_n\| \\
&= \|U(|T| - S_n)\| \\
&\leq \|U\|\||T| - S_n|\| \\
&< \|U\|\frac{\varepsilon}{\|U\|} \\
&= \varepsilon.
\end{aligned}$$

□

23 Homework 10

23.1 $(C[a,b], \|\cdot\|_\infty)$ is a Banach Space

Proposition 23.1. *Let $\|\cdot\|_\infty \colon C[a,b] \times C[a,b] \to \mathbb{R}$ be given by*

$$\|f\|_\infty = \sup\{|f(x)| \mid x \in [a,b]\} \tag{47}$$

for all $f \in C[a,b]$. Then $\|\cdot\|_\infty$ is a norm. Moreover, the pair $(C[a,b], \|\cdot\|_\infty)$ forms a Banach space.

Proof. Let us first show $\|\cdot\|_\infty$ is a norm. First note that the set $\{|f(x)| \mid x \in [a,b]\}$ is non-empty and bounded above (since f is continuous on a compact interval and hence attains a maximum). Therefore the supremum (47) exists.

For positive-definiteness, let $f \in C[a,b]$. Then

$$\begin{aligned}\|f\|_\infty &= \sup\{|f(x)| \mid x \in [a,b]\} \\ &\geq \sup\{0 \mid x \in [a,b]\} \\ &= 0.\end{aligned}$$

We have equality if and only if $|f(x)| = 0$ for all $x \in [a,b]$, and since $|\cdot|$ is positive-definite, this is equivalent to f being the zero function.

For absolute-homogeneity, let $f \in C[a,b]$ and $\alpha \in \mathbb{C}$. Then

$$\begin{aligned}\|\alpha f\|_\infty &= \sup\{|\alpha f(x)| \mid x \in [a,b]\} \\ &= \sup\{|\alpha||f(x)| \mid x \in [a,b]\} \\ &= |\alpha| \sup\{|f(x)| \mid x \in [a,b]\} \\ &= |\alpha|\|f\|_\infty,\end{aligned}$$

where the equality at the third line is justified by Proposition (23.10) (stated and proved in the Appendix).

For subadditivity, let $f, g \in C[a,b]$. Then

$$\begin{aligned}\|f+g\|_\infty &= \sup\{|f(x)+g(x)| \mid x \in [a,b]\} \\ &\leq \sup\{|f(x)|+|g(x)| \mid x \in [a,b]\} \\ &= \sup\{|f(x)| \mid x \in [a,b]\} + \sup\{|g(x)| \mid x \in [a,b]\} \\ &= \|f\|_\infty + \|g\|_\infty,\end{aligned}$$

where the equality at the third line is justified by Proposition (23.11) (stated and proved in the Appendix).

Finally, to show that $(C[a,b], \|\cdot\|_\infty)$ forms a Banach space, we need to show that every Cauchy sequence in $(C[a,b], \|\cdot\|_\infty)$ is convegent. Throughout the rest of the proof, we drop the notation $(C[a,b], \|\cdot\|_\infty)$ and simply write $C[a,b]$ instead. Let (f_n) be a Cauchy sequence in $C[a,b]$. We first make the observation that for each $x \in [a,b]$, the sequence $(f_n(x))$ forms a Cauchy sequence of complex numbers. Indeed, let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $m, n \geq N$ implies $\|f_n - f_m\|_\infty < \varepsilon$. In other words, $m, n \geq N$ implies

$$\sup\{|f_n(x) - f_m(x)| \mid x \in [a,b]\} < \varepsilon.$$

In particular $m, n \geq N$ implies

$$|f_n(x) - f_m(x)| < \varepsilon \tag{48}$$

for all $x \in [a,b]$. This proves our claim.

Since $\mathbb{C}$ is complete, we are justified in defining $f \colon [a,b] \to \mathbb{C}$ by

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

for all $x \in [a,b]$. By taking $m \to \infty$ in (48), we see that (f_n) converges *uniformly* to f. In particular, this implies f is continuous (by the usual $\varepsilon/3$ trick). Thus $f \in C[a,b]$. Finally, we note that convergence in $\|\cdot\|_\infty$ is equivalent to uniform convergence. Thus the Cauchy sequence (f_n) converges in the $\|\cdot\|_\infty$ norm to f. $\square$

23.2 Normed Linear Space + Parallelogram Law = Inner-Product Space

Proposition 23.2. *Let $(V, \|\cdot\|)$ be a normed linear space over $\mathbb{C}$ which satisfies the parallelogram law. Then the map $\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{C}$ defined by*

$$\langle x, y\rangle = \frac{1}{4}\left(\|x+y\|^2 + i\|x+iy\|^2 - \|x-y\|^2 - i\|x-iy\|^2\right) \tag{49}$$

for all $x, y \in V$ is an inner-product. Moreover, the norm induced by this inner-product is precisely $\|\cdot\|$. In other words, we have

$$\langle x, x\rangle = \|x\|^2$$

for all $x \in V$.

Proof. The most difficult part of this proof is showing that (49) is linear in the first argument. Before we do this, let us show that (59) is positive-definite and conjugate-symmetric.

For positive-definiteness, let $x \in V$. Then

$$\begin{aligned}
\langle x, x\rangle &= \frac{1}{4}\left(\|2x\|^2 + i\|x+ix\|^2 - i\|x-ix\|^2\right) \\
&= \frac{1}{4}\left(\|2x\|^2 + i((|1+i|^2 - |1-i|^2)\|x\|^2)\right) \\
&= \|x\|^2 \\
&\geq 0,
\end{aligned}$$

with equality if and only if $x = 0$. Note that this also gives us $\langle x, x\rangle = \|x\|^2$ for all $x \in V$.

For conjugate-symmetry, let $x, y \in V$. Then

$$\begin{aligned}
\overline{\langle y, x\rangle} &= \frac{1}{4}\overline{(\|y+x\|^2 + i\|y+ix\|^2 - \|y-x\|^2 - i\|y-ix\|^2)} \\
&= \frac{1}{4}\left(\|y+x\|^2 - i\|y+ix\|^2 - \|y-x\|^2 + i\|y-ix\|^2\right) \\
&= \frac{1}{4}\left(\|x+y\|^2 - i\|i(x-iy)\|^2 - \|x-y\|^2 + i\|i(x+iy)\|^2\right) \\
&= \frac{1}{4}\left(\|x+y\|^2 - i\|x-iy\|^2 - \|x-y\|^2 + i\|x+iy\|^2\right) \\
&= \frac{1}{4}\left(\|x+y\|^2 + i\|x+iy\|^2 - \|x-y\|^2 - i\|x-iy\|^2\right) \\
&= \langle x, y\rangle
\end{aligned}$$

Now we come to the difficult part, namely showing that (59) is linear in the first argument. We do this in several steps:

Step 1: We show that (59) is additive in the first argument (i.e. $\langle x+z, y\rangle = \langle x, y\rangle + \langle z, y\rangle$ for all $x, y, z \in V$). Let $x, y, z \in V$. First note that by the parallelogram law, we have

$$\begin{aligned}
\|x+z+y\|^2 - \|x+z-y\|^2 &= 2\|x+y\|^2 + 2\|z\|^2 - \|x+y-z\|^2 - 2\|x-y\|^2 - 2\|z\|^2 + \|x-y-z\|^2 \\
&= 2\|x+y\|^2 - 2\|x-y\|^2 - \|z-y-x\|^2 + \|z+y-x\|^2 \\
&= 2\|x+y\|^2 - 2\|x-y\|^2 - 2\|z-y\|^2 - 2\|x\|^2 + \|z-y+x\|^2 + 2\|z+y\|^2 + 2\|x\|^2 - \|z+y+x\|^2 \\
&= 2\|x+y\|^2 - 2\|x-y\|^2 + 2\|z+y\|^2 - 2\|z-y\|^2 + \|x+z-y\|^2 - \|x+z+y\|^2.
\end{aligned}$$

Adding $\|x+z-y\|^2 - \|x+z+y\|^2$ to both sides gives us

$$2(\|x+z+y\|^2 - \|x+z-y\|^2) = 2(\|x+y\|^2 - \|x-y\|^2 + \|z+y\|^2 - \|z-y\|^2),$$

and after cancelling 2 from both sides, we obtain

$$\|x+z+y\|^2 - \|x+z-y\|^2 = \|x+y\|^2 - \|x-y\|^2 + \|z+y\|^2 - \|z-y\|^2.$$

Therefore

$$\begin{aligned}
\langle x+z, y\rangle &= \frac{1}{4}\left(\|x+z+y\|^2 + i\|x+z+iy\|^2 - \|x+z-y\|^2 - i\|x+z-iy\|^2\right) \\
&= \frac{1}{4}\left(\|x+z+y\|^2 - \|x+z-y\|^2 + i(\|x+z+iy\|^2 - \|x+z-iy\|^2)\right) \\
&= \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2 + \|z+y\|^2 - \|z-y\|^2 + i(\|x+iy\|^2 - \|x-iy\|^2 + \|z+iy\|^2 - \|z-iy\|^2)\right) \\
&= \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2 + \|z+y\|^2 - \|z-y\|^2 + i(\|x+iy\|^2 - \|x-iy\|^2 + \|z+iy\|^2 - \|z-iy\|^2)\right) \\
&= \frac{1}{4}\left(\|x+y\|^2 + i\|x+iy\|^2 - \|x-y\|^2 - i\|x-iy\|^2 + \|z+y\|^2 + i\|z+iy\|^2 - \|z-y\|^2 - i\|z-iy\|^2\right) \\
&= \langle x, y\rangle + \langle z, y\rangle.
\end{aligned}$$

Thus we have additivity in the first argument.

Step 2: We show that (59) respects $\mathbb{Z}$-scaling in the first argument (i.e. $m\langle x, y\rangle = \langle mx, y\rangle$ for all integers $m \in \mathbb{Z}$ and for all $x, y \in V$). It suffices to show that (59) respects $\mathbb{N}$-scaling in the first argument since additivity implies

$$\begin{aligned}
0 &= \langle 0, y\rangle \\
&= \langle x - x, y\rangle \\
&= \langle x, y\rangle + \langle -x, y\rangle,
\end{aligned}$$

which implies $\langle -x, y\rangle = -\langle x, y\rangle$ for all $x, y \in V$. We prove (59) respects $\mathbb{N}$-scaling in the first argument using induction on $m \geq 2$. The base case $m = 2$ follows from Step 1. Now assume that for some $m \geq 2$ and for all $x, y \in V$, we have $\langle mx, y\rangle = m\langle x, y\rangle$. Then we have

$$\begin{aligned}
\langle (m+1)x, y\rangle &= \langle mx + x, y\rangle \\
&= \langle mx, y\rangle + \langle x, y\rangle \\
&= m\langle x, y\rangle + \langle x, y\rangle \\
&= (m+1)\langle x, y\rangle,
\end{aligned}$$

where we applied the induction step at the third line.

Step 3: We show that (59) respects $\mathbb{Q}$-scaling in the first argument. Let $\frac{m}{n} \in \mathbb{Q}$ and let $x, y \in V$. Then since (59) is additive in the first argument and since V is a $\mathbb{C}$-vector space, we have

$$\begin{aligned}
\frac{m}{n}\langle x, y\rangle &= \frac{m}{n}\left\langle \frac{n}{n}x, y\right\rangle \\
&= \frac{mn}{n}\left\langle \frac{1}{n}x, y\right\rangle \\
&= m\left\langle \frac{1}{n}x, y\right\rangle \\
&= \left\langle \frac{m}{n}x, y\right\rangle.
\end{aligned}$$

Therefore (59) respects $\mathbb{Q}$-scaling in the first argument.

Step 4: We show that (59) respects $\mathbb{R}$-scaling in the first argument. First note that for each $y \in V$, the map $\langle \cdot, y\rangle \colon V \to \mathbb{C}$ is continuous since the norm is continuous. Let $x, y \in V$ and let $r \in \mathbb{R}$. Choose a sequence (r_n) of rational numbers such that $r_n \to r$ (we can do this since $\mathbb{Q}$ is dense in $\mathbb{R}$). Then we have

$$\begin{aligned}
\langle rx, y\rangle &= \lim_{n\to\infty}\langle r_n x, y\rangle \\
&= \lim_{n\to\infty} r_n\langle x, y\rangle \\
&= r\langle x, y\rangle.
\end{aligned}$$

Therefore (59) respects $\mathbb{R}$-scaling in the first component.

Step 5: We show that (59) respects $\mathbb{C}$-scaling in the first component. We first show that $\langle ix, y\rangle = i\langle x, y\rangle$ for all $x, y \in V$.

Let $x, y \in V$. Then we have

$$\begin{aligned}
\langle ix, y\rangle &= \frac{1}{4}\left(\|ix+y\|^2 + i\|ix+iy\|^2 - \|ix-y\|^2 - i\|ix-iy\|^2\right) \\
&= \frac{1}{4}\left(\|x-iy\|^2 + i\|x+y\|^2 - \|x+iy\|^2 - i\|x-y\|^2\right) \\
&= \frac{1}{4}\left(i\|x+y\|^2 - \|x+iy\|^2 - i\|x-y\|^2 + \|x-iy\|^2\right) \\
&= \frac{i}{4}\left(\|x+y\|^2 + i\|x+iy\|^2 - \|x-y\|^2 - i\|x-iy\|^2\right) \\
&= i\langle x, y\rangle.
\end{aligned}$$

Now let $\lambda = r + is \in \mathbb{C}$. Then we have

$$\begin{aligned}
\langle \lambda x, y\rangle &= \langle (r+is)x, y\rangle \\
&= \langle rx + isx, y\rangle \\
&= \langle rx, y\rangle + \langle isx, y\rangle \\
&= r\langle x, y\rangle + s\langle ix, y\rangle \\
&= r\langle x, y\rangle + is\langle x, y\rangle \\
&= (r+is)\langle x, y\rangle \\
&= \lambda\langle x, y\rangle
\end{aligned}$$

for all $x, y \in V$. Therefore (59) respects $\mathbb{C}$-scaling in the first component. □

23.3 Example of Bounded Operator in $(C[0,1], \|\cdot\|_\infty)$

Proposition 23.3. *Consider $C[0,1]$ equipped with the supremum norm. Let $T\colon C[0,1] \to C[0,1]$ be the linear operator defined by*

$$(Tf)(x) = \int_0^x f(y)\mathrm{d}y$$

for all $x \in [0,1]$. Then T is bounded with $\|T\| = 1$.

Proof. Let $f \in C[0,1]$ such that $\|f\|_\infty \leq 1$. Then

$$\begin{aligned}
\|Tf\|_\infty &= \sup\{|(Tf)(x)| \mid x \in [0,1]\} \\
&= \sup\left\{\left|\int_0^x f(y)\mathrm{d}y\right| \mid x \in [0,1]\right\} \\
&\leq \sup\left\{\int_0^x |f(y)|\mathrm{d}y \mid x \in [0,1]\right\} \\
&\leq \sup\left\{\int_0^x \mathrm{d}y \mid x \in [0,1]\right\} \\
&= \sup\{x \mid x \in [0,1]\} \\
&= 1.
\end{aligned}$$

Thus $\|T\| \leq 1$. To see that $\|T\| = 1$, let $f\colon [0,1] \to \mathbb{C}$ be the constant function $f = 1$. Then $\|f\|_\infty = 1$ and

$$\begin{aligned}
\|Tf\|_\infty &= \sup\{|(Tf)(x)| \mid x \in [0,1]\} \\
&= \sup\left\{\left|\int_0^x \mathrm{d}y\right| \mid x \in [0,1]\right\} \\
&= \sup\{|x| \mid x \in [0,1]\} \\
&= \sup\{x \mid x \in [0,1]\} \\
&= 1.
\end{aligned}$$

□

23.4 Example of Linear Functional in $(C[a,b], \|\cdot\|_\infty)$

Proposition 23.4. *Consider $C[a,b]$ equipped with the supremum norm. Define a linear functional $\ell\colon C[a,b] \to \mathbb{R}$ by*

$$\ell(f) := f(a) - f(b).$$

for all $f \in C[a,b]$. Then ℓ is bounded. Moreover the set

$$\{f \in C[a,b] \mid f(a) = f(b)\}$$

is a closed subspace of $C[a,b]$.

Proof. Let $f \in C[a,b]$ such that $\|f\|_\infty \le 1$. Then

$$\begin{aligned} |\ell(f)| &= |f(a) - f(b)| \\ &\le |f(a)| + |f(b)| \\ &\le 1 + 1 \\ &= 2. \end{aligned}$$

Thus $\|\ell\| \le 2$. To see that $\|\ell\| = 2$, let $f\colon [a,b] \to \mathbb{C}$ be given by

$$f(x) = \frac{2}{b-a}(x-a) - 1$$

for all $x \in [a,b]$. So the graph of f is just the line segment from $(a,-1)$ to $(b,1)$. In particular, $\|f\|_\infty = 1$ and

$$\begin{aligned} |\ell(f)| &= |f(a) - f(b)| \\ &= |-1-1| \\ &= 2. \end{aligned}$$

The last part of the proposition follows from

$$\ker \ell = \{f \in C[a,b] \mid f(a) = f(b)\},$$

and $\ker \ell$ is a closed subspace since ℓ is a bounded linear operator. $\square$

23.5 Example of Linear Functional in $(C[a,b], \|\cdot\|_\infty)$ and Closed Subspace

Lemma 23.1. *Consider $C[a,b]$ equipped with the supremum norm. Let $[c,d] \subseteq [a,b]$ and define $\ell_{c,d}\colon C[a,b] \to \mathbb{C}$ by*

$$\ell_{c,d}(f) = \int_c^d f(t)\mathrm{d}t$$

for all $f \in C[a,b]$. Then $\ell_{c,d}$ is a bounded linear functional with $\|\ell_{c,d}\| = d - c$.

Proof. Linearity of $\ell_{c,d}$ follows from linearity of integration. So it suffices to check that $\ell_{c,d}$ is bounded. Let $f \in C[a,b]$ such that $\|f\|_\infty \le 1$. Then

$$\begin{aligned} |\ell_{c,d}(f)| &= \left|\int_c^d f(t)\mathrm{d}t\right| \\ &\le \int_c^d |f(t)|\mathrm{d}t \\ &\le \int_c^d \mathrm{d}t \\ &= d - c. \end{aligned}$$

Thus $\|\ell\| \le d - c$. To see that $\|\ell\| = d - c$, let $f\colon [a,b] \to \mathbb{C}$ be the constant function $f = 1$. Then $\|f\|_\infty = 1$ and

$$\begin{aligned} |\ell_{c,d}(f)| &= \left|\int_c^d f(t)\mathrm{d}t\right| \\ &= \left|\int_c^d \mathrm{d}t\right| \\ &= |d - c| \\ &= d - c. \end{aligned}$$

□

Proposition 23.5. *Consider $C[-1,1]$ equipped with the supremum norm. Let $\mathcal{Y}$ be the subset of $C[-1,1]$ consisting of all functions $g \in C[-1,1]$ such that*

$$\int_{-1}^{0} g(x)\mathrm{d}x = \int_{0}^{1} g(x)\mathrm{d}x = 0.$$

Then $\mathcal{Y}$ is a closed subspace.

Proof. Note that $\mathcal{Y} = \ker \ell_{-1,0} \cap \ker \ell_{0,1}$ is an intersection of two closed subspaces (since $\ell_{-1,0}$ and $\ell_{0,1}$ are bounded linear functionals by Lemma (23.1)). Thus $\mathcal{Y}$ is a closed subspace. □

Proposition 23.6. *With the notation as in Proposition (23.5) above, let $h \in C[-1,1]$ be given by*

$$h(x) = 2x$$

for all $x \in [-1,1]$. Then there does not exist a $g \in \mathcal{Y}$ such that $\|g-h\|_\infty = d(h,\mathcal{Y})$.

Proof.

Step 1: We will first show that $d(h,\mathcal{Y}) = 1$. To prove $d(h,\mathcal{Y}) \geq 1$, assume for a contradiction that $d(h,\mathcal{Y}) < 1$. Choose $\varepsilon > 0$ and $g \in \mathcal{Y}$ such that

$$\|g-h\|_\infty < 1-\varepsilon.$$

Write g in terms of its real and imaginary parts, say $g = u + iv$. Then

$$\begin{aligned} 0 &= \int_{-1}^{0} g(x)\mathrm{d}x \\ &= \int_{-1}^{0} u(x)\mathrm{d}x + i\int_{-1}^{0} v(x)\mathrm{d}x \end{aligned}$$

implies $\int_{-1}^{0} u(x)\mathrm{d}x = 0$ and $\int_{-1}^{0} v(x)\mathrm{d}x = 0$. Similarly,

$$\begin{aligned} 0 &= \int_{0}^{1} g(x)\mathrm{d}x \\ &= \int_{0}^{1} u(x)\mathrm{d}x + i\int_{0}^{1} v(x)\mathrm{d}x \end{aligned}$$

implies $\int_{0}^{1} u(x)\mathrm{d}x = 0$ and $\int_{0}^{1} v(x)\mathrm{d}x = 0$. Moreover, we have

$$\begin{aligned} 1-\varepsilon &> \|g-h\|_\infty \\ &= \sup_{x\in[-1,1]} \sqrt{(u(x)-h(x))^2 + v(x)^2} \\ &\geq \sup_{x\in[-1,1]} \sqrt{(u(x)-h(x))^2} \\ &= \|u-h\|_\infty. \end{aligned}$$

Therefore $u \in \mathcal{Y}$, $\|u-h\|_\infty < 1-\varepsilon$, and u is a real-valued function. Since $\|u-h\|_\infty < 1-\varepsilon$, $h(x) = 2x$ for all $x \in [-1,0]$, and both u and h are real-valued functions, we have

$$u(x) \leq 2x + 1 - \varepsilon$$

for all $x \in [-1,0]$. This implies

$$\begin{aligned} 0 &= \int_{-1}^{0} u(x)\mathrm{d}x \\ &\leq \int_{-1}^{0} (2x+1-\varepsilon)\mathrm{d}x \\ &= (x^2 + x - \varepsilon x)\big|_{-1}^{0} \\ &= \varepsilon \\ &> 0, \end{aligned}$$

which gives us our desired contradiction. Therefore $d(h, \mathcal{Y}) \geq 1$.

Now we will show that $d(h, \mathcal{Y}) \leq 1$. Let $t \in (0,1]$ and define $g_t \colon [-1,0] \to \mathbb{R}$ by the formula

$$g_t(x) = \begin{cases} 2x + 1 + t & \text{if } -1 \leq x \leq \frac{-2t}{1+t} \\ -\frac{(t-1)^2}{2t}x & \text{if } \frac{-2t}{1+t} \leq x \leq 0. \end{cases}$$

Extend g_t to all of $[-1,1]$ by the formula

$$g_t(x) = g_t(-x)$$

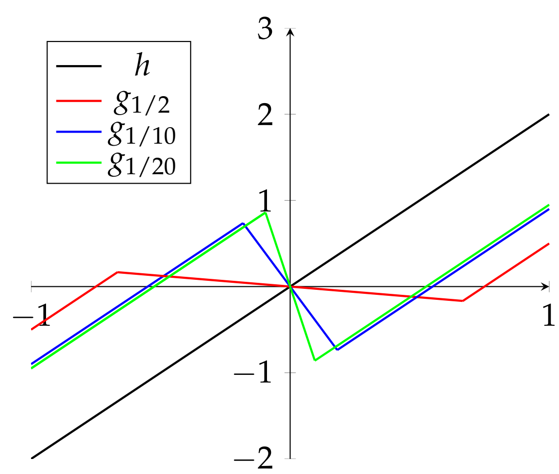
for all $x \in [0,1]$. So g_t is an odd function. Moreover g_t is continous since each segment of g_t is linear and since they agree on their boundaries:

$$\begin{aligned} 2\left(\frac{-2t}{1+t}\right) + 1 + t &= \frac{-4t}{1+t} + \frac{(1+t)^2}{1+t} \\ &= \frac{t^2 - 2t + 1}{1+t} \\ &= \frac{(t-1)^2}{1+t} \\ &= -\frac{(t-1)^2}{2t}\left(\frac{-2t}{1+t}\right) \end{aligned}$$

and

$$\begin{aligned} -\frac{(t-1)^2}{2t} \cdot 0 &= 0 \\ &= \frac{(t-1)^2}{2t} \cdot 0. \end{aligned}$$

The image below gives the graphs for h, $g_{1/2}$, and $g_{1/10}$:



Now observe that

$$\begin{aligned} \int_{-1}^{0} g_t(x)\mathrm{d}x &= \int_{-1}^{\frac{-2t}{1+t}} (2x + 1 + t)\mathrm{d}x + \int_{\frac{-2t}{1+t}}^{0} -\frac{(t-1)^2}{2t}x\mathrm{d}x \\ &= (x^2 + x + tx)|_{-1}^{\frac{-2t}{1+t}} + \left(\frac{-(t-1)^2}{4t}x^2\right)|_{\frac{-2t}{1+t}}^{0} \\ &= \left(\frac{2t}{1+t}\right)^2 + \left(\frac{-2t}{1+t}\right) + t\left(\frac{-2t}{1+t}\right) - (1 - 1 - t) + \frac{(t-1)^2}{4t}\left(\frac{2t}{1+t}\right)^2 \\ &= \left(\frac{(t-1)^2}{4t} + 1\right)\left(\frac{2t}{1+t}\right)^2 + (1+t)\left(\frac{-2t}{1+t}\right) + t \\ &= \frac{(t+1)^2}{4t}\frac{4t^2}{(1+t)^2} - t \\ &= t - t \\ &= 0. \end{aligned}$$

Therefore $g_t \in \mathcal{Y}$ for all $t \in (0,1]$. Moreover, by construction we have

$$\|g_t - h\|_\infty = 1 + t$$

for all $t \in (0,1]$. This implies $d(h, \mathcal{Y}) \leq 1$.

Step 2: We claim that there does not exist a $g \in \mathcal{Y}$ such that $\|g - h\|_\infty = 1$. Indeed, assume for a contradiction there does exist a $g \in \mathcal{Y}$ such that $\|g - h\|_\infty = 1$. Choose such a $g \in \mathcal{Y}$. We may assume that g is real-valued: if g is not real-valued, then we pass to its real-valued part u and as argued above we obtain $u \in \mathcal{Y}$ and

$$\begin{aligned}
1 &= \|g - h\|_\infty \\
&= \|u - h\|_\infty \\
&\geq 1.
\end{aligned}$$

Since g is real-valued and $\|g - h\|_\infty = 1$, we have

$$2x - 1 \leq g(x) \leq 2x + 1$$

for all $x \in [-1,1]$. Since g is continous, we cannot have

$$g(x) = \begin{cases} 2x + 1 & \text{for all } x \in (-1,0) \\ 2x - 1 & \text{for all } x \in (0,1). \end{cases}$$

Assume $g(x) \neq 2x - 1$ on the interval $(0,1)$. Choose $c \in (0,1)$ such that $g(c) \neq 2c - 1$. Since g is continuous and since $g(c) > 2c - 1$, there exists $\varepsilon > 0$ and $\delta > 0$ such that

$$g(x) > 2x - 1 + \varepsilon$$

for all $x \in (c - \delta, c + \delta)$. Choose such ε and δ so that $(c - \delta, c + \delta) \subset (0,1)$. Then

$$\begin{aligned}
0 &= \int_0^1 g(x)\mathrm{d}x \\
&= \int_0^1 g(x)\mathrm{d}x + \int_{c-\delta}^{c+\delta} g(x)\mathrm{d}x + \int_{c+\delta}^1 g(x)\mathrm{d}x \\
&> \int_0^{c-\delta} (2x - 1)\mathrm{d}x + \int_{c-\delta}^{c+\delta} (2x - 1 + \varepsilon)\mathrm{d}x + \int_{c+\delta}^1 (2x - 1)\mathrm{d}x \\
&= \int_0^1 (2x - 1)\mathrm{d}x + \int_{c-\delta}^{c+\delta} \varepsilon\mathrm{d}x \\
&= (x^2 - x)|_0^1 + \varepsilon x|_{c-\delta}^{c+\delta} \\
&= 2\varepsilon\delta \\
&> 0
\end{aligned}$$

gives us a contradiction.

Thus $g(x) \neq 2x + 1$ on the interval $(-1,0)$. Choose $c \in (-1,0)$ such that $g(c) \neq 2c + 1$. Then by a similar argument as above, we have

$$\begin{aligned}
0 &= \int_{-1}^0 g(x)\mathrm{d}x \\
&< \int_{-1}^0 (2x + 1)\mathrm{d}x - \int_{c-\delta}^{c+\delta} \varepsilon\mathrm{d}x \\
&= -2\varepsilon\delta \\
&< 0,
\end{aligned}$$

which also gives us a contradiction. Therefore there does not exist a $g \in \mathcal{Y}$ such that $\|g - h\|_\infty = 1$. $\square$

23.6 Closed Subspaces of $\mathcal{X}^*$ and $\mathcal{X}$

Definition 23.1. Let $\mathcal{X}$ be a normed linear space. For a set $A \subseteq \mathcal{X}$ we define $A^\perp$ to be the subset of $\mathcal{X}^*$ consisting of all $\ell \in \mathcal{X}^*$ such that $\ell(a) = 0$ for all $a \in A$. Similarly, for a set $M \subseteq \mathcal{X}^*$ we define $M_\perp$ to be the subset of $\mathcal{X}$ consisting of all vectors $x \in \mathcal{X}$ such that $\ell(x) = 0$ for all $\ell \in M$.

Proposition 23.7. *Let $\mathcal{X}$ be a normed linear space, let A be a subset of $\mathcal{X}$, and let M be a subset of $\mathcal{X}^*$. Then $A^\perp$ and $M_\perp$ are closed subspaces of $\mathcal{X}^*$ and $\mathcal{X}$ respectively.*

Proof. Let $x \in \mathcal{X}$. Define $\widehat{x}\colon \mathcal{X}^* \to \mathbb{C}$ by

$$\widehat{x}(\ell) = \ell(x)$$

for all $\ell \in \mathcal{X}^*$. We claim that $\widehat{x}$ is a bounded linear functional. To see that $\widehat{x}$ is linear, let $\ell, \ell' \in \mathcal{X}^*$ and let $\lambda, \lambda' \in \mathbb{C}$. Then

$$\begin{aligned}
\widehat{x}(\lambda\ell + \lambda'\ell') &= (\lambda\ell + \lambda'\ell')(x) \\
&= \lambda\ell(x) + \lambda'\ell'(x) \\
&= \lambda\widehat{x}(\ell) + \lambda'\widehat{x}(\ell').
\end{aligned}$$

To see that $\widehat{x}$ is bounded, let $\ell \in \mathcal{X}^*$. Then

$$\begin{aligned}
|\widehat{x}(\ell)| &= |\ell(x)| \\
&\leq \|x\|\|\ell\|.
\end{aligned}$$

Therefore $\widehat{x}$ is a bounded linear functional. In particular $\ker\widehat{x}$ is a closed subspace. Thus

$$A^\perp = \bigcap_{a \in A} \ker\widehat{a} \quad \text{and} \quad M_\perp = \bigcap_{\ell \in M} \ker\ell$$

are closed subspaces since an arbitrary intersection of closed subspaces is a closed subspace. □

Proposition 23.8. *Let $\mathcal{X}$ be a normed linear space, let A be a subset of $\mathcal{X}$, and let M be a subset of $\mathcal{X}^*$. Then $\overline{\text{span}}(A) \subseteq (A^\perp)_\perp$ and $\overline{\text{span}}(M) \subseteq (M_\perp)^\perp$.*

Proof. Proposition (23.7) implies $(A^\perp)_\perp$ and $(M_\perp)^\perp$ are closed subspace. Thus, it suffices to show

$$\text{span}(A) \subseteq (A^\perp)_\perp \quad \text{and} \quad \text{span}(M) \subseteq (M_\perp)^\perp.$$

First we show the former. Let $\lambda_1 a_1 + \cdots + \lambda_n a_n \in \text{span}(A)$ and let $\ell \in A^\perp$. Then since $\ell(a) = 0$ for all $a \in A$, we have

$$\begin{aligned}
\ell(\lambda_1 a_1 + \cdots + \lambda_n a_n) &= \lambda_1\ell(a_1) + \cdots + \lambda_n\ell(a_n) \\
&= \lambda_1 \cdot 0 + \cdots + \lambda_n \cdot 0 \\
&= 0.
\end{aligned}$$

Since ℓ was arbitrary, this implies $\lambda_1 a_1 + \cdots + \lambda_n a_n \in (A^\perp)_\perp$, and hence $\text{span}(A) \subseteq (A^\perp)_\perp$.

Now we show the latter. Let $\lambda_1\ell_1 + \cdots + \lambda_n\ell_n \in \text{span}(M)$ and let $x \in M_\perp$. Then since $\ell(x) = 0$ for all $\ell \in M$, we have

$$\begin{aligned}
(\lambda_1\ell_1 + \cdots + \lambda_n\ell_n)(x) &= \lambda_1\ell_1(x) + \cdots + \lambda_n\ell_n(x) \\
&= \lambda_1 \cdot 0 + \cdots + \lambda_n \cdot 0 \\
&= 0.
\end{aligned}$$

Since x was arbitrary, this implies $\lambda_1\ell_1 + \cdots + \lambda_n\ell_n \in (M_\perp)^\perp$, and hence $\text{span}(M) \subseteq (M_\perp)^\perp$. □

23.7 $(\ell^1)^*$ is isometrically isomorphic to ℓ^∞.

Proposition 23.9. *$(\ell^1)^*$ is isometrically isomorphic to ℓ^∞.*

Proof. For each $n \in \mathbb{N}$, let e^n denote the sequence with entry 1 in the nth component and entry 0 everywhere else. Define $\Phi\colon (\ell^1)^* \to \ell^\infty$ by

$$\Phi(\psi) = (\psi(e^n))$$

for all $\psi \in (\ell^1)^*$. Note that for any $\psi \in (\ell^1)^*$, we have $|(\psi(e^n))| \leq \|\psi\|$, and therefore $(\psi(e^n)) \in \ell^\infty$. We claim that $\|\psi\| = \|\Phi(\psi)\|_\infty$. Indeed,

$$\begin{aligned}
\|\Phi(\psi)\|_\infty &= \sup\{|\psi(e^n)| \mid n \in \mathbb{N}\} \\
&\leq \sup\left\{\left|\psi\left(\sum_{n=1}^{\infty} a_n e^n\right)\right| \mid \sum_{n=1}^{\infty} |a_n| \leq 1\right\} \\
&= \|\psi\|.
\end{aligned}$$

To prove the reverse inequality assume for a contradiction that $\|\psi\| > \|\Phi(\psi)\|_\infty$. Choose $\varepsilon > 0$ and $\sum_{n=1}^\infty a_n e^n \in \ell^1$ such that $\sum_{n=1}^\infty |a_n| \leq 1$ and

$$\left|\psi\left(\sum_{n=1}^{\infty} a_n e^n\right)\right| > \|\Phi(\psi)\|_\infty + \varepsilon. \tag{50}$$

Choose $N \in \mathbb{N}$ such that $\sum_{n=N}^\infty |a_n| < \varepsilon/\|\psi\|$ (we can find such an N since $\sum_{n=1}^\infty |a_n| < \infty$). Then

$$\begin{aligned}
\left|\psi\left(\sum_{n=1}^{\infty} a_n e^n\right)\right| &= \left|\psi\left(\sum_{n=1}^{N} a_n e^n + \sum_{n=N+1}^{\infty} a_n e^n\right)\right| \\
&= \left|\psi\left(\sum_{n=1}^{N} a_n e^n\right) + \psi\left(\sum_{n=N+1}^{\infty} a_n e^n\right)\right| \\
&= \left|\sum_{n=1}^{N} a_n \psi(e^n) + \psi\left(\sum_{n=N+1}^{\infty} a_n e^n\right)\right| \\
&\leq \left|\sum_{n=1}^{N} a_n \psi(e^n)\right| + \left|\psi\left(\sum_{n=N+1}^{\infty} a_n e^n\right)\right| \\
&\leq \sum_{n=1}^{N} |a_n||\psi(e^n)| + \|\psi\| \sum_{n=N+1}^{\infty} |a_n| \\
&< \|\Phi(\psi)\|_\infty \sum_{n=1}^{N} |a_n| + \|\psi\| \cdot \frac{\varepsilon}{\|\psi\|} \\
&\leq \|\Phi(\psi)\|_\infty + \varepsilon.
\end{aligned}$$

This contradicts (**??**).

Next we show Φ is linear. Let $\varphi, \psi \in (\ell^1)^*$ and let $\lambda, \mu \in \mathbb{C}$. Then

$$\begin{aligned}
\Phi(\lambda\varphi + \mu\psi) &= ((\lambda\varphi + \mu\psi)(e^n)) \\
&= \lambda(\varphi(e^n)) + \mu(\psi(e^n)) \\
&= \lambda\Phi(\varphi) + \mu\Phi(\psi).
\end{aligned}$$

Therefore Φ is an isometric embedding.

Now show that Φ is surjective, and hence an isometric isomorphism. Let $(a_n) \in \ell^\infty$, let $M = \sup\{|a_n|\}$, and let $E = \operatorname{span}\{e^n \mid n \in \mathbb{N}\}$. Define $\varphi\colon E \to \mathbb{C}$ to be the unique linear map such that

$$\varphi(e^n) = a_n$$

for all $n \in \mathbb{N}$. Let $x = x_{n_1}e^{n_1} + \cdots + x_{n_k}e^{n_k} \in E$ such that $|x_{n_1}| + \cdots + |x_{n_k}| \leq 1$. Then

$$\begin{aligned}
|\varphi(x_{n_1}e^{n_1} + \cdots + x_{n_k}e^{n_k})| &= |x_{n_1}\varphi(e^{n_1}) + \cdots + x_{n_k}\varphi(e^{n_k})| \\
&= |x_{n_1}a_{n_1} + \cdots + x_{n_k}a_{n_k}| \\
&\leq |x_{n_1}||a_{n_1}| + \cdots + |x_{n_k}||a_{n_k}| \\
&\leq |x_{n_1}|M + \cdots + |x_{n_k}||M \\
&= (|x_{n_1}| + \cdots + |x_{n_k}||)M \\
&\leq M
\end{aligned}$$

It follows that φ is bounded. By the Hahn-Banach Theorem, there exists a bounded linear functional $\widetilde{\varphi}$ defined on all of ℓ^1 such that $\widetilde{\varphi}|_E = \varphi$ and $\|\widetilde{\varphi}\| = \|\varphi\|$. Choose such a $\widetilde{\varphi} \in (\ell^1)^*$. Then clearly $\Phi(\widetilde{\varphi}) = (a_n)$. Therefore Φ is surjective, and hence an isometric isomorphism. □

Appendix

Problem 1

Proposition 23.10. *Let A be a non-emtpy set of real numbers which is bounded above and let λ be any non-negative real number. Then*

$$\sup(\lambda A) = \lambda \sup(A). \tag{51}$$

Proof. If $\lambda = 0$, then (51) is obvious, so assume $\lambda > 0$. Let α denote $\sup(A)$. Choose any element in λA, say λa where $a \in A$. Then since $a \leq \alpha$ and λ is non-negative, we have $\lambda a \leq \lambda\alpha$. This implies

$$\sup(\lambda A) \leq \lambda \sup(A).$$

For the reverse direction, observe that

$$\begin{aligned}
\sup(A) &= \sup(\lambda^{-1}\lambda A) \\
&\leq \lambda^{-1} \sup(\lambda A),
\end{aligned}$$

and this implies

$$\sup(\lambda A) \geq \lambda \sup(A).$$

□

Proposition 23.11. *Let A and B be non-empty sets of non-negative real numbers both of which are bounded above. Then*

$$\sup(A + B) = \sup(A) + \sup(B). \tag{52}$$

Proof. Let α denote $\sup(A)$, let β denote $\sup(B)$, and let $a + b$ be an arbitrary element in $A + B$. Then $a \leq \alpha$ and $b \leq \beta$ implies $a + b \leq \alpha + \beta$. Therefore

$$\sup(A + B) \leq \sup(A) + \sup(B). \tag{53}$$

To show the reverse inequality, we assume (for a contradiction) that the inequality (53) is strict

$$\sup(A + B) < \sup(A) + \sup(B).$$

Choose $\varepsilon > 0$ such that

$$\sup(A + B) < \sup(A) + \sup(B) - \varepsilon. \tag{54}$$

Choose $a \in A$ and $b \in B$ such that $a > \alpha - \varepsilon/2$ and $b > \beta - \varepsilon/2$. Then

$$\begin{aligned}
a + b &> \alpha - \frac{\varepsilon}{2} + \beta - \frac{\varepsilon}{2} \\
&= \alpha + \beta - \varepsilon.
\end{aligned}$$

But this contradicts (54). Therefore

$$\sup(A + B) \geq \sup(A) + \sup(B).$$

□

Appendix

24 Completion

Let $\mathcal{V}$ be an inner-product space. In this section, we describe a procedure called **completion** which constructs a Hibert space $\mathscr{C}_{\mathcal{V}}/\mathscr{C}_{\mathcal{V}}^0$ and an injective linear map $\iota\colon \mathcal{V} \to \mathscr{C}_{\mathcal{V}}/\mathscr{C}_{\mathcal{V}}^0$ such that ι respects then inner-product structure on both $\mathcal{V}$ and $\mathscr{C}_{\mathcal{V}}/\mathscr{C}_{\mathcal{V}}^0$ (namely we will show that ι is an isometry) and such that $\iota(\mathcal{V})$ is dense in $\mathscr{C}_{\mathcal{V}}/\mathscr{C}_{\mathcal{V}}^0$.

24.1 Constructing Completions

Let $\mathscr{C}_{\mathcal{V}}$ denote the set of all Cauchy sequences in $\mathcal{V}$. We can give $\mathscr{C}_{\mathcal{V}}$ the structure of a $\mathbb{C}$-vector space as follows: let $(x_n), (y_n) \in \mathscr{C}_{\mathcal{V}}$ and let $\lambda, \mu \in \mathbb{C}$. Then we define

$$a(x_n) + b(y_n) := (ax_n + by_n). \tag{55}$$

Scalar multiplication and addition as in (106) are easily seen to give $\mathscr{C}_{\mathcal{V}}$ the structure of a $\mathbb{C}$-vector space.

24.1.1 Pseudo Inner-Product

A natural contender for an inner-product on $\mathscr{C}_{\mathcal{V}}$ is the map $\langle \cdot, \cdot \rangle \colon \mathscr{C}_{\mathcal{V}} \times \mathscr{C}_{\mathcal{V}} \to \mathbb{C}$ defined by

$$\langle (x_n), (y_n) \rangle := \lim_{n\to\infty} \langle x_n, y_n \rangle \tag{56}$$

for all $(x_n), (y_n) \in \mathscr{C}_{\mathcal{V}}$. In fact, (107) will not be an inner-product, but rather a *pseudo* inner-product. Before we explain this however, let us first show that the righthand side of (107) converges in $\mathbb{C}$

Lemma 24.1. *Let (x_n) be a Cauchy sequence in $\mathcal{V}$. Then (x_n) is bounded.*

Proof. Let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n, m \geq N$ implies $\|x_n - x_m\| < \varepsilon$. Thus, fixing $m \in \mathbb{N}$, we see that $n \geq N$ implies

$$\|x_n\| < \|x_m\| + \varepsilon.$$

Now we let

$$M = \max\{\|x_1\|, \|x_2\|, \dots, \|x_m\| + \varepsilon\}.$$

Then M is a bound for (x_n). □

Proposition 24.1. *Let (x_n) and (y_n) be Cauchy sequences of vectors in $\mathcal{V}$. Then $(\langle x_n, y_n \rangle)$ is a Cauchy sequence of complex numbers in $\mathbb{C}$. In particular, (107) converges in $\mathbb{C}$.*

Proof. Let $\varepsilon > 0$. Choose M_x and M_y such that $\|x_n\| < M_x$ and $\|y_n\| < M_y$ for all $n \in \mathbb{N}$. We can do this by Lemma (24.1). Next, choose $N \in \mathbb{N}$ such that $n, m \geq N$ implies $\|x_n - x_m\| < \frac{\varepsilon}{2M_y}$ and $\|y_n - y_m\| < \frac{\varepsilon}{2M_x}$. Then $n, m \geq M$ implies

$$\begin{aligned}
|\langle x_n, y_n \rangle - \langle x_m, y_m \rangle| &= |\langle x_n, y_n \rangle - \langle x_m, y_n \rangle + \langle x_m, y_n \rangle - \langle x_m, y_m \rangle| \\
&= |\langle x_n - x_m, y_n \rangle + \langle x_m, y_n - y_m \rangle| \\
&\leq |\langle x_n - x_m, y_n \rangle| + |\langle x_m, y_n - y_m \rangle| \\
&\leq \|x_n - x_m\| \|y_n\| + \|x_m\| \|y_n - y_m\| \\
&\leq \|x_n - x_m\| M_y + M_x \|y_n - y_m\| \\
&< \varepsilon.
\end{aligned}$$

This implies $(\langle x_n, y_n \rangle)$ is a Cauchy sequence of complex numbers in $\mathbb{C}$. The latter statement in the proposition follows from the fact that $\mathbb{C}$ is complete. □

24.1.2 Quotienting Out To get an Inner-Product

As mentioned above, (107) is not an inner-product. It is what's called a pseudo inner-product:

Definition 24.1. Let V be a vector space over $\mathbb{C}$. A map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ is called a **pseudo inner-product on** V if it satisfies the following properties:

1. Linearity in the first argument: $\langle x, z\rangle + \langle y, z\rangle = \langle x+y, z\rangle$ and $\langle \lambda x, y\rangle = \lambda \langle x, y\rangle$ for all $x, y, z \in V$ and $\lambda \in \mathbb{C}$.

2. Conjugate symmetric: $\langle x, y\rangle = \overline{\langle y, x\rangle}$ for all $x, y \in V$.

3. Pseudo positive definiteness: $\langle x, x\rangle \geq 0$ for all nonzero $x \in V$.

A vector space equipped with a pseudo inner-product is called a **pseudo inner-product space**.

To see why (107) is a pseudo inner-product, note that linearity in the first argument and conjugate symmetry are clear. What makes (107) a pseudo inner-product and not an inner-product is that we have pseudo positive definiteness:

$$
\begin{aligned}
\langle (x_n), (x_n) \rangle &= \lim_{n\to\infty} \langle x_n, x_n \rangle \\
&= \lim_{n\to\infty} \|x_n\|^2 \\
&\geq 0.
\end{aligned}
$$

In particular, we may have $\langle (x_n), (x_n)\rangle = 0$ with $(x_n) \neq 0$. To remedy this situation, we define

$$
\mathscr{C}^0_{\mathcal{V}} := \{(x_n) \in \mathscr{C}_{\mathcal{V}} \mid x_n \to 0\}.
$$

Then $\mathscr{C}^0_{\mathcal{V}}$ is a subspace of $\mathscr{C}_{\mathcal{V}}$ (if $\lambda \in \mathbb{C}$ and $(x_n), (y_n) \in \mathscr{C}^0_{\mathcal{V}}$, then $(\lambda x_n + y_n) \to 0$ and hence $(\lambda x_n + y_n) \in \mathscr{C}^0_{\mathcal{V}}$). Therefore we obtain a quotient space $\mathscr{C}_{\mathcal{V}}/\mathscr{C}^0_{\mathcal{V}}$. Now we claim that the pseudo inner-product (107) induces a genuine inner-product, which we denote again by $\langle \cdot, \cdot \rangle$, on $\mathscr{C}_{\mathcal{V}}/\mathscr{C}^0_{\mathcal{V}}$, defined by

$$
\langle \overline{(x_n)}, \overline{(y_n)} \rangle := \lim_{n\to\infty} \langle x_n, y_n \rangle. \tag{57}
$$

for all $\overline{(x_n)}$[11] and $\overline{(y_n)}$ in $\mathscr{C}_{\mathcal{V}}/\mathscr{C}^0_{\mathcal{V}}$. We need to be sure that (108) is well-defined. Let (x'_n) and (y'_n) be different representatives of the cosets $\overline{(x_n)}$ and $\overline{(y_n)}$ respectively (so $x_n - x'_n \to 0$ and $y_n - y'_n \to 0$). Then

$$
\begin{aligned}
\langle \overline{(x'_n)}, \overline{(y'_n)} \rangle &= \lim_{n\to\infty} \langle x'_n, y'_n \rangle \\
&= \lim_{n\to\infty} \langle x'_n, y'_n \rangle + \lim_{n\to\infty} \langle x_n - x'_n, y'_n \rangle + \lim_{n\to\infty} \langle x_n, y_n - y'_n \rangle \\
&= \lim_{n\to\infty} (\langle x'_n, y'_n \rangle + \langle x_n - x'_n, y'_n \rangle + \langle x_n, y_n - y'_n \rangle) \\
&= \lim_{n\to\infty} (\langle x_n, y'_n \rangle + \langle x_n, y_n - y'_n \rangle) \\
&= \lim_{n\to\infty} \langle x_n, y_n \rangle \\
&= \langle \overline{(x_n)}, \overline{(y_n)} \rangle.
\end{aligned}
$$

Thus (108) is well-defined (meaning it is independent of the choice of representatives of cosets).

Now linearity in the first argument of (108) and conjugate symmetry of (108) are clear. This time however, we have positive definiteness: if $\overline{(x_n)} \in \mathscr{C}_{\mathcal{V}}/\mathscr{C}^0_{\mathcal{V}}$ such that $\langle \overline{(x_n)}, \overline{(x_n)} \rangle = 0$, then

$$
\begin{aligned}
\lim_{n\to\infty} \langle x_n, x_n \rangle &= \langle \overline{(x_n)}, \overline{(x_n)} \rangle \\
&= 0
\end{aligned}
$$

implies $x_n \to 0$, which implies $\overline{(x_n)} = 0$ in $\mathscr{C}_{\mathcal{V}}/\mathscr{C}^0_{\mathcal{V}}$.

[11] When we write $\overline{(x_n)}$ for a coset in $\mathscr{C}_{\mathcal{V}}/\mathscr{C}^0_{\mathcal{V}}$, then it is implicitly understood that (x_n) is an element $\mathscr{C}_{\mathcal{V}}$ which represents the coset $\overline{(x_n)}$ in $\mathscr{C}_{\mathcal{V}}/\mathscr{C}^0_{\mathcal{V}}$.

24.1.3 The map $\iota\colon \mathcal{V} \to \mathscr{C}_{\mathcal{V}}/\mathscr{C}_{\mathcal{V}}^0$

Let $\iota\colon \mathcal{V} \to \mathscr{C}_{\mathcal{V}}/\mathscr{C}_{\mathcal{V}}^0$ be defined by

$$\iota(x) = \overline{(x)}_{n\in\mathbb{N}}$$

for all $x \in \mathcal{V}$, where $(x)_{n\in\mathbb{N}}$ is a constant sequence in $\mathscr{C}_{\mathcal{V}}$.

Definition 24.2. An **isometry** between inner-product spaces $\mathcal{V}_1$ and $\mathcal{V}_2$ is an operator $T\colon \mathcal{V}_1 \to \mathcal{V}_2$ such that

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

for all $x, y \in \mathcal{V}_1$.

Remark 43. Note that an isometry is automatically injective. Indeed, let $x \in \mathrm{Ker}(T)$. Then

$$\begin{aligned}\langle x, y\rangle &= \langle Tx, Ty\rangle \\ &= \langle 0, Ty\rangle \\ &= 0\end{aligned}$$

for all $y \in \mathcal{V}_1$. It follows that $x = 0$.

Proposition 24.2. *The map $\iota\colon \mathcal{V} \to \mathscr{C}_{\mathcal{V}}/\mathscr{C}_{\mathcal{V}}^0$ is an isometry.*

Proof. Linearity of ι is clear. Let $x, y \in \mathcal{V}$. Then

$$\begin{aligned}\langle\langle \iota(x), \iota(y)\rangle &:= \lim_{n\to\infty} \langle x, y\rangle \\ &= \langle x, y\rangle.\end{aligned}$$

Thus ι is an isometry. □

Proposition 24.3. *The image of $\mathcal{V}$ under ι is dense in $\mathscr{C}_{\mathcal{V}}/\mathscr{C}_{\mathcal{V}}^0$. In other words, the closure of $\iota(\mathcal{V})$ in $\mathscr{C}_{\mathcal{V}}/\mathscr{C}_{\mathcal{V}}^0$ is all of $\mathscr{C}_{\mathcal{V}}/\mathscr{C}_{\mathcal{V}}^0$.*

Proof. Let $\overline{(x_n)}$ be a coset in $\mathscr{C}_{\mathcal{V}}/\mathscr{C}_{\mathcal{V}}^0$. To show that the closure of $\iota(\mathcal{V})$ is is all of $\mathscr{C}_{\mathcal{V}}/\mathscr{C}_{\mathcal{V}}^0$, we construct a sequence of cosets in $\iota(\mathcal{V})$ which converges to $\overline{(x_n)}$. Let $\varepsilon > 0$ and choose $N \in \mathbb{N}$ such that $n, m \geq N$ implies

$$\|x_n - x_m\| < \varepsilon/2.$$

Then $n, m \geq N$ implies

$$\begin{aligned}\|\iota(x_m) - \overline{(x_n)}\| &= \lim_{n\to\infty} \|x_m - x_n\| \\ &< \lim_{n\to\infty} \varepsilon \\ &= \varepsilon.\end{aligned}$$

Thus, $(\iota(x_m))$ is a sequence of cosets in $\iota(\mathcal{V})$ which converges to $\overline{(x_n)}$. □

24.1.4 $\mathscr{C}_{\mathcal{V}}/\mathscr{C}_{\mathcal{V}}^0$ is a Hilbert Space

Proposition 24.4. *$\mathscr{C}_{\mathcal{V}}/\mathscr{C}_{\mathcal{V}}^0$ is a Hilbert space.*

Proof. Let $(\overline{x}^n)$ be a Cauchy sequence of cosets in $\mathscr{C}_{\mathcal{V}}/\mathscr{C}_{\mathcal{V}}^0$ where

$$\overline{x}^n = \overline{(x_k^n)}_{k\in\mathbb{N}}$$

for each $n \in \mathbb{N}$. Throughout the remainder of this proof, let $\varepsilon > 0$.

Since each $x^n = (x_k^n)_{k\in\mathbb{N}}$ is a Cauchy sequence of elements in $\mathcal{V}$, there exists a $\pi(n) \in \mathbb{N}$ such that $k, l \geq \pi(n)$ implies

$$\|x_k^n - x_l^n\| < \frac{\varepsilon}{3}.$$

For each $n \in \mathbb{N}$, choose such $\pi(n) \in \mathbb{N}$ in such a way so $\pi(n) \geq \pi(m)$ whenever $n \geq m$.

Step 1: We show that the sequence $(x^n_{\pi(n)})$ of elements in $\mathcal{V}$ is a Cauchy sequence. Since $(\overline{x}^n)$ is a Cauchy sequence of cosets in $\mathscr{C}_{\mathcal{V}}/\mathscr{C}^0_{\mathcal{V}}$, there exists an $N \in \mathbb{N}$ such that $n, m \geq N$ implies

$$\|(\overline{x}^n) - (\overline{x}^m)\| = \lim_{k\to\infty} \|x^n_k - x^m_k\| < \frac{\varepsilon}{4}. \tag{58}$$

Choose such an $N \in \mathbb{N}$. It follows from (109) that for each $n \geq m \geq N$, there exists $\pi(n,m) \geq \pi(n)$ such that

$$\|x^n_k - x^m_k\| < \frac{\varepsilon}{3}$$

for all $k \geq \pi(n,m)$. Choose such $\pi(n,m)$ for each $n \geq m \geq N$. Then if $n \geq m \geq N$, we have

$$\begin{aligned}
\|x^n_{\pi(n)} - x^m_{\pi(m)}\| &= \|x^n_{\pi(n)} - x^n_{\pi(n,m)} + x^n_{\pi(n,m)} - x^m_{\pi(n,m)} + x^m_{\pi(n,m)} - x^m_{\pi(m)}\| \\
&\leq \|x^n_{\pi(n)} - x^n_{\pi(n,m)}\| + \|x^n_{\pi(n,m)} - x^m_{\pi(n,m)}\| + \|x^m_{\pi(n,m)} - x^m_{\pi(m)}\| \\
&< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 \\
&= \varepsilon.
\end{aligned}$$

Therefore $(x^n_{\pi(n)})$ is a Cauchy sequence of elements in $\mathcal{V}$ and hence represents a coset $\overline{(x^n_{\pi(n)})}$ in $\mathscr{C}_{\mathcal{V}}/\mathscr{C}^0_{\mathcal{V}}$.

Step 2: Let $x = (x^k_{\pi(k)})$[12]. We want to show that the sequence $(\overline{x}^n)$ of cosets in $\mathscr{C}_{\mathcal{V}}/\mathscr{C}^0_{\mathcal{V}}$ converges to the coset $\overline{x}$ in $\mathscr{C}_{\mathcal{V}}/\mathscr{C}^0_{\mathcal{V}}$. In particular, we need to find an $N \in \mathbb{N}$ such that $n \geq N$ implies

$$\|\overline{x}^n - \overline{x}\| = \lim_{k\to\infty} \|x^n_k - x^k_{\pi(k)}\| < \varepsilon$$

or in other words, $n \geq N$ implies

$$\|x^n_k - x^k_{\pi(k)}\| < \varepsilon$$

for all k sufficiently large.

Since x is a Cauchy sequence of elements in $\mathcal{V}$, there exists an $M \in \mathbb{N}$ such that $n, k \geq M$ implies

$$\|x^n_{\pi(n)} - x^k_{\pi(k)}\| < 2\varepsilon/3.$$

Choose such an $M \in \mathbb{N}$. Then $n \geq M$ implies

$$\begin{aligned}
\|x^n_k - x^k_{\pi(k)}\| &\leq \|x^n_k - x^n_{\pi(n)}\| + \|x^n_{\pi(n)} - x^k_{\pi(k)}\| \\
&< \frac{\varepsilon}{3} + \frac{2\varepsilon}{3} \\
&= \varepsilon.
\end{aligned}$$

for all $k \geq \max\{M, \pi(n)\}$. □

25 Normed Vector Spaces

Definition 25.1. Let V be a $\mathbb{C}$-vector space. A **norm** on V is a nonnegative-valued scalar function $\|\cdot\| \colon V \to [0,\infty)$ such that for all $\lambda \in \mathbb{C}$ and $u, v \in V$, we have

1. (Subadditivity) $\|u+v\| \leq \|u\| + \|v\|$,
2. (Absolutely Homogeneous) $\|\lambda v\| = |\lambda|\|v\|$,
3. (Positive-Definite) $\|v\| = 0$ if and only if $v = 0$.

We call the pair $(V, \|\cdot\|)$ a **normed vector space**.

[12]Note the change in index from n to k.

25.1 Bounded Linear Operators and Normed Vector Spaces

Definition 25.2. Let $\mathcal{V}$ and $\mathcal{W}$ be inner-product spaces. We define

$$\mathcal{B}(\mathcal{V},\mathcal{W}) := \{T \colon \mathcal{V} \to \mathcal{W} \mid T \text{ is a bounded linear operator}\}.$$

$\mathcal{B}(\mathcal{V},\mathcal{W})$ has the structure of a $\mathbb{C}$-vector space, where addition and scalar multiplication are defined by

$$(T+U)(x) = Tx + Ux \quad \text{and} \quad (\lambda T)(x) = T(\lambda x)$$

for all $T, U \in \mathcal{B}(\mathcal{V},\mathcal{W})$, $\lambda \in \mathbb{C}$, and $v \in \mathcal{V}$.

Proposition 25.1. *Let $\mathcal{V}$ and $\mathcal{W}$ be inner-product spaces. Then $(\mathcal{B}(\mathcal{V},\mathcal{W}), \|\cdot\|)$ is a normed vector space, where $\|\cdot\|$ is the map which sends a bounded linear operator T to its norm $\|T\|$.*

Proof. An easy exercise in linear algebra shows that $\mathcal{B}(\mathcal{V},\mathcal{W})$ has the structure of a $\mathbb{C}$-vector space, where addition and scalar multiplication are defined by

$$(T+U)(x) = T(x) + U(x) \quad \text{and} \quad (\lambda T)(x) = T(\lambda x)$$

for all $T, U \in \mathcal{B}(\mathcal{V},\mathcal{W})$, $\lambda \in \mathbb{C}$, and $v \in \mathcal{V}$. The details of this are left as an exercise. We are more interested in the fact that $\mathcal{B}(\mathcal{V},\mathcal{W})$ is a *normed* vector space. We just need to check that $\|\cdot\|$ satisfies the conditions laid out in Definition (25.1).

We first check for subadditivity. Let $T, U \in \mathcal{B}(\mathcal{V},\mathcal{W})$. Then

$$\begin{aligned}
\|(T+U)(x)\| &= \|Tx + Ux\| \\
&\leq \|Tx\| + \|Ux\| \\
&\leq \|T\|\|x\| + \|U\|\|x\| \\
&= (\|T\| + \|U\|)\|x\|
\end{aligned}$$

for all $x \in \mathcal{V}$. In particular, this implies $\|T+U\| \leq \|T\| + \|U\|$. Thus we have subadditivity.

Next we check that $\|\cdot\|$ is absolutely homogeneous. Let $T \in \mathcal{B}(\mathcal{V},\mathcal{W})$ and let $\lambda \in \mathbb{C}$. Then

$$\begin{aligned}
\|(\lambda T)(x)\| &= \|T(\lambda x)\| \\
&= \|\lambda Tx\| \\
&= |\lambda|\|Tx\|
\end{aligned}$$

for all $x \in \mathcal{V}$. In particular, this implies $\|\lambda T\| = |\lambda|\|T\|$. Thus $\|\cdot\|$ is absolutely homogeneous.

Finally we check for positive-definitiness. Let $T \in \mathcal{B}(\mathcal{V},\mathcal{W})$. Clearly $\|T\|$ is greater than or equal to 0 since it is the supremum of terms which are greater than or equal to 0. Suppose $\|T\| = 0$. Then

$$\begin{aligned}
\|Tx\| &\leq \|T\|\|x\| \\
&= 0 \cdot \|x\| \\
&= 0
\end{aligned}$$

for all $x \in \mathcal{V}$. In particular, this implies $Tx = 0$ for all $x \in \mathcal{V}$ (by positive-definiteness of the norm for $\mathcal{W}$). Therefore $T = 0$ since they agree on all $x \in \mathcal{V}$. □

25.2 Normed Vector Spaces Which Satisfy Parallelogram Law are Inner-Product Spaces

Proposition 25.2. *Let $(V, \|\cdot\|)$ be a normed vector space over $\mathbb{C}$ which satisfies the parallelogram law (6). Then the map $\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{C}$ defined by*

$$\langle x, y \rangle = \frac{1}{4}\left(\|x+y\|^2 + i\|x+iy\|^2 - \|x-y\|^2 - i\|x-iy\|^2\right) \tag{59}$$

for all $x, y \in V$ is an inner-product. Moreover, the norm induced by this inner-product is precisely $\|\cdot\|$. In other words, we have

$$\langle x, x \rangle = \|x\|^2$$

for all $x \in V$.

Proof. The most difficult part of this proof is showing that (59) is linear in the first argument. Before we do this, let us show that (59) is positive-definite and conjugate-symmetric.

For positive-definiteness, let $x \in V$. Then

$$\begin{aligned}
\langle x, x\rangle &= \frac{1}{4}\left(\|2x\|^2 + i\|x+ix\|^2 - i\|x-ix\|^2\right)\\
&= \frac{1}{4}\left(\|2x\|^2 + i((|1+i|^2 - |1-i|^2)\|x\|^2)\right)\\
&= \|x\|^2\\
&\geq 0,
\end{aligned}$$

with equality if and only if $x = 0$. Note that this also gives us $\langle x, x\rangle = \|x\|^2$ for all $x \in V$.

For conjugate-symmetry, let $x, y \in V$. Then

$$\begin{aligned}
\overline{\langle y, x\rangle} &= \frac{1}{4}\overline{(\|y+x\|^2 + i\|y+ix\|^2 - \|y-x\|^2 - i\|y-ix\|^2)}\\
&= \frac{1}{4}\left(\|y+x\|^2 - i\|y+ix\|^2 - \|y-x\|^2 + i\|y-ix\|^2\right)\\
&= \frac{1}{4}\left(\|x+y\|^2 - i\|i(x-iy)\|^2 - \|x-y\|^2 + i\|i(x+iy)\|^2\right)\\
&= \frac{1}{4}\left(\|x+y\|^2 - i\|x-iy\|^2 - \|x-y\|^2 + i\|x+iy\|^2\right)\\
&= \frac{1}{4}\left(\|x+y\|^2 + i\|x+iy\|^2 - \|x-y\|^2 - i\|x-iy\|^2\right)\\
&= \langle x, y\rangle
\end{aligned}$$

Now we come to the difficult part, namely showing that (59) is linear in the first argument. We do this in several steps:

Step 1: We show that (59) is additive in the first argument (i.e. $\langle x+z, y\rangle = \langle x, y\rangle + \langle z, y\rangle$ for all $x, y, z \in V$). Let $x, y, z \in V$. First note that by the parallelogram law (6), we have

$$\begin{aligned}
\|x+z+y\|^2 - \|x+z-y\|^2 &= 2\|x+y\|^2 + 2\|z\|^2 - \|x+y-z\|^2 - 2\|x-y\|^2 - 2\|z\|^2 + \|x-y-z\|^2\\
&= 2\|x+y\|^2 - 2\|x-y\|^2 - \|z-y-x\|^2 + \|z+y-x\|^2\\
&= 2\|x+y\|^2 - 2\|x-y\|^2 - 2\|z-y\|^2 - 2\|x\|^2 + \|z-y+x\|^2 + 2\|z+y\|^2 + 2\|x\|^2 - \|z+y+x\|^2\\
&= 2\|x+y\|^2 - 2\|x-y\|^2 + 2\|z+y\|^2 - 2\|z-y\|^2 + \|x+z-y\|^2 - \|x+z+y\|^2.
\end{aligned}$$

Adding $\|x+z-y\|^2 - \|x+z+y\|^2$ to both sides gives us

$$2(\|x+z+y\|^2 - \|x+z-y\|^2) = 2(\|x+y\|^2 - \|x-y\|^2 + \|z+y\|^2 - \|z-y\|^2),$$

and after cancelling 2 from both sides, we obtain

$$\|x+z+y\|^2 - \|x+z-y\|^2 = \|x+y\|^2 - \|x-y\|^2 + \|z+y\|^2 - \|z-y\|^2.$$

Therefore

$$\begin{aligned}
\langle x+z, y\rangle &= \frac{1}{4}\left(\|x+z+y\|^2 + i\|x+z+iy\|^2 - \|x+z-y\|^2 - i\|x+z-iy\|^2\right)\\
&= \frac{1}{4}\left(\|x+z+y\|^2 - \|x+z-y\|^2 + i(\|x+z+iy\|^2 - \|x+z-iy\|^2)\right)\\
&= \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2 + \|z+y\|^2 - \|z-y\|^2 + i(\|x+iy\|^2 - \|x-iy\|^2 + \|z+iy\|^2 - \|z-iy\|^2)\right)\\
&= \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2 + \|z+y\|^2 - \|z-y\|^2 + i(\|x+iy\|^2 - \|x-iy\|^2 + \|z+iy\|^2 - \|z-iy\|^2)\right)\\
&= \frac{1}{4}\left(\|x+y\|^2 + i\|x+iy\|^2 - \|x-y\|^2 - i\|x-iy\|^2 + \|z+y\|^2 + i\|z+iy\|^2 - \|z-y\|^2 - i\|z-iy\|^2\right)\\
&= \langle x, y\rangle + \langle z, y\rangle.
\end{aligned}$$

Thus we have additivity in the first argument.

Step 2: We show that (59) respects $\mathbb{Q}$-scaling in the first argument (i.e. $\frac{m}{n}\langle x,y\rangle = \langle \frac{m}{n}x, y\rangle$ for all rational numbers $\frac{m}{n} \in \mathbb{Q}$ and for all $x, y \in V$. Let $\frac{m}{n} \in \mathbb{Q}$ and let $x, y \in V$. Then since (59) is additive in the first argument and since V is a $\mathbb{C}$-vector space, we have

$$\begin{aligned}
\frac{m}{n}\langle x,y\rangle &= \frac{m}{n}\left\langle \frac{n}{n}x, y\right\rangle \\
&= \frac{mn}{n}\left\langle \frac{1}{n}x, y\right\rangle \\
&= m\left\langle \frac{1}{n}x, y\right\rangle \\
&= \left\langle \frac{m}{n}x, y\right\rangle.
\end{aligned}$$

Therefore (59) respects $\mathbb{Q}$-scaling in the first argument.

Step 3: We show that (59) respects $\mathbb{R}$-scaling in the first argument. First note that $y \in V$, the map $\langle \cdot, y\rangle \colon V \to \mathbb{C}$ is continuous. Let $x, y \in V$ and let $r \in \mathbb{R}$. Choose a sequence (r_n) of rational numbers such that $r_n \to r$ (we can do this since $\mathbb{Q}$ is dense in $\mathbb{R}$). Then we have

$$\begin{aligned}
\langle rx, y\rangle &= \lim_{n\to\infty}\langle r_n x, y\rangle \\
&= \lim_{n\to\infty} r_n\langle x,y\rangle \\
&= r\langle x,y\rangle.
\end{aligned}$$

Therefore (59) respects $\mathbb{R}$-scaling in the first component.
Step 4: We show that (59) respects $\mathbb{C}$-scaling in the first component. We first show that $\langle ix, y\rangle = i\langle x,y\rangle$ for all $x, y \in V$.

Let $x, y \in V$. Then

$$\begin{aligned}
\langle ix, y\rangle &= \frac{1}{4}\left(\|ix+y\|^2 + i\|ix+iy\|^2 - \|ix-y\|^2 - i\|ix-iy\|^2\right) \\
&= \frac{1}{4}\left(\|x-iy\|^2 + i\|x+y\|^2 - \|x+iy\|^2 - i\|x-y\|^2\right) \\
&= \frac{1}{4}\left(i\|x+y\|^2 - \|x+iy\|^2 - i\|x-y\|^2 + \|x-iy\|^2\right) \\
&= \frac{i}{4}\left(\|x+y\|^2 + i\|x+iy\|^2 - \|x-y\|^2 - i\|x-iy\|^2\right) \\
&= i\langle x,y\rangle.
\end{aligned}$$

Now let $\lambda = r + is \in \mathbb{C}$. Then we have

$$\begin{aligned}
\langle \lambda x, y\rangle &= \langle (r+is)x, y\rangle \\
&= \langle rx + isx, y\rangle \\
&= \langle rx, y\rangle + \langle isx, y\rangle \\
&= r\langle x,y\rangle + s\langle ix, y\rangle \\
&= r\langle x,y\rangle + is\langle x,y\rangle \\
&= (r+is)\langle x,y\rangle \\
&= \lambda\langle x,y\rangle
\end{aligned}$$

for all $x, y \in V$. Therefore (59) respects $\mathbb{C}$-scaling in the first component. □

25.3 Distances and Pseudo Normed Vector Spaces

Let $\mathcal{V}$ be an inner-product space and let $\mathcal{A}$ be a subspace of $\mathcal{V}$. We define

$$d(x, \mathcal{A}) = \inf\{\|x-a\| \mid a \in A\} \tag{60}$$

for all $x \in \mathcal{V}$. The map $d(-,\mathcal{A})$ is a good candidate for a norm on $\mathcal{V}$. It turns out however that $d(-,\mathcal{A})$ is just a **pseudo norm**.

Definition 25.3. Let V be a vector space over $\mathbb{C}$. A map $p\colon V \to [0,\infty)$ is called a **pseudo norm on** V if it satisfies the following properties:

1. (Subadditivity) $p(u+v) \leq p(u) + p(v)$ for all $u, v \in V$,
2. (Absolutely Homogeneous) $p(\lambda v) = |\lambda| p(v)$ for all $v \in V$ and $\lambda \in \mathbb{C}$.

A vector space equipped with a pseudo inner-product is called a **pseudo normed vector space**.

Remark 44. Thus a norm is just a pseudo norm with the positive-definiteness property.

25.3.1 Absolute Homogeneity of Distances

Proposition 25.3. *Let $\mathcal{V}$ be an inner-product space, let $\mathcal{A}$ be a subspace of $\mathcal{V}$, let $x \in \mathcal{V}$, and let $\lambda \in \mathbb{C}$. Then*

$$d(\lambda x, \mathcal{A}) = |\lambda| d(x, \mathcal{A}).$$

Proof. Choose a sequence (y_n) of elements in $\mathcal{A}$ such that

$$\|x - y_n\| < d(x,\mathcal{A}) + \frac{1}{|\lambda n|}$$

for all $n \in \mathbb{N}$. Then since $\mathcal{A}$ is a subspace, we have

$$\begin{aligned} d(\lambda x, \mathcal{A}) &\leq \|\lambda x - \lambda y_n\| \\ &= |\lambda| \|x - y_n\| \\ &< |\lambda| \left(d(x,\mathcal{A}) + \frac{1}{|\lambda n|} \right) \\ &= |\lambda| d(x,\mathcal{A}) + \frac{1}{n} \end{aligned}$$

for all $n \in \mathbb{N}$. In particular, this implies $d(\lambda x, \mathcal{A}) \leq |\lambda| d(x,\mathcal{A})$.

Conversely, choose a sequence (z_n) of elements in $\mathcal{A}$ such that

$$\|\lambda x - z_n\| < d(\lambda x, \mathcal{A}) + \frac{1}{n}$$

Then since $\mathcal{A}$ is a subspace, we have

$$\begin{aligned} |\lambda| d(x,\mathcal{A}) &\leq |\lambda| \|x - z_n/|\lambda|\| \\ &= \|\lambda x - z_n\| \\ &< d(\lambda x, \mathcal{A}) + \frac{1}{n} \end{aligned}$$

for all $n \in \mathbb{N}$. In particular, this implies $|\lambda| d(x,\mathcal{A}) \leq d(\lambda x, \mathcal{A})$. □

25.3.2 Subadditivity of Distances

Proposition 25.4. *Let $\mathcal{V}$ be an inner-product space, let $\mathcal{A}$ be a subspace of $\mathcal{V}$, and let $x, y \in \mathcal{V}$. Then*

$$d(x+y, \mathcal{A}) \leq d(x,\mathcal{A}) + d(y,\mathcal{A}).$$

Proof. Choose a sequences (w_n) and (z_n) of elements in $\mathcal{A}$ such that

$$\|x - w_n\| < d(x,\mathcal{A}) + \frac{1}{2n} \quad \text{and} \quad \|y - z_n\| < d(y,\mathcal{A}) + \frac{1}{2n}$$

for all $n \in \mathbb{N}$. Then since $\mathcal{A}$ is a subspace, we have

$$\begin{aligned} d(x+y, \mathcal{A}) &\leq \|(x+y) - (w_n + z_n)\| \\ &\leq \|x - w_n\| + \|y - z_n\| \\ &< d(x, \mathcal{A}) + \frac{1}{2n} + d(y, \mathcal{A}) + \frac{1}{2n} \\ &= d(x, \mathcal{A}) + d(y, \mathcal{A}) + \frac{1}{n} \end{aligned}$$

for all $n \in \mathbb{N}$. In particular, this implies $d(x+y, \mathcal{A}) \leq d(x, \mathcal{A}) + d(y, \mathcal{A})$. □

25.3.3 Quotienting out to get a Norm

To see why (60) is just a pseudo norm and not a norm, note that $d(x, \mathcal{A}) = 0$ if and only if $x \in \overline{\mathcal{A}}$. To remedy this situation, we quotient out by $\overline{\mathcal{A}}$. First we need a lemma.

Lemma 25.1. *Let $x \in \mathcal{V}$. Then $d(x, \mathcal{A}) = d(x, \overline{\mathcal{A}})$.*

Proof. We have $d(x, \mathcal{A}) \geq d(x, \overline{\mathcal{A}})$ since $\mathcal{A} \subseteq \overline{\mathcal{A}}$. For the reverse inequality, we assume (for a contradiction) that $d(x, \mathcal{A}) > d(x, \overline{\mathcal{A}})$. For the reverse inequality, let $\varepsilon > 0$. Choose $a \in \overline{\mathcal{A}}$ such that

$$\|x - a\| < d(x, \overline{\mathcal{A}}) + \varepsilon/2.$$

Choose $b \in \mathcal{A}$ such that $\|a - b\| < \varepsilon/2$. Then

$$\begin{aligned} d(x, \mathcal{A}) &\leq \|x - b\| \\ &\leq \|x - a\| + \|a - b\| \\ &< d(x, \overline{\mathcal{A}}) + \varepsilon/2 + \varepsilon/2 \\ &= d(x, \overline{\mathcal{A}}) + \varepsilon. \end{aligned}$$

Since this holds for all $\varepsilon > 0$, we have $d(x, \mathcal{A}) \leq d(x, \overline{\mathcal{A}})$. □

Proposition 25.5. *The pseudo norm $d(-, \mathcal{A})$ on $\mathcal{V}$ induces a well-defined norm $\|\cdot\|$ on $\mathcal{V}/\overline{\mathcal{A}}$, defined by*

$$\|\overline{x}\| = d(x, \mathcal{A}) \tag{61}$$

for all $\overline{x} \in \mathcal{V}/\overline{\mathcal{A}}$.

Proof. We first check that (61) is well-defined. Let $\overline{x} \in \mathcal{V} \backslash \overline{\mathcal{A}}$[13]. Choose another representative of the coset $\overline{x}$, say $x + a$ where $a \in \overline{\mathcal{A}}$. Then

$$\begin{aligned} \|\overline{x+a}\| &= d(x+a, \mathcal{A}) \\ &= d(x+a, \overline{\mathcal{A}}) \\ &= \inf\{\|x + a - b\| \mid b \in \overline{\mathcal{A}}\} \\ &= \inf\{\|x - c\| \mid c \in \overline{\mathcal{A}}\} \\ &= d(x, \mathcal{A}) \\ &= \|\overline{x}\|. \end{aligned}$$

Thus $\|\cdot\|$ is well-defined.

Now $\|\cdot\|$ is a pseudo norm on $\mathcal{V} \backslash \overline{\mathcal{A}}$ since it inherits these properties from $d(-, \mathcal{A})$ on $\mathcal{V}$. Indeed, for subadditivity, we have

$$\begin{aligned} \|\overline{x} + \overline{y}\| &= d(x+y, \mathcal{A}) \\ &\leq d(x, \mathcal{A}) + d(y, \mathcal{A}) \\ &= \|\overline{x}\| + \|\overline{y}\| \end{aligned}$$

[13]Do not confuse the overline over $\mathcal{A}$ with the overline over x. One denotes the closure of $\mathcal{A}$ and the other denotes a coset in $\mathcal{V}/\overline{\mathcal{A}}$ with a given represenative $x \in \mathcal{V}$.

for all $\overline{x}, \overline{y} \in \mathcal{V} \backslash \overline{\mathcal{A}}$, and for absolute homogeneity, we have

$$\begin{aligned}
\|\lambda \overline{x}\| &= d(\lambda x, \mathcal{A}) \\
&= |\lambda| d(x, \mathcal{A}) \\
&= |\lambda| \|\overline{x}\|
\end{aligned}$$

for all $\overline{x} \in \mathcal{V} \backslash \overline{\mathcal{A}}$ and $\lambda \in \mathbb{C}$.

Moreover $\| \cdot \|$ is a norm on $\mathcal{V} \backslash \overline{\mathcal{A}}$ since we also have positive-definiteness. Indeed, let $\overline{x} \in \mathcal{V} / \overline{\mathcal{A}}$. Then

$$\begin{aligned}
\overline{x} = 0 &\iff x \in \overline{\mathcal{A}} \\
&\iff d(x, \mathcal{A}) = 0 \\
&\iff \|\overline{x}\| = 0.
\end{aligned}$$

□

26 Duality

Let K be a field, let V be a K-vector space with basis $\beta = \{\beta_1, \ldots, \beta_m\}$, and let W be a K-vector space with basis $\gamma = \{\gamma_1, \ldots, \gamma_n\}$. Recall from linear algebra that we define the **(algebraic) dual** of V to be the K-vector space

$$V^\star := \{\varphi \colon V \to K \mid \varphi \text{ is linear}\}.$$

where addition and scalar multiplication are defined by

$$(\varphi + \psi)(v) = \varphi(v) + \psi(v) \quad \text{and} \quad (\lambda\varphi)(v) = \varphi(\lambda v)$$

for all $\varphi, \psi \in V^\star$, $\lambda \in \mathbb{C}$, and $v \in V$. The **(algebraic) dual** of β is defined to be the basis of $V^\star$ given by $\beta^\star := \{\beta_1^\star, \ldots, \beta_m^\star\}$, where each $\beta_i^\star$ is uniquely determined by

$$\beta_i^\star(\beta_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases}$$

If $T \colon V \to W$ is a linear map, we define its **(algebraic) dual** to be the linear map $T^\star \colon W^\star \to V^\star$ defined by

$$T^\star(\varphi) = \varphi \circ T$$

for all $\varphi \in W^\star$. One learns in linear algebra that the transpose of the matrix representation of T with respect to the bases β and γ is equal to the matrix representation of $T^\star$ with respect to the bases $\gamma^\star$ and $\beta^\star$. In terms of notation, this is written as

$$([T]_\beta^\gamma)^\top = [T^\star]_{\gamma^\star}^{\beta^\star}.$$

We want to describe an analog of this situation for inner-product spaces over $\mathbb{C}$.

Definition 26.1. Let $\mathcal{V}$ be an inner-product space over $\mathbb{C}$. We define its **(continuous) dual space**[a] to be

$$\begin{aligned}
\mathcal{V}^* &:= \{\ell \colon \mathcal{V} \to \mathbb{C} \mid \ell \text{ is linear and continuous}\}. \\
&= \{\ell \colon \mathcal{V} \to \mathbb{C} \mid \ell \text{ is a bounded operator}\}.
\end{aligned}$$

[a]When speaking about the dual space of an inner-product space, we will always mean its continuous dual.

Remark 45. Thus $\mathcal{V}^*$ captures both topological and linear aspects of $\mathcal{V}$.

26.1 Riesz Representation Theorem Revisited

Theorem 26.1. *(Riesz Representation Theorem) Let $\mathcal{H}$ be a Hilbert space. Then there exists an isometric antiisomorphism $\Phi \colon \mathcal{H} \to \mathcal{H}^*$.*

Proof. Define $\Phi \colon \mathcal{H} \to \mathcal{H}^*$ by

$$\Phi(x) = \langle \cdot, x \rangle$$

for all $x \in \mathcal{H}$. We first show Φ is antilinear. Let $x, y \in \mathcal{H}$ and let $\lambda, \mu \in \mathbb{C}$. Then

$$\begin{aligned}\Phi(\lambda x + \mu y)(z) &= \langle z, \lambda x + \mu y \rangle \\ &= \overline{\lambda}\langle z, x \rangle + \overline{\mu}\langle z, y \rangle \\ &= \overline{\lambda}\Phi(x)(z) + \overline{\mu}\Phi(y)(z)\end{aligned}$$

for all $z \in \mathcal{H}$. Therefore Φ is antilinear.

Note Φ is an injective antilinear map since the inner-product is positive-definite. Also, the Riesz representation theorem implies Φ is surjective. Finally Example (**??**) implies Φ is an isometry. Therefore Φ is an isometric antiisomorphism. $\square$

27 Limit Infimum

Let (a_n) be a sequence of positive real numbers. We define the **limit infimum** of (a_n), denoted $\liminf(a_n)$, to be the limit

$$\liminf(a_n) := \lim_{N \to \infty} (\inf\{a_n \mid n \geq N\}). \tag{62}$$

Since the sequence $(\inf\{a_n \mid n \geq N\})_{N \in \mathbb{N}}$ is a monotone increasing sequence in N, the limit (62) always exists or equals $-\infty$.

Proposition 27.1. *Let (a_n) be a sequence of positive real-valued numbers.*

1. *If $\liminf(a_n) = A$, then for each $\varepsilon > 0$ and $N \in \mathbb{N}$, there exists some $n \geq N$ such that $a_n > A + \varepsilon$. In other words, for all $\varepsilon > 0$, the sequence (a_n) is frequently strictly less than $A + \varepsilon$.*

2. *If $\liminf(a_n) = A$, then for each $\varepsilon > 0$, there exists $N \in N$ such that $a_n > A - \varepsilon$ for all $n \geq N$. In other words, for all $\varepsilon > 0$, the sequence (a_n) is eventually strictly greater than $A - \varepsilon$.*

3. *Conversely, if $A \geq 0$ satisfies 1 and 2, then liminf$(a_n) = A$.*

Proof.

1. We prove this by contradiction. To obtain a contradiction, assume that there there exists $\varepsilon > 0$ and $N \in \mathbb{N}$ such that there does not exist any $n \geq N$ with $a_n < A + \varepsilon$. Choose such $\varepsilon > 0$ and $N \in \mathbb{N}$. Then $a_n \geq A + \varepsilon$ for all $n \geq N$. This implies $\inf\{a_n \mid n \geq N\} > A + \varepsilon$. Since $\inf\{a_n \mid n \geq N\}$ is a monotone increasing function of N, this implies $\liminf(a_n) \geq A + \varepsilon$, which is a contradiction.

2. We prove this by contradiction. To obtain a contradiction, assume that there exists $\varepsilon > 0$ such that there does not exist an $N \in \mathbb{N}$ with $a_n > A - \varepsilon$ for all $n \geq N$. Choose such $\varepsilon > 0$ and let $N \in \mathbb{N}$. Then there exists $n \geq N$ such that $a_n \leq A - \varepsilon$. In particular, this implies

$$\inf\{a_n \mid n \geq N\} \leq A - \varepsilon$$

Since N is arbitrary, this further implies

$$\begin{aligned}\liminf(a_n) &= \lim_{N \to \infty} (\inf\{a_n \mid n \geq N\}) \\ &\leq A - \varepsilon,\end{aligned}$$

which contradicts the fact that $\liminf(a_n) = A$.

3. Let $A \geq 0$ satisfy 1 and 2 and let $A' = \liminf(a_n)$. We prove by contradiction that $A = A'$. Assume for a contradiction that $A < A'$. Let $\varepsilon = A' - A$. Since A' satisfies 2, the sequence (a_n) is eventually greater than $A' - \varepsilon/2$. On the other hand, since A satisfies 1, the sequence (a_n) is frequently less than $A + \varepsilon/2 = A' - \varepsilon/2$. This is a contradiciton. An analagous argument gives a contradiction when we assume $A > A'$. Therefore $A = A'$. $\square$

Lemma 27.1. *Let (a_n) and (b_n) be two sequences of positive real numbers such that* $\liminf(a_n) = A$ *and* $\liminf(b_n) = B$. *Then*

1. $\liminf(a_n b_n) = AB$
2. $\liminf(a_n + b_n) = A + B$

Proof.

1. Let $\varepsilon > 0$. Since the sequence (a_n) is eventually greater than $A - \varepsilon/(A+B)$ and since the sequence (b_n) is eventually greater than $B - \varepsilon/(A+B)$, the sequence $(a_n b_n)$ is eventually greater than

$$\begin{aligned}\left(A - \frac{\varepsilon}{A+B}\right)\left(B - \frac{\varepsilon}{A+B}\right) &= AB - \frac{\varepsilon(A+B)}{(A+B)} + \frac{\varepsilon^2}{(A+B)^2} \\ &\geq AB - \varepsilon.\end{aligned}$$

An analagous argument shows that $(a_n b_n)$ is frequently less than $AB + \varepsilon$.

2. This is proved in the same way as 1. □

Part III
Measure Theory

28 Introduction

Measure theory is a central subject in analysis. Let us consider four problems which helped lead to the development of measure theory.

1) When we studied inner-products in Hilbert spaces, we considered the inner-product space $(C[a,b], \langle \cdot, \cdot \rangle_1)$, where $C[a,b]$ is the $\mathbb{C}$-vector space of all continuous functions defined on the interval $[a,b]$ and $\langle \cdot, \cdot \rangle_1$ is the inner-product defined by

$$\langle f, g \rangle = \int_a^b f(x)\overline{g(x)} \mathrm{d}x \tag{63}$$

for all $f, g \in C[a,b]$. One of the problems that we ran into when studying this inner-product space is that it is *not* a Hilbert space. In other words, $C[a,b]$ is not complete with respect to topology induced by this inner-product. It turns out however that just as how $\mathbb{R}$ can be viewed as the "completion" of $\mathbb{Q}$ with respect to the usual topology on $\mathbb{Q}$ induced by the usual absolute value $|\cdot|$, there is a similar space which can be viewed as the "completion" of $C[a,b]$ with respect to the topology induced by the inner-product $\langle \cdot, \cdot \rangle_1$. Measure theory will give us a nice interpretation of what this completed space looks like.

2) Now consider the normed linear space $(C[a,b], \|\cdot\|_\infty)$, where $\|\cdot\|_\infty$ is the supremum norm. This time $C[a,b]$ *is* complete with respect to the topology induced by this norm. In other words, $(C[a,b], \|\cdot\|_\infty)$ is a Banach space. In this case, we'd like to know what the dual space looks like. Recall that the dual of a normed linear space $(\mathcal{X}, \|\cdot\|)$ is defined to be the space $(\mathcal{X}^*, \|\cdot\|)$ where

$$\mathcal{X}^* := \{\ell \colon \mathcal{X} \to \mathbb{C} \mid \ell \text{ is a bounded linear functional}\}$$

and where the norm, denoted $\|\cdot\|$ again, is defined by

$$\|\ell\| = \sup\{|\ell(x)| \mid \|x\| \leq 1\}$$

for all $\ell \in \mathcal{X}^*$. For instance, here are three examples of a bounded linear functionals on $C[a,b]$ (with respect to the sup norm)

$$\begin{aligned}\ell_1(f) &= \int_a^b f(x)\mathrm{d}x \\ \ell_2(f) &= f(a) \\ \ell_3(f) &= f(a) + f(b).\end{aligned}$$

Here again we will find that measure theory will give us a natural interpretation of what the dual space of $C[a,b]$ looks like. We shall that see that, in a certain sense, the bounded linear functionals on $C[a,b]$ will correspond to different ways you can integrate on $[a,b]$.

3) The third problem we consider goes as follows: let $T\colon \mathcal{H} \to \mathcal{H}$ be a bounded linear map from a Hilbert space to itself. If T is compact and self-adjoint, then the Spectral Theorem tells us that there exists an orthonormal basis of eigenvectors (e_n) with eigenvalues (λ_n) such that

$$Tx = \sum_{n=1}^{\infty} \lambda_n \langle x, e_n \rangle e_n$$

for all $x \in \mathcal{H}$. Can we still get a spectral theorem if we remove the compactness condition? It turns out that there is a way to do this. We may not be able to completely show this in this class (perhaps in a functional analysis class), but we will make progress using measure theory.

4) The foundations of probability theory is based on measure theory. We will better understand the connection between probability theory and measure theory.

Let's go back to $C[a,b]$ equipped with the inner-product $\langle \cdot, \cdot \rangle_1$ described above. As we said, the main problem with $(C[a,b], \langle \cdot, \cdot \rangle_1)$ is that this space is not complete. Ideally, we would like to be able to measure "the length" of any subset of $\mathbb{R}$ such that

1. the length of an interval (c,d) is $d-c$,
2. if we translate a set, its length should stay the same,
3. if (E_n) is a sequence of pairwise disjoint subsets of $\mathbb{R}$ with lengths $\mu(E_n)$, then the length of their union satisfy
$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n).$$

29 Intervals, Algebras, and Measures

Throughout the rest of this section, we consider the normed linear space $(C[a,b], \|\cdot\|)$[14] where, unless otherwise specified, $\|\cdot\|$ is the L_1 norm:
$$\|f\| := \int_a^b |f(x)| \mathrm{d}x$$
for all $f \in C[a,b]$ where the integral is understood to be the Riemann integral. It is easy to prove that $C[a,b]$ equipped with the L_1 norm is a normed linear space, but it is not complete. As shown in the Appendix, we know that we can construct the completion of $C[a,b]$ using Cauchy sequences, but this space is a little too abstract for our desires. We would like to describe the completion of $C[a,b]$ in a more concrete way. More specifically, we'd like to realize the completion of $C[a,b]$ as a quotient of a certain space of functions (rather than as a quotient of a space of Cauchy sequences).

Indeed, if $\mathcal{Y}$ is a dense subspace of a normed linear space $\mathcal{X}$, then every bounded linear functional $\ell\colon \mathcal{Y} \to \mathbb{C}$ can be extended in a unique way to a bounded linear functional $\widetilde{\ell}\colon \mathcal{X} \to \mathbb{C}$ with the same norm. That is, $\widetilde{\ell}|_{\mathcal{Y}} = \ell$ and $\|\widetilde{\ell}\| = \|\ell\|$. The linear functional $\ell\colon C[a,b] \to \mathbb{C}$ defined by

$$\ell(f) = \int_a^b f(x) \mathrm{d}x$$

for all $f \in C[a.b]$ is bounded with $\|\ell\| = 1$. So if we can define the completion of $C[a,b]$ as a quotient of a space of functions, then we will be able to compute $\widetilde{\ell}(f)$ for every f. In turn, this can be naturally viewed as computing $\int_a^b f(x)\mathrm{d}x$. This new integral will be called the **Lebesgue integral** and the representative functions in this space will be called **Lebesgue measurable functions**. Let us denote the completion of $C[a,b]$ with respect to $\|\cdot\|$ by $(\mathscr{C}[a,b], \|\cdot\|)$, or more simply $\mathscr{C}[a,b]$.

[14] Unless otherwise specified, we will often simply write $C[a,b]$ instead of $(C[a,b], \|\cdot\|)$ and view $C[a,b]$ as a normed linear space equipped with the L_1 norm.

29.1 Intervals

Unless otherwise specified, by a a **subinterval of** $[a,b]$, we shall mean a subinterval of $[a,b]$ of the following types: (c,d), $(c,b]$, $[c,b)$, and $[c,d]$, where

$$-\infty < a \leq c \leq d \leq b < \infty.$$

Note that $(c,c) = \emptyset$ and $[c,c] = \{c\}$ are considered intervals as well. If I is a subinterval of $[a,b]$, then the **indicator function** $1_I \colon [a,b] \to \{0,1\}$ with respect to I is defined by

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

for all $x \in [a,b]$. If I is a subinterval of $[a,b]$, then it closure in $[a,b]$ has the form

$$\overline{I} = [c,d]$$

for some $a \leq c \leq d \leq b$. In this case, we define the **length** of I to be

$$\text{length}(I) = d - c.$$

We denote by $\mathcal{I}_{[a,b]}$ (or more simply just $\mathcal{I}$ if context is clear) to be the collection of all finite unions of subintervals of $[a,b]$.

29.1.1 Collection of all Subintervals of $[a,b]$ Forms a Semialgebra of Sets

Definition 29.1. A nonempty collection $\mathcal{E}$ of subsets of X is said to be a **semialgebra** of sets if it satisfies the following properties:

1. $\emptyset \in \mathcal{E}$;
2. If $E, F \in \mathcal{E}$, then $E \cap F \in \mathcal{E}$;
3. If $E \in \mathcal{E}$, then E^c is a finite disjoint union of sets from $\mathcal{E}$.

Proposition 29.1. *The collection of all subintervals of $[a,b]$ forms a semialgebra of sets.*

Proof. Let $\mathcal{I}$ denote the collection of all subintervals of $[a,b]$. We have $\emptyset \in \mathcal{I}$ since $\emptyset = (c,c)$ for any $c \in [a,b]$. Now we show $\mathcal{I}$ is closed under finite intersections. Let I_1 and I_2 be subintervals of $[a,b]$. Taking the closure of I_1 and I_2 gives us closed intervals, say

$$\overline{I}_1 = [c_1, d_1] \quad \text{and} \quad \overline{I}_2 = [c_2, d_2].$$

Assume without loss of generality that $c_1 \leq c_2$. If $d_1 < c_2$, then $I_1 \cap I_2 = \emptyset$, so assume that $d_1 \geq c_2$. If $d_1 \geq d_2$, then $I_1 \cap I_2 = I_2$, so assume that $d_1 < d_2$. If $c_1 = c_2$, then $I_1 \cap I_2 = I_1$, so assume that $c_2 > c_1$. So we have reduced the case to where

$$c_1 < c_2 \leq d_1 < d_2.$$

With these assumptions in mind, we now consider four cases:

Case 1: If $I_1 = [c_1, d_1]$ or $I_1 = (c_1, d_1]$ and $I_2 = [c_2, d_2]$ or $I_2 = [c_2, d_2)$, then $I_1 \cap I_2 = [c_2, d_1]$.

Case 2: If $I_1 = [c_1, d_1)$ or $I_1 = (c_1, d_1)$ and $I_2 = [c_2, d_2]$ or $I_2 = [c_2, d_2)$, then $I_1 \cap I_2 = [c_2, d_1)$.

Case 3: If $I_1 = [c_1, d_1]$ or $I_1 = (c_1, d_1)$ and $I_2 = (c_2, d_2]$ or $I_2 = (c_2, d_2)$, then $I_1 \cap I_2 = (c_2, d_1)$.

Case 4: If $I_1 = [c_1, d_1)$ or $I_1 = (c_1, d_1)$ and $I_2 = (c_2, d_2]$ or $I_2 = (c_2, d_2)$, then $I_1 \cap I_2 = (c_2, d_1)$.

In all cases, we see that $I_1 \cap I_2$ is a subinterval of $[a,b]$.

Now we show that compliments can be expressed as finite disjoint unions. Let I be a subinterval of $[a,b]$ and write $\overline{I} = [c,d]$. We consider four cases:

Case 1: If $I = [c, d]$, then $I^c = [a, c) \cup (d, b]$.

Case 2: If $I = (c, d]$, then $I^c = [a, c] \cup (d, b]$.

Case 3: If $I = [c, d)$, then $I^c = [a, c) \cup [d, b]$.

Case 4: If $I = (c, d)$, then $I^c = [a, c] \cup [d, b]$.

Thus in all cases, we can express I^c as a disjoint union of intervals since $a \leq c \leq d \leq b$. $\square$

29.1.2 Indicator Function 1_I Represents an Element in $\mathscr{C}[a, b]$

The next proposition tells us that 1_I should represent an element in $\mathscr{C}[a, b]$ for all subintervals I of $[a, b]$ and that $\|1_I\| = \text{length}(I)$.

Proposition 29.2. *Let I be a subinterval of $[a, b]$. Then there exists a Cauchy sequence (f_n) in $(C[a, b], \|\cdot\|)$ such that (f_n) converges pointwise to 1_I on $[a, b]$. Moreover we have*

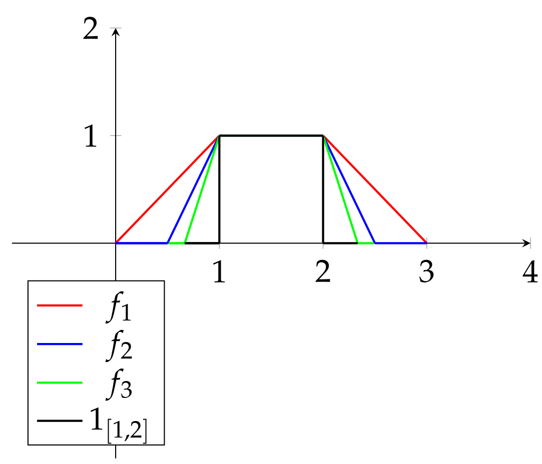
$$\lim_{n \to \infty} \|f_n\| = \text{length}(I).$$

Proof. If $I = \emptyset$, then we take $f_n = 0$ for all $n \in \mathbb{N}$. Thus assume I is a nonempty subinterval of $[a, b]$. We consider two cases; namely $I = (c, d)$ and $I = [c, d]$. The other cases ($I = (c, d]$ and $I = [c, d)$) will easily be seen to be a mixture of these two cases.

Case 1: Suppose $I = [c, d]$. For each $n \in \mathbb{N}$, define

$$f_n(x) = \begin{cases} 0 & \text{if } a \leq x < c - \left(\frac{c-a}{n}\right) \\ \frac{n}{c-a}(x - c) + 1 & \text{if } c - \left(\frac{c-a}{n}\right) \leq x < c \\ 1 & \text{if } c \leq x \leq d \\ \frac{n}{d-b}(x - d) + 1 & \text{if } d < x \leq d + \left(\frac{b-d}{n}\right) \\ 0 & \text{if } d + \left(\frac{b-d}{n}\right) < x \leq b \end{cases}$$

The image below gives the graphs for f_1, f_2, and f_3 in the case where $[a, b] = [0, 3]$ and $[c, d] = [1, 2]$.



For each $n \in \mathbb{N}$, the function f_n is continuous since each of its segments is continuous and are equal on their boundaries.

Let us check that (f_n) converges pointwise to 1_I: If $x \in [a, c)$, then we choose $N \in \mathbb{N}$ such that

$$x \leq c - \left(\frac{c-a}{N}\right).$$

Then $f_n(x) = 0$ for all $n \geq N$. Thus

$$\lim_{n \to \infty} f_n(x) = 0 = 1_I(x).$$

Similarly, if $x \in (d, b]$, then we choose $N \in \mathbb{N}$ such that

$$x \geq d + \left(\frac{b-d}{N}\right).$$

Then $f_n(x) = 0$ for all $n \geq N$. Thus

$$\lim_{n\to\infty} f_n(x) = 0 = 1_I(x).$$

Finally, if $x \in [c, d]$, then $f_n(x) = 1$ for all $n \in \mathbb{N}$ by definition and thus

$$\lim_{n\to\infty} f_n(x) = 0 = 1_I(x).$$

Let us check that (f_n) is Cauchy in $(C[a,b], \|\cdot\|_1)$. Let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that

$$\frac{c-a+b-d}{n} < \varepsilon$$

for all $n \geq N$. Then $n \geq m \geq N$ implies

$$\begin{aligned}
\|f_n - f_m\|_1 &= \int_a^b |f_n(x) - f_m(x)| \mathrm{d}x \\
&= \int_a^b (f_n(x) - f_m(x)) \mathrm{d}x \\
&= \int_{c-\left(\frac{c-a}{m}\right)}^{c} (f_n(x) - f_m(x)) \mathrm{d}x + \int_d^{d+\left(\frac{b-d}{m}\right)} (f_n(x) - f_m(x)) \mathrm{d}x \\
&\leq \int_{c-\left(\frac{c-a}{m}\right)}^{c} \mathrm{d}x + \int_d^{d+\left(\frac{b-d}{m}\right)} \mathrm{d}x \\
&= \frac{c-a}{m} + \frac{b-d}{m} \\
&= \frac{c-a+b-d}{m} \\
&< \varepsilon.
\end{aligned}$$

Thus the sequence (f_n) is Cauchy in $(C[a,b], \|\cdot\|_1)$.

Finally, we check that $\|f_n\|_1 \to \text{length}(I)$ as $n \to \infty$. We have

$$\begin{aligned}
d - c &\leq \|f_n\|_1 \\
&= \int_a^b |f_n(x)| \mathrm{d}x \\
&= \int_a^b f_n(x) \mathrm{d}x \\
&= \int_{c-\left(\frac{c-a}{n}\right)}^{c} f_n(x) \mathrm{d}x + \int_c^d \mathrm{d}x + \int_d^{d+\left(\frac{b-d}{n}\right)} f_n(x) \mathrm{d}x \\
&\leq \int_{c-\left(\frac{c-a}{n}\right)}^{c} \mathrm{d}x + \int_c^d \mathrm{d}x + \int_d^{d+\left(\frac{b-d}{n}\right)} \mathrm{d}x \\
&= \frac{c-a}{n} + d - c + \frac{b-d}{n} \\
&\to d - c.
\end{aligned}$$

Thus for each $n \in \mathbb{N}$, we have

$$d - c \leq \|f_n\|_1 \leq d - c + \frac{c-a+b-d}{n}. \tag{64}$$

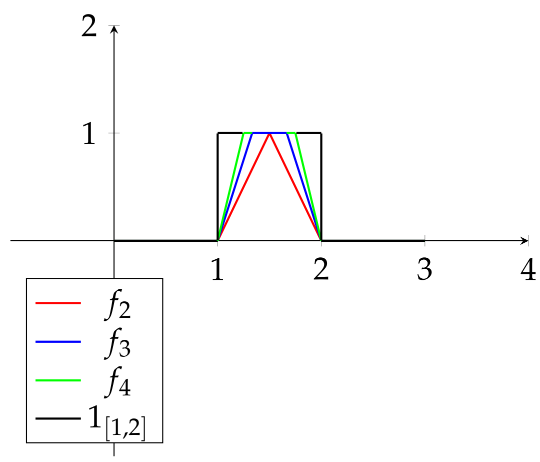
By taking $n \to \infty$ in (64), we see that

$$\begin{aligned}
\lim_{n\to\infty} \|f_n\|_1 &= d - c \\
&= \text{length}(I).
\end{aligned}$$

Case 2: Suppose $I = (c, d)$. For each $n \geq 2$, define

$$f_n(x) = \begin{cases} 0 & \text{if } a \leq x \leq c \\ \frac{n}{d-c}(x-c) & \text{if } c < x \leq c + \left(\frac{d-c}{n}\right) \\ 1 & \text{if } c + \left(\frac{d-c}{n}\right) \leq x \leq d - \left(\frac{d-c}{n}\right) \\ \frac{n}{c-d}(x-d) & \text{if } d - \left(\frac{d-c}{n}\right) \leq x \leq d \\ 0 & \text{if } d \leq x \leq b \end{cases}$$

The image below gives the graphs for f_2, f_3, and f_4 in the case where $[a, b] = [0, 3]$ and $(c, d) = (1, 2)$.



That (f_n) is a Cauchy sequence of continuous funtions in $(C[a, b], \|\cdot\|_1)$ which converges pointwise to 1_I and $\|f_n\|_1 \to \text{length}(I)$ as $n \to \infty$ follows from similar arguments used in case 1. $\square$

Remark 46. Note that $1_{[c,c]}$ is not the zero function, even though $\|1_{[c,c]}\| = 0$. Thus we need to identify the function $1_{[c,c]}$ and the constant zero function in $\mathscr{C}[a, b]$. In other words, $1_{[c,c]}$ and the constant zero function should represent the same element in $\mathscr{C}[a, b]$. Similarly, we have

$$\begin{aligned} \|1_{(c,d]} - 1_{(c,d)}\| &= \|1_{[d,d]}\| \\ &= 0, \end{aligned}$$

and so the functions $1_{(c,d]}$ and $1_{(c,d)}$ should represent the same element in $\mathscr{C}[a, b]$.

29.1.3 Finite Sums of Indicator Functions of Intervals Represents Elements in $\mathscr{C}[a, b]$

Let E be any finite union of intervals. Since the collection of all intervals forms a semialgebra, we can express E as a finite union of disjoint intervals, say

$$E = \bigcup_{i=1}^{k} I_i$$

where $I_i \cap I_j = \emptyset$ whenever $i \neq j$. The next proposition tells us that 1_E should represent an element in $\mathscr{C}[a, b]$ and that

$$\|1_E\| = \sum_{i=1}^{k} \text{length}(I_i)$$

.

Proposition 29.3. *With the notation as above, there exists a Cauchy sequence (f_n) in $(C[a, b], \|\cdot\|)$ such that (f_n) converges pointwise to 1_E on $[a, b]$. Moreover we have*

$$\lim_{n\to\infty} \|f_n\| = \sum_{i=1}^{k} \text{length}(I_i). \tag{65}$$

Proof. For each $1 \le i \le k$, choose a Cauchy sequence $(f_n^i)_{n\in\mathbb{N}}$ such that (f_n^i) converges pointwise to 1_{I_i} on $[a,b]$ and satisfies

$$\lim_{n\to\infty} \|f_n^i\| = \text{length}(I_i).$$

For each $n \in \mathbb{N}$, set $f_n = \sum_{i=1}^k f_n^i$. We claim that (f_n) is a Cauchy sequence that converges pointwise to 1_E and satisfies (65). Indeed, it is Cauchy since a sum of Cauchy sequences is Cauchy. Let us check that it converges pointwise to 1_E. Let $x \in E$ and let $\varepsilon > 0$. Then $x \in I_i$ for some $1 \le i \le k$, and so there exists an $N \in \mathbb{N}$ such that $n \ge N$ implies

$$|f_n^i(x) - 1_{I^i}(x)| < \frac{\varepsilon}{k}$$

for all $1 \le i \le n$. Choose such an $N \in \mathbb{N}$. Then $n \ge N$ implies

$$\begin{aligned}
|f_n(x) - 1_E(x)| &= \left| \sum_{i=1}^k f_n^i(x) - \sum_{i=1}^k 1_{I_i}(x) \right| \\
&\le \sum_{i=1}^k |f_n^i(x) - 1_{I_i}(x)| \\
&< \sum_{i=1}^k \frac{\varepsilon}{k} \\
&= \varepsilon.
\end{aligned}$$

It follows that f_n converges pointwise to 1_E. Finally, let us check that (65) holds. Let $\varepsilon > 0$. Without loss of generality, we may assume that $\|f_n^i\| \ge \text{length}(I_i)$ and $\|f_n\| \ge \text{length}(E)$ for all $1 \le i \le k$ and for all $n \in \mathbb{N}$. Choose $N \in \mathbb{N}$ such that $n \ge N$ implies

$$\|f_n^i\| - \text{length}(I_i) < \frac{\varepsilon}{k}$$

for all $i = 1, \dots, k$. Then $n \ge N$ implies

$$\begin{aligned}
\|f_n\| - \text{length}(E) &= \left\| \sum_{i=1}^k f_n^i \right\| - \text{length}\left(\bigcup_{i=1}^k I_i \right) \\
&= \left\| \sum_{i=1}^k f_n^i \right\| - \sum_{i=1}^k \text{length}(I_i) \\
&\le \sum_{i=1}^k \|f_n^i\| - \sum_{i=1}^k \text{length}(I_i) \\
&< \varepsilon.
\end{aligned}$$

It follows that (65) holds. ∎

29.1.4 Well-Definedness of Length

Given the result above, it seems logical that the indiciator function of any set E which can be expressed as a finite disjoint union of intervals should represent an element in $\mathscr{C}[a,b]$. Moreover, the norm of 1_E (or the measure of E) ought to be the sums of the norms of 1_{I_i} (or the sum of the lengths of I_i). We need to be careful though! We may have two different ways of expressing E as a finite disjoint union, say

$$E = \bigcup_{i=1}^k I_i \quad \text{and} \quad E = \bigcup_{i'=1}^{k'} I'_{i'}.$$

We must check that if this is the case, then

$$\sum_{i=1}^k \text{length}(I_i) = \sum_{i'=1}^{k'} \text{length}(I'_{i'}).$$

It turns out that this is indeed true, but we will leave the proof of this to the reader.

29.2 Algebras

Definition 29.2. Let $\mathcal{A}$ be a nonempty collection of subsets of X. We make the following definitions.

1. We say $\mathcal{A}$ is an **algebra** if it is closed under finite intersections and is closed under complements. To be closed under finite intersections means that if $A, B \in \mathcal{A}$, the $A \cap B \in \mathcal{A}$. To be closed under complements means that if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$. Note that in this case, we automatically have $\emptyset \in \mathcal{A}$. Indeed, since $\mathcal{A}$ is nonempty, we can choose $A \in \mathcal{A}$. Then $\emptyset = A \cap A^c \in \mathcal{A}$. Note also that $\mathcal{A}$ is closed under finite unions. This means that if $A, B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$. Indeed, we have

$$\begin{aligned} A \cup B &= ((A \cup B)^c)^c \\ &= (A^c \cap B^c)^c \\ &\in \mathcal{A}. \end{aligned}$$

2. We say $\mathcal{A}$ is a **semialgebra** if it contains the emptyset, is closed under finite intersections, and complements can be expressed by finite disjoint unions of members of $\mathcal{A}$. The last part means if $A \in \mathcal{A}$, then there exists a pairwise disjoint sequence $A_1, \dots, A_n \in \mathcal{A}$ such that $A^c = \bigcup_{i=1}^n A_i$. Here we need include $\emptyset \in \mathcal{A}$ as part of the definition, since we don't necessarily get this from the other two axioms. However something that we do get from the other two aximos is that *relative* complements can be expressed by finite disjoint unions of members of $\mathcal{A}$. Indeed, let $A, B \in \mathcal{A}$. Choose a pairwise disjoint sequence $B_1, \dots, B_n \in \mathcal{A}$ such that $B^c = \bigcup_{i=1}^n B_i$. Then we have

$$\begin{aligned} A \backslash B &= A \cap B^c \\ &= A \cap \left(\bigcup_{i=1}^n B_i \right) \\ &= \bigcup_{i=1}^n A \cap B_i, \end{aligned}$$

where $A \cap B_1, \dots A \cap B_n$ is a pairwise disjoint sequence of members of $\mathcal{A}$. Clearly every algebra is a semialgebra.

3. We say $\mathcal{A}$ is a σ**-algebra** if it is closed under *countable* intersections and is closed under complements. To be closed under countable intersections means that if (A_n) is a sequence of members of $\mathcal{A}$, the $\bigcup_{n=1}^\infty A_n \in \mathcal{A}$. Clearly every σ-algebra is an algebra.

Remark 47. We typically use $\mathcal{E}$ to denote a semialgebra, $\mathcal{A}$ to denote an algebra, and $\mathcal{M}$ to denote a σ-algebra.

29.2.1 Obtaining an Algebra from a Semialgebra

Proposition 29.4. *Let $\mathcal{E}$ be a semialgebra of subsets of X. Then the collection $\mathcal{A}$ consisting of all sets which are finite disjoint union of sets in $\mathcal{E}$ forms an algebra of sets.*

Proof. Let us show that $\mathcal{A}$ is closed under finite intersections. Let $A, A' \in \mathcal{A}$ and express A and A' as a finite disjoint union of members of $\mathcal{E}$, say

$$A = E_1 \cup \cdots \cup E_n \quad \text{and} \quad A' = E'_1 \cup \cdots \cup E'_{n'}.$$

Then we have

$$
\begin{aligned}
A \cap A' &= \left(\bigcup_{i=1}^{n} E_i \right) \cap \left(\bigcup_{i'=1}^{n'} E'_{i'} \right) \\
&= \bigcup_{i'=1}^{n'} \left(\left(\bigcup_{i=1}^{n} E_i \right) \cap E'_{i'} \right) \\
&= \bigcup_{i'=1}^{n'} \left(\bigcup_{i=1}^{n} E_i \cap E'_{i'} \right) \\
&= \bigcup_{\substack{1 \leq i \leq n \\ 1 \leq i' \leq n'}} E_i \cap E'_{i'}
\end{aligned}
$$

where the union is disjoint since the E_i and $E'_{i'}$ are disjoint from one another whenever $i \neq i'$. Thus $A \cap A' \in \mathcal{A}$, and hence $\mathcal{A}$ is closed under finite intersections.

Now let us show that $\mathcal{A}$ is closed under complements. Let $A \in \mathcal{A}$ and express A as a finite disjoint union of members of $\mathcal{E}$, say

$$
A = \bigcup_{i=1}^{n} E_i.
$$

For each E_i, express E_i^c as a finite disjoint union of members of E, say

$$
E_i^c = \bigcup_{j_i=1}^{n_i} E_{i,j_i}.
$$

Then we have

$$
\begin{aligned}
A^c &= \left(\bigcup_{i=1}^{n} E_i \right)^c \\
&= \bigcap_{i=1}^{n} E_i^c \\
&= \bigcap_{i=1}^{n} \bigcup_{j_i=1}^{n_i} E_{i,j_i} \\
&= \bigcup_{j_i=1}^{n_i} \bigcap_{i=1}^{n} E_{i,j_i}
\end{aligned}
$$

where we were allowed to commute the union with the intersection from the third line to the fourth line since these are finite unions and finite intersections. Also, the union is disjoint since the E_{i,j_i} and E_{i,j'_i} are disjoint from one another whenever $j_i \neq j'_i$. Thus $A^c \in \mathcal{A}$, and hence $\mathcal{A}$ is closed complements. $\square$

29.2.2 Ascendification, Contractifcation, and Disjointification

Let $\mathcal{A}$ be an algebra of subsets of X and let (A_n) be a sequence in $\mathcal{A}$. There are several operations we can perform on (A_n) to obtain a new sequence in $\mathcal{A}$. They are called **ascendification**, **contractification**, and **disjointification**.

Definition 29.3. Let $\mathcal{A}$ be an algebra of subsets of X and let (A_n) be a sequence in $\mathcal{A}$.

1. Define the sequence (B_n) as follows: for all $n \in \mathbb{N}$, we set $B_n = \bigcup_{i=1}^n A_i$. Since $\mathcal{A}$ is closed under finite unions, it is clear that (B_n) is a sequence in $\mathcal{A}$. The (B_n) is an **ascending sequence**, which means $B_n \subseteq B_{n+1}$ for all $n \in \mathbb{N}$. We call (B_n) the **ascendification** of the sequence (A_n). Note that

$$\bigcup_{n=1}^{N} A_n = \bigcup_{n=1}^{N} B_n$$

for all $N \in \mathbb{N}$, and thus in particular, we have

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n.$$

2. Define the sequence (C_n) as follows: for all $n \in \mathbb{N}$, we set $C_n = \bigcap_{i=1}^n A_i$. Since $\mathcal{A}$ is closed under finite intersections, it is clear that (C_n) is a sequence in $\mathcal{A}$. The (C_n) is a **contracting sequence**, which means $C_n \supseteq C_{n+1}$ for all $n \in \mathbb{N}$. We call (C_n) the **ascendification** of the sequence (A_n). Note that

$$\bigcap_{n=1}^{N} A_n = \bigcap_{n=1}^{N} C_n$$

for all $N \in \mathbb{N}$, and thus in particular, we have

$$\bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} C_n.$$

3. Define the sequence (D_n) as follows: for all $n \in \mathbb{N}$, we set $D_n = A_n \setminus \left(\bigcup_{i=1}^{n-1} A_i\right)$. Since $\mathcal{A}$ is closed under finite unions and is closed under relative complements, it is clear that (D_n) is a sequence in $\mathcal{A}$. The (D_n) is a **pairwise disjoint sequence**, which means $D_m \cap D_n = \emptyset$ whenever $m \neq n$. . We call (D_n) the **disjointification** of the sequence (A_n). Note that

$$\bigcup_{n=1}^{N} A_n = \bigcup_{n=1}^{N} D_n$$

for all $N \in \mathbb{N}$, and thus in particular, we have

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} D_n.$$

29.2.3 Equivalent Definitions For σ-Algebra

Proposition 29.5. *Let $\mathcal{A}$ be an algebra of subsets of X. Then $\mathcal{A}$ is a σ-algebra if and only if $\mathcal{A}$ is closed under ascending unions: if (A_n) is an ascending sequence of members of $\mathcal{A}$, then $\bigcup_{n=1}^\infty A_n \in \mathcal{A}$.*

Proof. Clearly, if $\mathcal{A}$ is a σ-algebra, then it is closed under ascending unions since it is closed under countable unions. Conversely, suppose $\mathcal{A}$ is an algebra and that is is closed under ascending unions. Let (A_n) be a sequence in $\mathcal{A}$. Let (B_n) be the ascendification of the sequence (A_n). Then we have

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n \in \mathcal{A}.$$

Thus $\mathcal{A}$ is closed under countable unions. It follows that $\mathcal{A}$ is a σ-algebra. □

29.2.4 Generating σ-Algebra from a Collection of Subsets

Proposition 29.6. *Let X be a set and let $\mathcal{C}$ be a nonempty collection of subsets of X. Then there exists a smallest σ-algebra which contains $\mathcal{C}$. It is called the **σ-algebra generated by** $\mathcal{C}$ and is denoted by $\sigma(\mathcal{C})$.*

Proof. Let $\mathscr{F}$ be the family of all σ-algebras $\mathcal{F}$ such that $\mathcal{F}$ contains $\mathcal{C}$. The family $\mathscr{F}$ is nonempty since the power set $\mathcal{P}(X)$ of X is a σ-algebra which contains $\mathcal{C}$. Define

$$\sigma(\mathcal{C}) := \bigcap_{\mathcal{F}\in\mathscr{F}} \mathcal{F}.$$

We claim that $\sigma(\mathcal{C})$ is the smallest σ-algebra which contains $\mathcal{C}$.

Let us first show that $\sigma(\mathcal{C})$ is a σ-algebra. Let (A_n) be a sequence of sets in $\sigma(\mathcal{C})$ and let A be a set in $\sigma(\mathcal{C})$. Then (A_n) is a sequence of sets in $\mathcal{F}$ and A is a set in $\mathcal{F}$ for all $\mathcal{F} \in \mathscr{F}$. Therefore $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ and $X\backslash A \in \mathcal{F}$ for all $\mathcal{F} \in \mathscr{F}$ (as each $\mathcal{F}$ is a σ-algebra). Therefore $\bigcup_{n=1}^{\infty} A_n \in \sigma(\mathcal{C})$ and $X\backslash A \in \sigma(\mathcal{C})$.

Now we will show that $\sigma(\mathcal{C})$ is the smallest algebra which contains $\mathcal{C}$. Suppose Σ' is a another σ-algebra which contains $\mathcal{C}$. Then $\Sigma' \in \mathscr{F}$, hence

$$\sigma(\mathcal{C}) \subseteq \bigcap_{\mathcal{F}\in\mathscr{F}} \mathcal{F} \subseteq \Sigma'.$$

□

Definition 29.4. The smallest σ-algebra containing $\mathcal{I}$ is called the **Borel σ-algebra** and the elements of this σ-algebra are called **Borel sets**.

29.3 Premeasures and Measures

Definition 29.5. Let $\mathcal{A}$ be an algebra of subsets of X and let $\mu\colon \mathcal{A} \to [0,\infty]$. We make the following definitions:

1. The function μ is **monotone** if $\mu(A) \leq \mu(B)$ whenever $A, B \in \mathcal{A}$ such that $A \subseteq B$. In this case, we say μ is **finite** if $\mu(X) < \infty$. Thus if μ is finite, we have $\mu(A) < \infty$ for all $A \in \mathcal{A}$ by monotonicity of μ.

2. The function μ is **finitely additive** if $\mu(\emptyset) = 0$ and

$$\mu(A \cup B) = \mu(A) + \mu(B)$$

for all $A, B \in \mathcal{A}$ such that $A \cap B = \emptyset$. In this case, μ is automatically monotone. Indeed, assume that $A, B \in \mathcal{A}$ such that $A \subseteq B$. Then

$$\begin{aligned}\mu(B) &= \mu((B\backslash A) \cup A) \\ &= \mu(B\backslash A) + \mu(A).\end{aligned}$$

Since $\mu(B\backslash A) \in [0,\infty]$, we conclude that $\mu(A) \leq \mu(B)$. If moreover $\mu(A) < \infty$, then we may subtract $\mu(A)$ from both sides to obtain

$$\mu(B\backslash A) = \mu(B) - \mu(A).$$

3. The function μ is **countably subadditive** if $\mu(\emptyset) = 0$ and

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n)$$

for all sequences (A_n) in $\mathcal{A}$ such that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

4. The function μ is **countably additive** if $\mu(\emptyset) = 0$ and

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

for all pairwise disjoint sequences (A_n) in $\mathcal{A}$ such that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$. In this case, we call the function μ a **premeasure** and we call the triple $(X, \mathcal{A}, \mu)$ a **premeasure space**. If $\mathcal{A}$ is a σ-algebra, then we call μ a **measure** and we call the triple $(X, \mathcal{A}, \mu)$ a **measure space**.

29.3.1 Equivalent Definitions for Premeasure

Proposition 29.7. *Let $\mathcal{A}$ be an algebra of subsets of X and let $\mu\colon \mathcal{A} \to [0,\infty]$. The following statements are equivalent.*

1. *μ is a premeasure;*
2. *μ is finitely additive and countably subadditive;*
3. *μ is finitely additive and satisfies*
$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n). \tag{66}$$
for all ascending sequences (A_n) in $\mathcal{A}$ such that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Proof. We first show 1 implies 2. Suppose that μ is a premeasure. Then $\mu(\varnothing) = 0$ by definition. Let $A, B \in \mathcal{A}$ such that $A \cap B = \varnothing$. Then set $A_1 = A$, $A_2 = B$, and $A_n = \varnothing$ for all $n \geq 3$. Then

$$\begin{aligned}
\mu(A \cup B) &= \mu\left(\bigcup_{n=1}^{\infty} A_n\right) \\
&= \sum_{n=1}^{\infty} \mu(A_n) \\
&= \mu(A_1) + \mu(A_2).
\end{aligned}$$

Thus μ is finitely additive. It is also countably subadditive. Indeed, let (A_n) be any sequence in $\mathcal{A}$. Disjointify the sequence (A_n) to the sequence (D_n): set $D_1 = A_1$ and $D_n = A_n \backslash \bigcup_{m=1}^{n-1} A_m$ for all $n > 1$. Then

$$\begin{aligned}
\mu\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mu\left(\bigcup_{n=1}^{\infty} D_n\right) \\
&= \sum_{n=1}^{\infty} \mu(D_n) \\
&\leq \sum_{n=1}^{\infty} \mu(A_n).
\end{aligned}$$

Thus μ is countably subadditive.

Now we show 2 implies 1. Suppose that μ is finitely additive and countably subadditive. Then $\mu(\varnothing) = 0$ by definition. Let (A_n) be a sequence of pairwise disjoint sets in $\mathcal{A}$ such that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$. Countable subadditivity of μ gives us
$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n).$$
For the reverse inequality, observe that for each $N \in \mathbb{N}$, finite additivity of μ gives us

$$\begin{aligned}
\mu\left(\bigcup_{n=1}^{\infty} A_n\right) &\geq \mu\left(\bigcup_{n=1}^{N} A_n\right) \\
&= \sum_{n=1}^{N} \mu(A_n).
\end{aligned}$$

Taking $N \to \infty$, we find that
$$\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) \geq \sum_{n=1}^{\infty} \mu^*(A_n).$$

Thus μ is a premeasure.

Now we will show 1 implies 3. Suppose μ is a premeasure. Let $A, B \in \mathcal{A}$ such that $A \cap B = \emptyset$. Then set $A_1 = A$, $A_2 = B$, and $A_n = \emptyset$ for all $n \geq 3$. Then

$$\begin{aligned}
\mu(A \cup B) &= \mu\left(\bigcup_{n=1}^{\infty} A_n\right) \\
&= \sum_{n=1}^{\infty} \mu(A_n) \\
&= \mu(A_1) + \mu(A_2).
\end{aligned}$$

Thus μ is finitely additive. Next let (A_n) be an ascending sequence in $\mathcal{A}$ such that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$. Disjointify (A_n) into the sequence (D_n): let $D_1 = A_1$ and $D_n = A_n \backslash \bigcup_{m=1}^{n-1} A_m$ for all $n \in \mathbb{N}$. Then we have

$$\begin{aligned}
\mu\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mu\left(\bigcup_{n=1}^{\infty} D_n\right) \\
&= \sum_{n=1}^{\infty} \mu(D_n) \\
&= \lim_{m\to\infty} \sum_{m=1}^{n} \mu(D_m) \\
&= \lim_{m\to\infty} \mu\left(\bigcup_{m=1}^{n} D_m\right) \\
&= \lim_{n\to\infty} \mu(A_n).
\end{aligned}$$

Therefore μ satisfies (66).

Now we will show 3 implies 1. Suppose μ is finitely additive and satisfied (66). Let (A_n) be a sequence of pairwise disjoint sets in $\mathcal{A}$ such that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$. Construct an ascending sequence (B_n) in $\mathcal{A}$ as follows: we set $B_1 = A_1$ and $B_n = \bigcup_{m=1}^{n} A_n$ for all $n \in \mathbb{N}$. Clearly $B_n \in \mathcal{A}$ for all $n \in \mathbb{N}$, and their union is equal to $\bigcup_{n=1}^{\infty} A_n$. Thus

$$\begin{aligned}
\mu\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mu\left(\bigcup_{n=1}^{\infty} B_n\right) \\
&= \lim_{n\to\infty} \mu(B_n) \\
&= \lim_{n\to\infty} \mu\left(\bigcup_{m=1}^{n} A_m\right) \\
&= \lim_{n\to\infty} \sum_{m=1}^{n} \mu(A_m) \\
&= \sum_{m=1}^{\infty} \mu(A_m).
\end{aligned}$$

Thus μ is a premeasure. □

29.3.2 Measure of descending sequences may not commute with limits

It seems reasonable to expect that, if (X, Σ, μ) is a measure space and (A_n) is a descending sequence of sets in Σ, then

$$\lim_{n\to\infty} \mu(A_n) = \mu\left(\bigcap_{n=1}^{\infty} A_n\right).$$

Unfortunately, this assertion can fail to be true. For example, consider the case where $X = \mathbb{Z}_{\geq 0}$, $\Sigma = \mathcal{P}(\mathbb{Z}_{\geq 0})$, and μ is the counting measure. Define $A_m := \{n \in \mathbb{Z}_{\geq 0} \mid n \geq m\}$ for all $m \in \mathbb{N}$. Then

$$\begin{aligned}\lim_{m\to\infty} \mu(A_m) &= \lim_{m\to\infty} \infty \\ &= \infty \\ &\neq 0 \\ &= \mu(\emptyset) \\ &= \mu\left(\bigcap_{m=1}^{\infty} A_m\right).\end{aligned}$$

On the other hand, there is a positive statement we can make about descending sequences. This is done in the following proposition.

Proposition 29.8. *Let (X, Σ, μ) be a measure space and let (A_n) be a descending sequence of sets in Σ such that $\mu(A_1) < \infty$. Then*

$$\lim_{n\to\infty} \mu(A_n) = \mu\left(\bigcap_{n=1}^{\infty} A_n\right).$$

Proof. The sequence $(A_1 \backslash A_n)$ is an ascending sequence, hence

$$\begin{aligned}\mu(A_1) - \lim_{n\to\infty} \mu(A_n) &= \lim_{n\to\infty} \mu(A_1) - \lim_{n\to\infty} \mu(A_n) \\ &= \lim_{n\to\infty} (\mu(A_1) - \mu(A_n)) \\ &= \lim_{n\to\infty} \mu(A_1 \backslash A_n) \\ &= \mu\left(\bigcup_{n=1}^{\infty} (A_1 \backslash A_n)\right) \\ &= \mu\left(A_1 \backslash \left(\bigcap_{n=1}^{\infty} A_n\right)\right) \\ &= \mu(A_1) - \mu\left(\bigcap_{n=1}^{\infty} A_n\right),\end{aligned}$$

where we used the fact that $\mu(A_1) < \infty$ to get from line 2 to line 3. Also since $\mu(A_1) < \infty$, we can subtract $\mu(A_1)$ from both sides to obtain

$$\lim_{n\to\infty} \mu(A_n) = \mu\left(\bigcap_{n=1}^{\infty} A_n\right).$$

□

29.3.3 Outer Measure

Definition 29.6. Let $\mu \colon \mathcal{P}(X) \to [0, \infty]$. We say μ is an **outer measure** if μ is monotone and countably subadditive.

Proposition 29.9. *Let $(X, \mathcal{A}, \mu)$ be a premeasure space. Define $\mu^* \colon \mathcal{P}(X) \to [0, \infty]$ by*

$$\mu^*(S) = \inf\left\{\sum_{n=1}^{\infty} \mu(A_n) \mid \{A_n\} \subseteq \mathcal{A} \text{ covers } S\text{, that is, } S \subseteq \bigcup_{n=1}^{\infty} A_n\right\}.$$

Then μ^* is an outer measure.

Proof. We first show μ^* is monotone. Let $S, T \in \mathcal{P}(X)$ such that $S \subseteq T$. Suppose that $\{A_n\} \subseteq \mathcal{A}$ covers T. Then $\{A_n\} \subseteq \mathcal{A}$ covers S too since $S \subseteq T$. Since the covering $\{A_n\}$ was arbitrary, we see that

$$\begin{aligned}\mu^*(S) &\leq \inf\left\{\sum_{n=1}^{\infty} \mu(A_n) \mid \{A_n\} \subseteq \mathcal{A} \text{ covers } T\right\} \\ &= \mu^*(T).\end{aligned}$$

Now we will show that μ^* is countably subadditive. First observe that $\{\emptyset\}$ is a covering of $\emptyset$. Thus

$$\begin{aligned}0 &\leq \mu^*(\emptyset)\\ &\leq \mu(\emptyset)\\ &= 0\end{aligned}$$

implies $\mu^*(\emptyset) = 0$. Now let (S_n) be a sequence in $\mathcal{P}(X)$ and let $\varepsilon > 0$. For each $n \in \mathbb{N}$ choose a covering $\{A_{n,k}\}_{k\in\mathbb{N}} \subseteq \mathcal{A}$ of S_n such that

$$\sum_{k=1}^{\infty} \mu(A_{n,k}) \leq \mu^*(S_n) + \frac{\varepsilon}{2^n}.$$

Then observe that $\{A_{n,k}\}_{k,n\in\mathbb{N}} \subseteq \mathcal{A}$ is a covering of $\bigcup_{n=1}^{\infty} S_n$, and so we have

$$\begin{aligned}\mu^*\left(\bigcup_{n=1}^{\infty} S_n\right) &\leq \sum_{n=1}^{\infty}\sum_{k=1}^{\infty} \mu(A_{n,k})\\ &\leq \sum_{n=1}^{\infty}\left(\mu^*(S_n) + \frac{\varepsilon}{2^n}\right)\\ &= \sum_{n=1}^{\infty} \mu^*(S_n) + \varepsilon.\end{aligned}$$

Taking $\varepsilon \to 0$ gives us our desired result. □

30 Extending Finite Premeasures

Throughout this section, let $(X, \mathcal{A}, \mu)$ be a finite premeasure space. Our goal in this section is to show that μ can be *uniquely* extended to a measure on $\sigma(\mathcal{A})$. In other words, there exists a measure $\widetilde{\mu}$ on $\supset(\mathcal{A})$ such that $\widetilde{\mu}|_{\mathcal{A}} = \mu$. Furthermore, if ν is any other measure on $\sigma(\mathcal{A})$ such that $\nu|_{\mathcal{A}} = \mu$, then $\widetilde{\mu} = \nu$. Let us state this as a theorem up front.

Theorem 30.1. *(Caratheodory, Hahn, Kolmogorov) Let $(X, \mathcal{A}, \mu)$ be a finite premeasure space. Then μ has a unique extension to a measure on $\sigma(\mathcal{A})$.*

Our proof of Theorem (30.1) will involve Topology, where much is known about unique extensions of continuous functions. In particular, we will realize $\mathcal{A}$ as a topological space and we will realize $\mu\colon \mathcal{A} \to [0, \infty)$ as a continuous function defined on $\mathcal{A}$. In this setting, it will be easy to show that μ can be uniquely extended to a continuous function on the larger space $\sigma(\mathcal{A})$.

30.1 Defining a Pseudometric on $\mathcal{P}(X)$

The first step is to construct a pseudometric on $\mathcal{P}(X)$.[15] In particular, we define $\mathrm{d}_\mu\colon \mathcal{P}(X) \times \mathcal{P}(X) \to [0, \infty]$ by

$$\mathrm{d}_\mu(A, B) = \mu^*(A \Delta B)$$

for all $A, B \in \mathcal{P}(X)$.

Proposition 30.1. d_μ *is a pseudometric on $\mathcal{P}(X)$.*

Proof. We first check reflexivity of d_μ. Let $A \in \mathcal{P}(X)$, then have

$$\begin{aligned}\mathrm{d}_\mu(A, A) &= \mu^*(A \Delta A)\\ &= \mu^*(\emptyset)\\ &= 0.\end{aligned}$$

Next we check symmetry of d_μ. Let $A, B \in \mathcal{P}(X)$. Then we have

$$\begin{aligned}\mathrm{d}_\mu(A, B) &= \mu^*(A \Delta B)\\ &= \mu^*(B \Delta A)\\ &= \mathrm{d}_\mu(B, A).\end{aligned}$$

[15] See the Appendix for more details on pseudometric spaces.

Finally, we check triangle inequality. Let $A, B, C \in \mathcal{P}(X)$. Then we have

$$\begin{aligned}
\mathrm{d}_\mu(A, C) &= \mu^*(A\Delta C) \\
&= \mu^*(A\Delta B\Delta B\Delta C) \\
&\leq \mu^*((A\Delta B) \cup (B\Delta C)) \\
&\leq \mu^*(A\Delta B) + \mu^*(B\Delta C) \\
&= \mathrm{d}_\mu(A, B) + \mathrm{d}_\mu(B, C),
\end{aligned}$$

where we obtained the third line from the second line by monotonicity of μ^*, and where we obtained the fourth line from the third line by finite subadditivity of μ^*. □

30.1.1 Metric Space Induced by Pseudometric Space

Proposition (30.1) tells us that $(\mathcal{P}(X), \mathrm{d}_\mu)$ is a pseudometric space. The reason d_μ is a pseudometric and not a metric is because we not have identity of indiscernibles: we may have $\mu^*(A\Delta B) = 0$ with $A \neq B$. All is not lost however as every pseudometric space induces a metric space in a natural way. Let us briefly describe the metric space induced by the pseudometric space $(\mathcal{P}(X), \mathrm{d}_\mu)$. More details can be found in the Appendix. We introduce an equivalence relation $\sim$ on $\mathcal{P}(X)$ as follows: let $A, B \in \mathcal{P}(X)$. Then

$$A \sim B \text{ if and only if } \mathrm{d}_\mu(A, B) = 0.$$

One checks that $\sim$ is an equivalence relation on $\mathcal{P}(X)$ and so we may consider quotient space

$$[\mathcal{P}(X)] := \mathcal{P}(X)/\sim .$$

We shall use the notation $[A]$ to denote a coset in $[A]$ with $A \in \mathcal{P}(X)$ as a particular representative. We define a metric $[\mathrm{d}_\mu]$ on $[\mathcal{P}(X)]$ by

$$[\mathrm{d}_\mu]([A], [B]) = \mathrm{d}_\mu(A, B) \tag{67}$$

One checks that (67) is well-defined and satisfies all of the properties required for it to be a metric. Furthermore, one shows that the quotient topology on $[\mathcal{P}(X)]$ is the same as the topology induced by the metric $[\mathrm{d}_\mu]$. In particular, the projection map

$$\pi \colon \mathcal{P}(X) \to [\mathcal{P}(X)]$$

is continuous, and for any topological space Y (such as $[0, \infty]$!) we have a bijection

$$\left\{ \begin{array}{c} \text{continuous functions from } \mathcal{P}(X) \text{ to } Y \\ \text{which are constant on equivalence classes} \end{array} \right\} \cong \{\text{continuous functions from } [\mathcal{P}(X)] \text{ to } Y\} .$$

Indeed, if $\nu \colon [\mathcal{P}(X)] \to Y$ is continuous, then the function $\nu \circ \pi \colon \mathcal{P}(X) \to Y$ is continuous since it is a composition of continuous functions and it is constant on equivalence classes: if $A \sim B$, then

$$\begin{aligned}
(\nu \circ \pi)(A) &= \nu(\pi(A)) \\
&= \nu(\pi(B)) \\
&= (\nu \circ \pi)(B).
\end{aligned}$$

Convsersely, if $\eta \colon \mathcal{P}(X) \to Y$ is continuous and constant on equivalence classes, then it induces a unique continuous function $\nu \colon [\mathcal{P}(X)] \to Y$ such that $\nu \circ \pi = \eta$.

There many other properties which are both shared by $(\mathcal{P}(X), \mathrm{d}_\mu)$ and $([\mathcal{P}(X)], [\mathrm{d}_\mu])$. For instance, $(\mathcal{P}(X), \mathrm{d}_\mu)$ is complete if and only if $([\mathcal{P}(X)], [\mathrm{d}_\mu])$ complete. For this and many other reasons, we choose to work in the pseudometric space $(\mathcal{P}(X), \mathrm{d}_\mu)$ rather than the metric space $([\mathcal{P}(X)], [\mathrm{d}_\mu])$.

30.1.2 Complement Map is Isometry

Proposition 30.2. *The complement map $-^c \colon \mathcal{P}(X) \to \mathcal{P}(X)$ given by*

$$-^c(A) = A^c$$

for all $A \in \mathcal{P}(X)$ is an isometry on $[\mathcal{P}(X)]$.

Proof. We first check that $-^c$ is constant on equivalence classes. Suppose $A, A' \in \mathcal{P}(X)$ with $A \sim A'$ (so $A \Delta A' = \emptyset$). Then

$$\begin{aligned} A^c \Delta A'^c &= A \Delta A \\ &= \emptyset. \end{aligned}$$

Thus $A^c \sim A'^c$, and so the complement map is constant on equivalence classes. Now we check that it is an isometry. Let $A, B \in \mathcal{P}(X)$. Then

$$\begin{aligned} \mathrm{d}_\mu(A, B) &= \mu^*(A \Delta B) \\ &= \mu^*(A^c \Delta B^c) \\ &= \mathrm{d}_\mu(A^c, B^c). \end{aligned}$$

Thus $-^c$ is an isometry. □

30.1.3 Continuity of Finite Unions and Finite Intersections

Proposition 30.3. *The union map* $\cup \colon \mathcal{P}(X) \times \mathcal{P}(X) \to \mathcal{P}(X)$*, defined by*

$$\cup(A, B) = A \cup B$$

for all $(A, B) \in \mathcal{P}(X)$*, is continuous on* $[\mathcal{P}(X)] \times [\mathcal{P}(X)]$*.*

Proof. We first check that the union map is constant on equivalence classes. Suppose $A \sim A'$ and $B \sim B'$ where $A, A', B, B' \in \mathcal{P}(X)$. Thus $A \Delta A' = 0$ and $B \Delta B' = 0$. Then

$$\begin{aligned} (A \cup B) \Delta (A' \cup B') &\subseteq (A \Delta A') \cup (B \Delta B') \\ &= \emptyset \cup \emptyset \\ &= \emptyset. \end{aligned}$$

It follows that $A \cup B \sim A' \cup B'$. Now we will show that the union map is continuous. Suppose $A_n \to A$ and $B_n \to B$ in $(\mathcal{P}(X), \mathrm{d}_\mu)$. Let $\varepsilon > 0$ and choose $N \in \mathbb{N}$ such that

$$\mathrm{d}_\mu(A_n, A) < \frac{\varepsilon}{2} \quad \text{and} \quad \mathrm{d}_\mu(B_n, B) < \frac{\varepsilon}{2}.$$

Then $n \geq N$ implies

$$\begin{aligned} \mathrm{d}_\mu((A_n \cup B_n), (A \cup B)) &= \mu^*((A_n \cup B_n) \Delta (A \cup B)) \\ &= \mu^*((A_n \cup B_n) \Delta (A \cup B)) \\ &\leq \mu^*((A_n \Delta A) \cup (B_n \Delta B)) \\ &\leq \mu^*(A_n \Delta A) + \mu^*(B_n \Delta B) \\ &< \mathrm{d}_\mu(A_n, A) + \mathrm{d}_\mu(B_n, B) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon. \end{aligned}$$

It follows that the union map is continuous on $[\mathcal{P}(X)] \times [\mathcal{P}(X)]$. □

Proposition 30.4. *The intersection map* $\cap \colon \mathcal{P}(X) \times \mathcal{P}(X) \to \mathcal{P}(X)$*, defined by*

$$\cap(A, B) = A \cap B$$

for all $(A, B) \in \mathcal{P}(X)$*, is continuous on* $[\mathcal{P}(X)] \times [\mathcal{P}(X)]$*.*

Proof. The intersection is a composition of the union map with the complement map. Thus it is a composition of continuous functions, and hence must be continuous. □

30.1.4 Uniform Continuity of μ^*

Proposition 30.5. *The function $\mu^*: \mathcal{P}(X) \to [0,\infty]$ is Lipschitz continuous on $[\mathcal{P}(X)]$.*

Proof. Let $A, B \in \mathcal{P}(X)$. Then

$$\begin{aligned}|\mu^*(A) - \mu^*(B)| &\leq \max\{\mu^*(A\backslash B), \mu^*(B\backslash A)\} \\ &\leq \mu^*((A\backslash B) \cup (B\backslash A)) \\ &= \mu^*(A\Delta B) \\ &= \mathrm{d}_\mu(A,B).\end{aligned}$$

It follows that μ^* is Lipschitz continuous on $\mathcal{P}(X)$. To see that it is Lipschitz continuous on $[\mathcal{P}(X)]$, we just need to check that it is constant on equivalence classes. Let $A, A' \in \mathcal{P}(X)$ such that $A \sim A'$ (so $\mu^*(A\Delta A') = 0$). Then

$$\begin{aligned}\mu^*(A') &= \mu^*((A\Delta A)\Delta A') \\ &= \mu^*(A\Delta(A\Delta A')) \\ &\leq \mu^*(A \cup (A\Delta A')) \\ &\leq \mu^*(A) + \mu^*(A\Delta A') \\ &= \mu^*(A).\end{aligned}$$

By a similar argument, we also have $\mu^*(A) \geq \mu^*(A')$. Thus μ^* is constant on equivalence classes. □

Remark 48. Since $\mu^*|_{\mathcal{A}} = \mu$, we have also shown that μ is continuous on $[\mathcal{A}]$ and that μ^* is a continuous extension of μ.

30.2 Completion of $(\mathcal{A}, \mathrm{d}_\mu)$

The pseudometric d_μ on $\mathcal{P}(X)$ restricts to a pseudometric on $\sigma(\mathcal{A})$ and it also restricts to a pseudometric on $\mathcal{A}$. Formally, we should denote these restrictions by $\mathrm{d}_\mu|_{\sigma(\mathcal{A})\times\sigma(\mathcal{A})}$ and $\mathrm{d}_\mu|_{\mathcal{A}\times\mathcal{A}}$ respectively, however in order to clean notation, we will simply denote these restricts by d_μ. Thus we have the following inclusions of pseudometric spaces

$$(\mathcal{A}, \mathrm{d}_\mu) \subseteq (\sigma(\mathcal{A}), \mathrm{d}_\mu) \subseteq (\mathcal{P}(X), \mathrm{d}_\mu).$$

Our goal in this subsection is to show that $(\mathcal{A}, \mathrm{d}_\mu)$ is dense in $(\sigma(\mathcal{A}), \mathrm{d}_\mu)$ and that $(\sigma(\mathcal{A}), \mathrm{d}_\mu)$ is complete. Thus $(\sigma(\mathcal{A}), \mathrm{d}_\mu)$ is the completion of $(\mathcal{A}, \mathrm{d}_\mu)$. Before doing this, we introduce the limsup and liminf of a sequence of sets.

30.2.1 Limit Supremum and Limit Infimum

Recall that if (a_n) is a sequence of real numbers, we define its limit supremum by the formula

$$\limsup a_n = \inf_{N\geq 1} \sup_{n\geq N} a_n.$$

Similarly, we define its limit infimum by the formula

$$\liminf a_n = \sup_{N\geq 1} \inf_{n\geq N} a_n.$$

There is an analagous notion of limit supremum and limit infimum of a sequence of sets.

Definition 30.1. Let (A_n) be a sequence of sets. The **limit supremum** of (A_n) is defined to be

$$\limsup A_n = \bigcap_{N\geq 1} \bigcup_{n\geq N} A_n.$$

The **limit infimum** of (A_n) is defined by

$$\liminf A_n = \bigcup_{N\geq 1} \bigcap_{n\geq N} A_n.$$

30.2.2 $(\sigma(\mathcal{A}), \mathrm{d}_\mu)$ is complete

We now want to show that $(\sigma(\mathcal{A}), \mathrm{d}_\mu)$ is complete. In fact, we will do better than this. We will show that (Σ, d_μ) is complete, where Σ is any σ-algebra which contains $\mathcal{A}$. Recall that this means that every Cauchy sequence in (Σ, d_μ) converges to a limit in (Σ, d_μ) . Before we prove this, we need to establish two lemmas.

Lemma 30.2. *Let (X, d) be a pseudometric space and let (x_n) be a Cauchy sequence in X. Suppose there exists a subsequence $(x_{\pi(n)})$ of the sequence (x_n) such that $x_{\pi(n)} \to x$ for some $x \in X$. Then $x_n \to x$.*

Proof. Let $\varepsilon > 0$. Since $(x_{\pi(n)})$ is convergent, there exists an $N \in \mathbb{N}$ such that $\pi(n) \geq N$ implies

$$\mathrm{d}(x_{\pi(n)}, x) < \frac{\varepsilon}{2}.$$

Since (x_n) is Cauchy, there exists $M \in \mathbb{N}$ such that $m, n \geq M$ implies

$$\mathrm{d}(x_m, x_n) < \frac{\varepsilon}{2}.$$

Choose such M and N and assume without loss of generality that $N \geq M$. Then $\pi(n) \geq n \geq N$ implies

$$\begin{aligned}
\mathrm{d}(x_n, x) &\leq \mathrm{d}(x_{\pi(n)}, x_n) + \mathrm{d}(x_{\pi(n)}, x) \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\
&= \varepsilon.
\end{aligned}$$

It follows that $x_n \to x$. □

Lemma 30.3. *Let (A_n) be a sequence in $\mathcal{P}(X)$. Then*

$$\bigcup_{n=1}^{\infty} (A_n \Delta A_{n+1}) = \bigcup_{n=1}^{\infty} A_n \backslash \bigcap_{m=1}^{\infty} A_m.$$

Proof. Suppose $x \in \bigcup_{n=1}^{\infty} (A_n \Delta A_{n+1})$. Choose $n \in \mathbb{N}$ such that $x \in A_n \Delta A_{n+1}$. Thus either $x \in A_n \backslash A_{n+1}$ or $x \in A_{n+1} \backslash A_n$. Without loss of generality, say $x \in A_n \backslash A_{n+1}$. Then since $x \in A_n$, we see that $x \in \bigcup_{n=1}^{\infty} A_n$ and since $x \notin A_{n+1}$, we see that $x \notin \bigcap_{n=1}^{\infty} A_n$. Therefore $x \in \bigcup_{n=1}^{\infty} A_n \backslash \bigcap_{m=1}^{\infty} A_m$. This implies

$$\bigcup_{n=1}^{\infty} (A_n \Delta A_{n+1}) \subseteq \bigcup_{n=1}^{\infty} A_n \backslash \bigcap_{m=1}^{\infty} A_m.$$

Conversely, suppose $x \in \bigcup_{n=1}^{\infty} A_n \backslash \bigcap_{m=1}^{\infty} A_m$. Since $x \in \bigcup_{n=1}^{\infty} A_n$, there exists some $n \in \mathbb{N}$ such that $x \in A_n$. Since $x \notin \bigcap_{m=1}^{\infty} A_m$, there exists some $k \in \mathbb{N}$ such that $x \notin A_k$. Assume without loss of generality that $k < n$. Choose m to be the least natural number number such that $x \in A_m$, $x \notin A_{m-1}$, and $k < m \leq n$. Clearly this number exists since $x \notin A_k$ and $x \in A_n$. Then $x \in A_m \Delta A_{m-1}$, which implies $x \in \bigcup_{n=1}^{\infty} (A_n \Delta A_{n+1})$. Thus

$$\bigcup_{n=1}^{\infty} (A_n \Delta A_{n+1}) \supseteq \bigcup_{n=1}^{\infty} A_n \backslash \bigcap_{m=1}^{\infty} A_m.$$

□

Theorem 30.4. *Let Σ be any σ-algebra which contains $\mathcal{A}$. Then (Σ, d_μ) is complete.*

Proof. Let (A_n) be a Cauchy sequence in (Σ, d_μ) . To show that (A_n) is convergent, it suffices to show that a subsequence of (A_n) is convergent, by Lemma (39.1). We construct such a subsequence as follows: Since (A_n) is Cauchy, for each $n \geq 1$ there exists a $\pi(n) \geq n$ such that $k, m \geq \pi(n)$ implies

$$\mathrm{d}_\mu(A_k, A_m) < \frac{1}{2^n}.$$

Choose such $\pi(n) \geq n$ such that $m < n$ implies $\pi(m) < \pi(n)$. In particular, for each $n \geq 1$, we have

$$\mathrm{d}_\mu(A_{\pi(n)}, A_{\pi(n+1)}) < \frac{1}{2^n}. \tag{68}$$

By passing to the subsequence $(A_{\pi(n)})$ of (A_n) if necessary, we may as well assume that

$$\mathrm{d}_\mu(A_n, A_{n+1}) < \frac{1}{2^n}$$

for all $n \geq 1$. Now set $A = \limsup A_n$. We claim that $A_n \to A$. Indeed, let $\varepsilon > 0$ and choose $N \geq 1$ such that $2^{2-N} < \varepsilon$. Then $n \geq N$ implies

$$\begin{aligned}
\mathrm{d}_\mu(A, A_n) &= \mu^*(A \Delta A_n) \\
&\leq \mu^*(A \backslash A_n) + \mu^*(A_n \backslash A) && \text{finite subadditivity of } \mu^* \\
&\leq \mu^*\left(\bigcup_{m \geq N} A_m \backslash \bigcap_{m \geq N} A_m\right) + \mu^*\left(\bigcup_{m \geq N} A_m \backslash \bigcap_{m \geq N} A_m\right) && \text{monotonicity of } \mu^* \\
&= 2\mu^*\left(\bigcup_{m \geq N} A_m \backslash \bigcap_{m \geq N} A_m\right) \\
&= 2\mu^*\left(\bigcup_{m \geq N} (A_m \Delta A_{m+1})\right) \\
&\leq 2 \sum_{m \geq N} \mu^*(A_m \Delta A_{m+1}) \\
&= 2 \sum_{m \geq N} \mathrm{d}_\mu(A_m \Delta A_{m+1}) \\
&< 2 \sum_{m \geq N} \frac{1}{2^m} \\
&= 2^{2-N} \\
&< \varepsilon.
\end{aligned}$$

It follows that (A_n) converges to A. Finally, note that

$$\begin{aligned}
A &= \limsup A_n \\
&= \bigcap_{N \geq 1} \bigcup_{n \geq N} A_n \\
&\in \Sigma.
\end{aligned}$$

Thus every Cauchy sequence of sets in (Σ, d_μ) converges to a set in (Σ, d_μ). Therefore (Σ, d_μ) is complete. $\square$

30.2.3 $(\mathcal{A}, \mathrm{d}_\mu)$ is dense in $(\sigma(\mathcal{A}), \mathrm{d}_\mu)$

Now we want to show that $(\mathcal{A}, \mathrm{d}_\mu)$ is dense subset in $(\sigma(\mathcal{A}), \mathrm{d}_\mu)$. In other words, we want to show that $\overline{\mathcal{A}} = \sigma(\mathcal{A})$, where $\overline{\mathcal{A}}$ denotes the closure of $\mathcal{A}$ in the pseudometric space $(\mathcal{P}(X), \mathrm{d}_\mu)$. Recall that $\overline{\mathcal{A}}$ is the *smallest* closed set which contains $\mathcal{A}$. Since $\sigma(\mathcal{A})$ is complete, it is certainly closed in $(\mathcal{P}(X), \mathrm{d}_\mu)$, and so we have

$$\overline{\mathcal{A}} \subseteq \sigma(\mathcal{A}).$$

On the other hand, $\sigma(\mathcal{A})$ is the *smallest* σ-algebra which contains $\mathcal{A}$. Thus if we can show that $\overline{\mathcal{A}}$ is a σ-algebra, then we will have the reverse inclusion.

Proposition 30.6. *$\overline{\mathcal{A}}$ is a σ-algebra.*

Proof. The proof consists of three steps.

Step 1: We first show that $\overline{\mathcal{A}}$ is an algebra. We have $\emptyset \in \mathcal{A} \subseteq \overline{\mathcal{A}}$. Next, let $A \in \overline{\mathcal{A}}$. Choose a sequence (A_n) in $\mathcal{A}$ such that $A_n \to A$. Then since taking complements is continuous, we have $A_n^c \to A^c$, which implies $A^c \in \overline{\mathcal{A}}$. Finally, let $A, B \in \overline{\mathcal{A}}$. Choose sequences (A_n) and (B_n) in $\mathcal{A}$ such that $A_n \to A$ and $B_n \to B$. Then since taking unions is continuous, we have $A_n \cup B_n \to A \cup B$, which implies $A \cup B \in \overline{\mathcal{A}}$.

Step 2: We show that μ^* restricts to a measure on $\overline{\mathcal{A}}$. It suffices to show that μ^* is finitely additive on $\overline{\mathcal{A}}$ since we already know it is countably subadditive. Let A and B be two disjoint members of $\overline{\mathcal{A}}$ and choose sequences

(A_n) and (B_n) in $\mathcal{A}$ such that $A_n \to A$ and $B_n \to B$. Then $A_n \cup B_n \to A \cup B$, and so it follows by continuity of μ^* that

$$\begin{aligned}
\mu^*(A \cup B) &= \lim_{n\to\infty} \mu^*(A_n \cup B_n) \\
&= \lim_{n\to\infty} \mu(A_n \cup B_n) \\
&= \lim_{n\to\infty} (\mu(A_n) + \mu(B_n) - \mu(A_n \cap B_n)) \\
&= \lim_{n\to\infty} (\mu^*(A_n) + \mu^*(B_n) - \mu^*(A_n \cap B_n)) \\
&= \mu^*(A) + \mu^*(B) - \mu^*(A \cap B) \\
&= \mu^*(A) + \mu^*(B) - \mu^*(\emptyset) \\
&= \mu^*(A) + \mu^*(B).
\end{aligned}$$

where we needed to use the fact that μ is a finite measure in order to get the third line from the second line.

Step 3: We now show that $\overline{\mathcal{A}}$ is a σ-algebra. Let (A_n) be a sequence of sets in $\overline{\mathcal{A}}$. We want to show $\bigcup_{n=1}^{\infty} A_n \in \overline{\mathcal{A}}$. To do this, we will shows that $\bigcup_{n=1}^{N} A_n \to \bigcup_{n=1}^{\infty} A_n$ as $N \to \infty$. Then since $\bigcup_{n=1}^{N} A_n \in \overline{\mathcal{A}}$ since $\overline{\mathcal{A}}$ is an algebra by step 2, it will then follow that $\bigcup_{n=1}^{\infty} A_n \in \overline{\mathcal{A}}$.

Disjointify the sequence (A_n) to the sequence (B_n): set $B_1 = A_1$ and $B_n = A_n \backslash \bigcup_{i=1}^{n-1} A_i$ for all $n > 1$. Then (B_n) is a sequence of pairwise disjoint sets in $\overline{\mathcal{A}}$ since $\overline{\mathcal{A}}$ is an algebra by step 2. Moreover, for each $N \in \mathbb{N}$, we have

$$\begin{aligned}
\mathrm{d}_\mu\left(\bigcup_{n=1}^{\infty} A_n, \bigcup_{n=1}^{N} A_n\right) &= \mathrm{d}_\mu\left(\bigcup_{n=1}^{\infty} B_n, \bigcup_{n=1}^{N} B_n\right) \\
&= \mu^*\left(\left(\bigcup_{n=1}^{\infty} B_n\right) \Delta \left(\bigcup_{n=1}^{N} B_n\right)\right) \\
&= \mu^*\left(\bigcup_{n=N+1}^{\infty} B_n\right) \\
&= \sum_{n=N+1}^{\infty} \mu^*(B_n),
\end{aligned}$$

where the last term tends to 0 as $N \to \infty$ since it is the tail of a convergent series:

$$\begin{aligned}
\sum_{n=1}^{\infty} \mu^*(B_n) &= \mu^*\left(\bigcup_{n=1}^{\infty} B_n\right) \\
&\leq \mu^*(X) \\
&< \infty.
\end{aligned}$$

Therefore $\bigcup_{n=1}^{N} A_n \to \bigcup_{n=1}^{\infty} A_n$ as $N \to \infty$. □

30.3 Uniqueness of Measure

Recall that in the proof of Proposition (30.6) we showed that μ^* restricts to a measure on $\overline{\mathcal{A}} = \sigma(\mathcal{A})$. In particular, since $\mu^*|_{\mathcal{A}} = \mu$, we see that the measure $\mu^*|_{\sigma(\mathcal{A})}$ is an extension of μ to all of $\sigma(\mathcal{A})$. This gives us precisely one example of a measure which extends μ to all of $\sigma(\mathcal{A})$. It is not at all clear however that this is the *only* such extension. If μ is not finite, then there may be more than one extension. Indeed, see problem 6 in homework 2 for an example of this. In our case however where μ is finite, we shall see that this extension is in fact unique. We will use ideas from topology to show this; all of the topological work we just did is about to payoff!

30.3.1 Uniqueness Extensions of Continuous Functions

We first begin with a result from topology.

Proposition 30.7. *Let X be a topological space, let A be a dense subspace of X, let Y be a Hausdorff space, and let $f\colon A \to Y$ be a continuous function. If there exists a continuous extension of f to all of X, then it must be unique, that is, if $\widetilde{f}_1\colon X \to Y$ and $\widetilde{f}_2\colon X \to Y$ are two continuous functions such that*

$$\widetilde{f}_1|_A = f = \widetilde{f}_2|_A,$$

then $\widetilde{f}_1 = \widetilde{f}_2$.

Proof. Assume for a contradiction that $\widetilde{f}_1\colon X \to Y$ and $\widetilde{f}_2\colon X \to Y$ are two continuous extensions of f such that $\widetilde{f}_1 \neq \widetilde{f}_2$. Choose $x \in X$ such that $\widetilde{f}_1(x) \neq \widetilde{f}_2(x)$. Since Y is Hausdorff, we may choose open neighborhoods V_1 and V_2 of $\widetilde{f}_1(x)$ and $\widetilde{f}_2(x)$ respectively such that $V_1 \cap V_2 = \emptyset$. Then $\widetilde{f}_1^{-1}(V_1) \cap \widetilde{f}_2^{-1}(V_2)$ is an open neighborhood of x, and so it must have a nonempty intersection with A. Choose $a \in A \cap \widetilde{f}_1^{-1}(V_1) \cap \widetilde{f}_2^{-1}(V_2)$. Then

$$f(a) = \widetilde{f}_1(a) \in V_1.$$

Similarly,

$$f(a) = \widetilde{f}_2(a) \in V_2.$$

Thus $f(a) \in V_1 \cap V_2$, which is a contradiction since V_1 and V_2 were chosen to disjoint from one another. □

30.3.2 Continuity of Finite Measure

Lemma 30.5. *Let $\mathcal{A}$ be an algebra and let μ be a measure on $\sigma(\mathcal{A})$. Then*

$$(\mu|_{\mathcal{A}})^*(A) \geq \mu(A)$$

for all $A \in \sigma(\mathcal{A})$.

Proof. Let $A \in \sigma(\mathcal{A})$. Then

$$\begin{aligned}
(\mu|_{\mathcal{A}})^*(A) &= \inf\left\{ \sum_{n=1}^{\infty} (\mu|_{\mathcal{A}})(E_n) \mid (E_n) \text{ is a sequence in } \mathcal{A} \text{ such that } A \subseteq \bigcup_{n=1}^{\infty} E_n \right\} \\
&= \inf\left\{ \sum_{n=1}^{\infty} \mu(E_n) \mid (E_n) \text{ is a sequence in } \mathcal{A} \text{ such that } A \subseteq \bigcup_{n=1}^{\infty} E_n \right\} \\
&\geq \inf\left\{ \mu\left(\bigcup_{n=1}^{\infty} E_n \right) \mid (E_n) \text{ is a sequence in } \mathcal{A} \text{ such that } A \subseteq \bigcup_{n=1}^{\infty} E_n \right\} \\
&\geq \mu(A),
\end{aligned}$$

where we used countable subadditivity of μ to get from the second line to the third line, and where we used monotonicity of μ to get from the third line to the fourth line. □

Proposition 30.8. *Let $\mathcal{A}$ be an algebra and let μ be a finite measure on $\sigma(\mathcal{A})$. Then μ is Lipschitz continuous with respect to $\mathrm{d}_{\mu|_{\mathcal{A}}}$.*

Proof. Let $A, B \in \sigma(\mathcal{A})$. Assume without loss of generality that $\mu(A) \geq \mu(B)$. Then

$$\begin{aligned}
\mu(A) - \mu(B) &\leq \mu(A \backslash B) \\
&\leq \mu((A \backslash B) \cup (B \backslash A)) \\
&= \mu(A \Delta B) \\
&\leq (\mu|_{\mathcal{A}})^*(A \Delta B) \\
&= \mathrm{d}_{\mu|_{\mathcal{A}}}(A, B),
\end{aligned}$$

where we used the fact that μ is finite in the first line. □

30.3.3 Uniqueness of Extension for Finite Measures

Theorem 30.6. *Let μ and ν be two finite measures defined on $\sigma(\mathcal{A})$ which coincide on $\mathcal{A}$. Then $\mu = \nu$.*

Proof. We first note that $\mathrm{d}_{\mu|_{\mathcal{A}}} = \mathrm{d}_{\nu|_{\mathcal{A}}}$ since μ and ν agree on $\mathcal{A}$. Indeed, let $A, B \in \sigma(\mathcal{A})$. Then we have

$$\begin{aligned}
\mathrm{d}_{\mu|_{\mathcal{A}}}(A, B) &= (\mu|_{\mathcal{A}})^*(A\Delta B) \\
&= \inf\left\{ \sum_{n=1}^{\infty} (\mu|_{\mathcal{A}})(E_n) \mid (E_n) \text{ is a sequence in } \mathcal{A} \text{ such that } A\Delta B \subseteq \bigcup_{n=1}^{\infty} E_n \right\} \\
&= \inf\left\{ \sum_{n=1}^{\infty} (\nu|_{\mathcal{A}})(E_n) \mid (E_n) \text{ is a sequence in } \mathcal{A} \text{ such that } A\Delta B \subseteq \bigcup_{n=1}^{\infty} E_n \right\} \\
&= (\nu|_{\mathcal{A}})^*(A\Delta B) \\
&= \mathrm{d}_{\nu|_{\mathcal{A}}}(A, B).
\end{aligned}$$

Therefore $\mathrm{d}_{\mu|_{\mathcal{A}}}$ and $\mathrm{d}_{\nu|_{\mathcal{A}}}$ induce a common topology on $\sigma(\mathcal{A})$. Both $\mu\colon \sigma(\mathcal{A}) \to [0, \infty]$ and $\nu\colon \sigma(\mathcal{A}) \to [0, \infty]$ are continuous extensions of $\mu|_{\mathcal{A}} = \nu|_{\mathcal{A}}$ with respect to this common topology by Proposition (30.8). Since $[0, \infty]$ is Hausdorff and since $\mathcal{A}$ is dense in $\sigma(\mathcal{A})$ with respect to this common topology, it follows from Proposition (30.7) that $\mu = \nu$. □

30.4 Measurability

We've just shown that any finite premeasure space $(X, \mathcal{A}, \mu)$ can be uniquely extended to the finite measure space $(X, \sigma(\mathcal{A}), \mu^*|_{\sigma(\mathcal{A})})$. In fact, since $\mu^*|_{\sigma(\mathcal{A})}$ is unique, we may as well simply denote it by μ. It turns out that $(X, \mathcal{A}, \mu)$ can be uniquely extended to an even bigger measure space $(X, \mathcal{M}, \mu^*|_{\mathcal{M}})$. Let us define what this bigger measure space is.

Definition 30.2. Let μ be an outer measure on a set X. A subset A of X is said to be μ-**measurable** if

$$\mu^*(S) \geq \mu^*(S \cap A) + \mu^*(S \backslash A) \tag{69}$$

holds for every subset S of X. Note that (69) is equivalent to the equation

$$\mu^*(S) = \mu^*(S \cap A) + \mu^*(S \backslash A)$$

since μ^* is finitely subadditive. Note also that (69) holds for every subset S of X is equivalent to the assertion that this inequality holds for every subset S of X for which $\mu^*(S) < \infty$. We denote by $\mathcal{M}_\mu$ (or more simply just $\mathcal{M}$ if μ is clear from context) to be the set of all μ-measurable sets.

30.4.1 $\mathcal{M}_{\mu^*}$ contains $\mathcal{A}$

Proposition 30.9. *Let $(X, \mathcal{A}, \mu)$ be a premeasure space. Then $\mathcal{A} \subseteq \mathcal{M}_{\mu^*}$.*

Proof. Let $A \in \mathcal{A}$, let $S \in \mathcal{P}(X)$, and let $\varepsilon > 0$. Choose a covering $\{A_n\} \in \mathcal{A}$ of S such that

$$\sum_{n=1}^{\infty} \mu(A_n) \leq \mu^*(S) + \varepsilon.$$

Then $\{A_n \cap A\}$ is a covering of $S \cap A$ and $\{A_n \backslash A\}$ is a covering of $S \backslash A$, and so

$$\begin{aligned}
\mu^*(S) &\geq \sum_{n=1}^{\infty} \mu(A_n) - \varepsilon \\
&= \sum_{n=1}^{\infty} \mu((A_n \cap A) \cup (A_n \backslash A)) - \varepsilon \\
&= \sum_{n=1}^{\infty} \mu(A_n \cap A) + \sum_{n=1}^{\infty} \mu(A_n \backslash A) - \varepsilon \\
&\geq \mu^*(S \cap A) + \mu^*(S \backslash A) - \varepsilon.
\end{aligned}$$

Taking $\varepsilon \to 0$ gives us

$$\mu^*(S) \geq \mu^*(S \cap A) + \mu^*(S \backslash A).$$

Therefore A is μ^*-measurable since S was arbitrary. $\square$

30.4.2 $(X, \mathcal{M}_{\mu^*}, \mu^*|_{\mathcal{M}_{\mu^*}})$ is a measure space

Proposition 30.10. *Let $(X, \mathcal{A}, \mu)$ be a premeasure space. Then $(X, \mathcal{M}_{\mu^*}, \mu^*|_{\mathcal{M}_{\mu^*}})$ is a measure space.*

Proof. To clean notation in what follows, we denote $\mathcal{M} = \mathcal{M}_{\mu^*}$ and $\mu^* = \mu^*|_{\mathcal{M}_{\mu^*}}$. We prove this proposition in several steps:

Step 1: We first show $\mathcal{M}$ is an algebra. First we show it is closed under finite unions. Let $A, B \in \mathcal{M}$ and let $S \in \mathcal{P}(X)$. Then

$$\begin{aligned}
\mu^*(S) &= \mu^*\left(S \cap A\right) + \mu^*\left(S \backslash A\right) \\
&= \mu^*\left(S \cap A\right) + \mu^*\left((S \backslash A) \cap B\right) + \mu^*\left((S \backslash A) \backslash B\right) \\
&\geq \mu^*\left((S \cap A) \cup ((S \backslash A) \cap B)\right) + \mu^*\left((S \backslash A) \backslash B\right) \\
&= \mu^*\left(S \cap (A \cup B)\right) + \mu^*\left(S \backslash (A \cup B)\right)
\end{aligned}$$

Therefore $A \cap B \in \mathcal{M}$.

Next we shows it is closed under complements. Let $A \in \mathcal{M}$ and let $S \in \mathcal{P}(X)$. Then

$$\begin{aligned}
\mu^*(S) &\geq \mu^*(S \cap A) + \mu^*(S \backslash A) \\
&= \mu^*\left(S \backslash (X \backslash A)\right) + \mu^*(S \backslash A) \\
&= \mu^*\left(S \backslash (X \backslash A)\right) + \mu^*\left(S \cap (X \backslash A)\right).
\end{aligned}$$

Therefore $X \backslash A \in \mathcal{M}$.

Step 2: We show that μ^* restricts to a measure on $\mathcal{M}$. To do this, we just need to show that μ^* is finitely additive on $\mathcal{M}$. In fact, we claim that for any $S \in \mathcal{P}(X)$ and pairwise disjoint $A_1, \ldots, A_n \in \mathcal{M}$, we have

$$\mu^*\left(S \cap \left(\bigcup_{m=1}^{n} A_m\right)\right) = \sum_{m=1}^{n} \mu^*\left(S \cap A_m\right). \tag{70}$$

We prove (70) by induction on n. The equality holds trivially for $n = 1$. For the induction step, assume that it holds for some $n \geq 1$. Let S be a subset of X and let$A_1, \ldots, A_{n+1}$ be a finite sequence of members in $\mathcal{M}$. Then

$$\begin{aligned}
\mu^*\left(S \cap \left(\bigcup_{m=1}^{n+1} A_m\right)\right) &\geq \mu^*\left(S \cap \left(\bigcup_{m=1}^{n+1} A_m\right) \cap A_{n+1}\right) + \mu^*\left(S \cap \left(\bigcup_{m=1}^{n+1} A_m\right) \cap (X \backslash A_{n+1})\right) \\
&= \mu^*\left(S \cap A_{n+1}\right) + \mu^*\left(S \cap \left(\bigcup_{m=1}^{n} A_m\right)\right) \\
&= \mu^*\left(S \cap A_{n+1}\right) + \sum_{m=1}^{n} \mu^*(S \cap A_m) \\
&= \sum_{m=1}^{n+1} \mu^*(S \cap A_m).
\end{aligned}$$

This establishes (70). Setting $S = X$ in (70) gives us finite additivity of μ^* on $\mathcal{M}$.

Step 3: We prove that $\mathcal{M}$ is a σ-algebra. Since $\mathcal{M}$ was already shown to be an algebra, it suffices to show that $\mathcal{M}$ is closed under countable unions. Let (A_n) be a sequence in $\mathcal{M}$. Disjointify the sequence (A_n) to the sequence

(D_n): set $D_1 = A_1$ and $D_n = A_n \backslash \bigcup_{m=1}^{n-1} A_m$ for all $n > 1$. Note that (D_n) is a sequence in $\mathcal{M}$ since $\mathcal{M}$ is algebra. Let $S \in \mathcal{P}(X)$ and $n \in \mathbb{N}$. Observe that

$$\begin{aligned}
\mu^*(S) &\geq \mu^*\left(S \cap \left(\bigcup_{m=1}^{n} D_m\right)\right) + \mu^*\left(S \backslash \left(\bigcup_{m=1}^{n} D_m\right)\right) \\
&\geq \mu^*\left(S \cap \left(\bigcup_{m=1}^{n} D_m\right)\right) + \mu^*\left(S \backslash \left(\bigcup_{n\in\mathbb{N}} A_n\right)\right) \\
&= \sum_{m=1}^{n} \mu^*\left(S \cap D_m\right) + \mu^*\left(S \backslash \left(\bigcup_{n\in\mathbb{N}} A_n\right)\right),
\end{aligned}$$

where we applied finite-additivity of μ^* to the first term on the right-hand side and we applied monotonicity of μ^* to the second term on the right-hand side. Taking the limit as $n \to \infty$. We obtain

$$\begin{aligned}
\mu^*(S) &\geq \sum_{m=1}^{\infty} \mu^*\left(S \cap D_m\right) + \mu^*\left(S \backslash \left(\bigcup_{n\in\mathbb{N}} A_n\right)\right) \\
&\geq \mu^*\left(\bigcup_{n\in\mathbb{N}} \left(S \cap D_m\right)\right) + \mu^*\left(S \backslash \left(\bigcup_{n\in\mathbb{N}} A_n\right)\right) \\
&= \mu^*\left(S \cap \bigcup_{n\in\mathbb{N}} D_m\right) + \mu^*\left(S \backslash \left(\bigcup_{n\in\mathbb{N}} A_n\right)\right) \\
&= \mu^*\left(S \cap \left(\bigcup_{n\in\mathbb{N}} A_m\right)\right) + \mu^*\left(S \backslash \left(\bigcup_{n\in\mathbb{N}} A_n\right)\right),
\end{aligned}$$

where we applied countable subadditivity of μ^* to the first expression on the right-hand side. Thus $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{M}$. $\square$

Remark 49. One should compare the proof that $(X, \sigma(\mathcal{A}), \mu^*|_{\sigma(\mathcal{A})})$ is a measure space with the proof that $(X, \mathcal{M}, \mu^*|_{\mathcal{M}})$ is a measure space. They both have three similar steps. However in the former case, we needed to use continuity of μ^* to prove these steps, whereas in the latter case, we needed to use μ^*-measurability to prove these steps. Actually, we did prove a slightly stronger result in the former case, namely that $\overline{\mathcal{A}} = \sigma(\mathcal{A})$.

In general, $\mathcal{M}_{\mu^*}$ strictly contains $\sigma(\mathcal{A})$. The σ-algebra $\mathcal{M}$ has the following additional property: if E is a member of $\mathcal{M}$ and $\mu^*(E) = 0$, then every subset of E is also a member of $\mathcal{M}$. Indeed, if $S \subseteq E$, then for any $A \in \mathcal{A}$, we have

$$\begin{aligned}
\mu^*(S \cap A) + \mu^*(S \backslash A) &\leq \mu^*(E) + \mu^*(S \backslash A) \\
&= \mu^*(S \backslash A) \\
&\leq \mu^*(S),
\end{aligned}$$

which implies $S \in \mathcal{M}$. This property is not necessarily true for $\sigma(\mathcal{A})$. In the special case $\mathcal{A} = \mathcal{I}$ interval algebra, we call this σ-algebra $\mathcal{M}$ the σ-algebra of **Lebesgue measurable sets**. Recall that $\sigma(\mathcal{I})$ is called the σ-algebra of **Borel measurable sets**. It is known that not every Lebesgue measurable set is Borel measurable. The extension of the length measure m from $\mathcal{I}$ to $\sigma(\mathcal{I})$ or $\mathcal{M}$ is called *the* **Lebesgue measure**. By *a* **Borel measure**, we mean any measure defined on $\sigma(\mathcal{I})$. There are many Borel measures but there is only one Lebesgue measure.

30.5 The Borel Algebra

For an algebra $\mathcal{A}$, we define $\mathcal{A}_\sigma$ to be the collection of all countable unions of elements in $\mathcal{A}$. We define $\mathcal{A}_\delta$ to be the collection of all countable intersections of elements in $\mathcal{A}$. We define $\mathcal{A}_{\delta\sigma} = (\mathcal{A}_\delta)_\sigma$ and in the same way, we define $\mathcal{A}_{\sigma\delta} = (\mathcal{A}_\sigma)_\delta$. In general,

$$\mathcal{A}_{\sigma\delta\sigma\delta\cdots} \quad \text{and} \quad \mathcal{A}_{\delta\sigma\delta\sigma\cdots}$$

are not equal to $\sigma(\mathcal{A})$.

Proposition 30.11. *Let $E \in \sigma(\mathcal{A})$. Then*

1. *for all $\varepsilon > 0$, there exists $G \in \mathcal{A}_\delta$ and $F \in A_\supset$ such that*

$$G \subseteq E \subseteq F$$

   and $\mu(F \backslash E) < \varepsilon$ and $\mu(E \backslash G) < \varepsilon$.

2. *there exists $B \in \mathcal{A}_{\supset\circ}$ such that $E \subseteq B$ and $\mu(E) = \mu(B)$.*

Proof. 1. Let $\varepsilon > 0$. Choose a cover $\{A_n\} \subseteq \mathcal{A}$ of E such that

$$\sum_{n=1}^{\infty} \mu(A_n) < \mu(E) + \varepsilon.$$

where we denote by $\mu\colon \sigma(\mathcal{A}) \to [0, \infty]$ to be the unique extension of $\mu\colon \mathcal{A} \to [0, \infty]$. Now set

$$F = \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}_\sigma.$$

Then $E \subseteq F$ and

$$\begin{aligned}
\mu(F \backslash E) &= \mu(F) - \mu(E) \\
&= \mu\left(\bigcup_{n=1}^{\infty} A_n\right) - \mu(E) \\
&\leq \sum_{n=1}^{\infty} \mu(A_n) - \mu(E) \\
&< \varepsilon.
\end{aligned}$$

Now we apply what we just proved to the set $E^c \in \sigma(\mathcal{A})$. This means we can find a set $F \in \mathcal{A}_\sigma$ such that $E^c \subseteq F$ and

$$\mu(F \backslash E^c) < \varepsilon.$$

Now we set $G = F^c$. Then observe that $G \in \mathcal{A}_\delta$ and

$$\begin{aligned}
\mu(E \backslash G) &= \mu(E \cap G^c) \\
&= \mu(E \cap (F^c)^c) \\
&= \mu(F \cap (E^c)^c) \\
&= \mu(F \backslash E^c) \\
&< \varepsilon.
\end{aligned}$$

2. For each $n \in \mathbb{N}$, choose $F_n \in \mathcal{A}_\sigma$ such that $E \subseteq F_n$ and

$$\mu(F_n \backslash E) < \frac{1}{n}.$$

Let $B := \bigcap_{n=1}^{\infty} F_n$. Then $B \in \mathcal{A}_{\sigma\delta}$ and $E \subseteq B$ since $E \subseteq F_n$ for each $n \in \mathbb{N}$. Finally, observe that

$$\begin{aligned}
\mu(B \backslash E) &\leq \mu(F_n \backslash E) \\
&< \frac{1}{n}
\end{aligned}$$

for all $n \in \mathbb{N}$. This implies $\mu(B \backslash E) = 0$. $\square$

30.5.1 Borel σ-Algebra on $\mathbb{R}$

Definition 30.3. The **Borel σ-algebra** on $\mathbb{R}$ is defined to be

$$\mathcal{B}(\mathbb{R}) = \{E \subseteq \mathbb{R} \mid E \cap [n, n+1) \text{ is a Borel set in } [n, n+1) \text{ for all } n \in \mathbb{Z}\}.$$

For $E \in \mathcal{B}(\mathbb{R})$, we define

$$\mathrm{m}(E) = \sum_{n \in \mathbb{Z}} \mathrm{m}(E \cap [n, n+1)).$$

Denote by $\mathcal{I}_n$ the interval algebra on $[n, n+1)$. Denote by $\mathcal{B}_n = \sigma(\mathcal{I}_n)$ the Borel σ-algebra on $[n, n+1)$. It is easy to show that

$$\mathcal{B}(\mathbb{R}) = \sigma\left(\bigcup_{n=-\infty}^{\infty} \mathcal{B}_n\right).$$

Indeed, we have

$$\mathcal{B}(\mathbb{R}) \supseteq \sigma\left(\bigcup_{n=-\infty}^{\infty} \mathcal{B}_n\right),$$

since $\bigcup_n \mathcal{B}_n \subseteq \mathcal{B}(\mathbb{R})$ for all $n \in \mathbb{Z}$ and $\sigma\left(\bigcup_n \mathcal{B}_n\right)$ is the smallest σ-algebra which contains all the $\bigcup_n \mathcal{B}_n$. For the reverse inclusion, let $E \in \mathcal{B}(\mathbb{R})$. Then $E \cap [n, n+1) \in \mathcal{B}_n$ for all $n \in \mathbb{Z}$. Therefore

$$\begin{aligned} E &= \bigcup_{n=-\infty}^{\infty} (E \cap [n, n+1)) \\ &\in \bigcup_{n=-\infty}^{\infty} \mathcal{B}_n. \end{aligned}$$

Thus

$$\mathcal{B}(\mathbb{R}) \subseteq \sigma\left(\bigcup_{n=-\infty}^{\infty} \mathcal{B}_n\right).$$

30.5.2 Translation Invariance

Proposition 30.12. *For all $E \in \mathcal{B}(\mathbb{R})$ and $t \in \mathbb{R}$, we have*

$$\mathrm{m}(E) = \mathrm{m}(E + t).$$

Proof. Define $\mu \colon \mathcal{B}(\mathbb{R}) \to [0, \infty]$ by

$$\mu(E) = \mathrm{m}(E + t)$$

for all $E \in \mathcal{B}(\mathbb{R})$. We need to show that $\mathrm{m}(E) = \mu(E)$ for all $E \in \mathcal{B}(\mathbb{R})$. Let $[c, d] \subseteq [n, n+1)$ (where $[c, d]$ can be open or half-open). Then $\mathrm{m}([c, d]) = d - c$. Note

$$\begin{aligned} \mu([c, d]) &= \mathrm{m}([c + t, d + t]) \\ &= \sum_{k=-\infty}^{\infty} \mathrm{m}([c + t, d + t] \cap [n, n+1)). \end{aligned}$$

Suppose that $[c + t, d + t] \subseteq [k, k+1)$ for some $k \in \mathbb{Z}$. Then

$$\mathrm{m}([c + t, d + t] \cap [n, n+1)) = \begin{cases} d - c & \text{if } n = k \\ 0 & \text{if } n \neq k. \end{cases}$$

So $\mu([c, d]) = d - c$. Now suppose that $[c + t, d + t] \subseteq [k, k+1) \cup [k+1, k+2)$ for some $k \in \mathbb{N}$. Then

$$\begin{aligned} \mathrm{m}([c + t, d + t]) &= \mathrm{m}([c + t, d + t] \cap [k, k+1]) + \mathrm{m}([c + t, d + t] \cap [k+1, k+2]) \\ &= \mathrm{m}([c + t, k+1]) + \mathrm{m}([k+1, d + t]) \\ &= k + 1 - (c + t) + d + t - (k+1) \\ &= d - c \\ &= \mathrm{m}([c, d]). \end{aligned}$$

So for each interval $[c, d] \subseteq [n, n+1)$, we have

$$\mu([c, d]) = \mathrm{m}([c, d]).$$

By finite additivity of m and μ, this implies $\mu(E) = \mathrm{m}(E)$ for all $E \in \mathcal{I}_n$. Since $\mu|_{\mathcal{B}_n}$ and $\mathrm{m}|_{\mathcal{B}_n}$ are both finite measures that coincide on $\mathcal{I}_n$, they must coincide on $\sigma(\mathcal{I}_n) = \mathcal{B}_n$ by the uniqueness part of the extension theorem. Now for every $E \in \mathcal{B}(\mathbb{R})$, we have

$$\begin{aligned}
\mathrm{m}(E) &= \sum_{n \in \mathbb{Z}} \mathrm{m}(E \cap [n, n+1)) \\
&= \sum_{n \in \mathbb{Z}} \mu(E \cap [n, n+1)) \\
&= \mu(E).
\end{aligned}$$

□

30.5.3 Existence of non-Borel sets

Lemma 30.7. *Let $E \subseteq [0, 1)$ be a Borel set. For $t \in [0, 1)$, define the **cyclic translation***

$$E \oplus t = \{x \oplus t \mid x \in E\}$$

where

$$x \oplus t = \begin{cases} x + t & \text{if } x + t < 1 \\ x + t - 1 & \text{if } x + t \geq 1 \end{cases}$$

The set $E \oplus t$ is a Borel set for all $t \in [0, 1)$, and

$$\mathrm{m}(E \oplus t) = \mathrm{m}(E).$$

Proof. Let $E_1 = E \cap [1 - t, 1)$ and $E_2 = E \cap [0, 1 - t)$. Clearly $E = E_1 \cup E_2$ and E_1 and E_2 are disjoint. Since $E_2 + t \subseteq [0, 1)$, we have

$$\begin{aligned}
\mathrm{m}(E_2 + t) &= \mathrm{m}(E_2 \oplus t) \\
&= \mathrm{m}(E_2)
\end{aligned}$$

Also $E_1 + t = [1, 1 + t)$ implies $E_1 \oplus t = E_1 + t - 1$, which implies

$$\begin{aligned}
\mathrm{m}(E_1 \oplus t) &= \mathrm{m}(E_1 + (t - 1)) \\
&= \mathrm{m}(E_1).
\end{aligned}$$

Thus since $(E_1 \oplus t) \cap (E_2 \oplus t) = \emptyset$, we have

$$\begin{aligned}
\mathrm{m}(E \oplus t) &= \mathrm{m}(E_1 \oplus t) + \mathrm{m}(E_2 \oplus t) \\
&= \mathrm{m}(E_1) + \mathrm{m}(E_2) \\
&= \mathrm{m}(E_1 \cup E_2) \\
&= \mathrm{m}(E).
\end{aligned}$$

□

Proposition 30.13. *There exists a set $N \subseteq [0, 1)$ which is not a Borel set.*

Proof. Define a relation $\sim$ on $[0, 1)$ by $x \sim y$ if $x - y \in \mathbb{Q}$. It's easy to see that $\sim$ is an equivalence relation. Define a set N by picking one element in each equivalence class of this relation. Denote by (q_n) the sequence of all rational numbers in $[0, 1)$. Notice that

$$(N \oplus q_n) \cap (N \oplus q_m) = \emptyset$$

for all $q_n \neq q_m$. Furthermore, we have

$$[0, 1) = \bigcup_{n=1}^{\infty} (N \oplus q_n).$$

Indeed, one inclusion is clear, so let's show the other inclusion. Assume for a contradiction that $[0, 1)$ is not contained in $\bigcup_n N \oplus q_n$. Choose $x \in [0, 1)$ such that $x \notin \bigcup_n (N \oplus q_n)$. Let $y \in N$ such that $y \sim x$. □

31 Integration

31.1 Simple Functions

Definition 31.1. Let $(X, \mathcal{M})$ be a measurable space. A function $\varphi\colon X \to \mathbb{R}$ is said to be a **simple function** (with respect to the measurable space $(X, \mathcal{M})$) if it is an $\mathbb{R}$-linear combination of indicators of measurable sets, or in other words, it has the form

$$\varphi = \sum_{i=1}^{n} a_i 1_{A_i} \tag{71}$$

where $a_i \in \mathbb{R}$ and $A_i \in \mathcal{M}$ for all $1 \leq i \leq n$. We call the expression (71) a **representation** of φ. If $i \neq i'$ implies $A_i \cap A_{i'} = \emptyset$, then we call (71) a **disjoint representation** of φ. If $i \neq i'$ implies $a_i \neq a_{i'}$ and $A_i \cap A_{i'} = \emptyset$, then we call (71) *the* **canonical representation** of φ. It is clear that this representation uniquely determines φ (which is why we say it is the *canonical* representation).

31.1.1 Integrating nonnegative simple functions

Definition 31.2. Let $(X, \mathcal{M}, \mu)$ be a measure space and let $\varphi\colon X \to [0, \infty)$ be a nonnegative simple function with canonical representation

$$\varphi = \sum_{i=1}^{n} a_i 1_{A_i}.$$

We define the **integral of** φ by

$$\int_X \varphi \mathrm{d}\mu := \sum_{i=1}^{n} a_i \mu(A_i). \tag{72}$$

Note that there are no issues with well-definedness of (72) since the canonical representation is uniquely determined by φ. It will actually turn out that it doesn't matter how we represent φ; if $\varphi = \sum_{i=1}^{m} c_i 1_{E_i}$ is any representation of φ (canonical or not) then we will have $\int_X \varphi \mathrm{d}\mu = \sum_{i=1}^{m} c_i \mu(E_i)$. Let us show this for disjoint representations of φ.

Lemma 31.1. *Let φ be a simple function and suppose*

$$\varphi = \sum_{i=1}^{m} c_i 1_{E_i} \tag{73}$$

is a disjoint representation of φ. Then

$$\int_X \varphi \mathrm{d}\mu = \sum_{i=1}^{m} c_i \mu(E_i).$$

Proof. Note that the representation (73) is not necessarily canonical since we may have $c_i = c_{i'}$ for some $i \neq i'$. So express φ in terms of its canonical representation, say $\varphi = \sum_{j=1}^{n} a_j 1_{A_j}$. Then we have

$$\begin{aligned}
\int_X \varphi \mathrm{d}\mu(x) &= \sum_{j=1}^{n} a_j \mu(A_j) \\
&= \sum_{j=1}^{n} a_j \mu\left(\bigcup_{i \mid c_i = a_j} E_i \right) \\
&= \sum_{j=1}^{n} a_j \sum_{i \mid c_i = a_j} \mu(E_i) \\
&= \sum_{j=1}^{n} \sum_{i \mid c_i = a_j} c_i \mu(E_i) \\
&= \sum_{i=1}^{m} c_i \mu(E_i).
\end{aligned}$$

□

31.1.2 $\mathbb{R}_{\geq 0}$-Scaling, Monotonicity, and Additivity of Integration for Nonnegative Simple Functions

Proposition 31.1. *Let $\varphi, \psi \colon X \to [0, \infty)$ be nonnegative simple functions and let $a \geq 0$. Then we have*

1. *$\mathbb{R}_{\geq 0}$-scaling of integration for nonnegative simple functions.:*

$$\int_X a\varphi \mathrm{d}\mu = a \int_X \varphi \mathrm{d}\mu. \tag{74}$$

2. *Monotonicity of integration for nonnegative simple functions: if $\varphi \leq \psi$, then*

$$\int_X \varphi \mathrm{d}\mu \leq \int_X \psi \mathrm{d}\mu.$$

3. *Additivity of integration for nonnegative simple functions:*

$$\int_X (\varphi + \psi) \mathrm{d}\mu = \int_X \varphi \mathrm{d}\mu + \int_X \psi \mathrm{d}\mu.$$

Proof. 1. Express φ in terms of its canonical representation, say $\varphi = \sum_{i=1}^n a_i 1_{A_i}$. Then $a\varphi$ is a nonnegative simple function with canonical representation $a\varphi = \sum_{i=1}^n aa_i 1_{A_i}$. Therefore

$$\begin{aligned} a \int_X \varphi \mathrm{d}\mu &= a \sum_{i=1}^n a_i \mu(A_i) \\ &= \sum_{i=1}^n aa_i \mu(A_i) \\ &= \int_X a\varphi \mathrm{d}\mu \end{aligned}$$

Thus we have $\mathbb{R}_{\geq 0}$-scaling of integration for nonnegative simple functions.

2. Assume $\varphi \leq \psi$. Express φ and ψ in terms of their canonical representations, say $\varphi = \sum_{i=1}^m a_i 1_{A_i}$ and $\psi = \sum_{j=1}^n b_j 1_{B_j}$. Since $\varphi \leq \psi$, we have

$$\begin{aligned} \varphi &= \sum_{i=1}^m a_i 1_{A_i} \\ &= \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} \min(a_i, b_j) 1_{A_i \cap B_j} \\ &\leq \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} \max(a_i, b_j) 1_{A_i \cap B_j} \\ &= \sum_{j=1}^n b_j 1_{B_j} \\ &= \psi. \end{aligned}$$

Therefore

$$\begin{aligned} \int_X \varphi \mathrm{d}\mu &= \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} \min(a_i, b_j) \mu(A_i \cap B_j) \\ &\leq \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} \max(a_i, b_j) \mu(A_i \cap B_j) \\ &= \int_X \psi \mathrm{d}\mu. \end{aligned}$$

Thus we have monotonicity of integration for nonnegative simple functions.

3. Express φ and ψ in terms of their canonical representations, say $\varphi = \sum_{i=1}^m a_i 1_{A_i}$ and $\psi = \sum_{j=1}^n b_j 1_{B_j}$. Then $\varphi + \psi$ is a nonnegative simple function with a disjoint representation given by

$$\varphi + \psi = \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} (a_i + b_j) 1_{A_i \cap B_j}$$

Therefore

$$\begin{aligned}
\int_X (\varphi + \psi) \mathrm{d}\mu &= \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} (a_i + b_j) \mu(A_i \cap B_j) \\
&= \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} a_i \mu(A_i \cap B_j) + \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} b_j \mu(A_i \cap B_j) \\
&= \sum_{1 \le i \le m} a_i \mu\left(\bigcup_{j=1}^n (A_i \cap B_j) \right) + \sum_{1 \le j \le n} b_j \mu\left(\bigcup_{i=1}^m A_i \cap B_j \right) \\
&= \sum_{1 \le i \le m} a_i \mu(A_i) + \sum_{1 \le j \le n} b_j \mu(B_j) \\
&= \int_X \varphi \mathrm{d}\mu + \int_X \psi \mathrm{d}\mu.
\end{aligned}$$

Thus we have additivity of integration for nonnegative simple functions. □

31.2 Measurable Functions

We want to be able to integrate more general nonnegative functions. The functions that we want to consider now are called measurable functions. Before we explain what these functions are how to integrate them, let's define them in a more general context first.

Definition 31.3. Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces and let $f\colon X \to Y$ be a function. We say f is **measurable with respect to** $\mathcal{M}$ **and** $\mathcal{N}$ if $f^{-1}(\mathcal{N}) \subseteq \mathcal{M}$ where

$$f^{-1}(\mathcal{N}) = \{f^{-1}(B) \mid B \in \mathcal{N}\}.$$

In other words, f is measurable with respect to $\mathcal{M}$ and $\mathcal{N}$ if for all $B \in \mathcal{N}$ we have

$$f^{-1}(B) = \{x \in X \mid f(x) \in B\} \in \mathcal{M}.$$

Thus f is measurable with respect to $\mathcal{M}$ and $\mathcal{N}$ if the inverse image of a measurable set is a measurable set. If $\mathcal{M} = \mathcal{N}$, then we will just say f is measurable with respect to $\mathcal{M}$. If the σ-algebras $\mathcal{M}$ and $\mathcal{N}$ are clear from context, then we will just say f is measurable.

For our purposes, we will mostly be interested in the case where $Y = \mathbb{R}$ and $\mathcal{N} = \mathcal{B}(\mathbb{R})$ (or $Y = [0, \infty]$ and $\mathcal{N} = \mathcal{B}[0, \infty]$). We want to find another criterion for a function $f\colon X \to \mathbb{R}$ to be measurable (rather than the inverse image of a Borel-measurable set is a $\mathcal{M}$-measurable set). We state this criterion as a corollary following the next two propositions:

Proposition 31.2. *Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces and let $f\colon X \to Y$ be a function. Suppose that $\mathcal{N}$ is generated as a σ-algebra by the collection $\mathcal{C}$ of subsets of Y. Then $f^{-1}(\mathcal{N}) \subseteq \mathcal{M}$ if and only if $f^{-1}(\mathcal{C}) \subseteq \mathcal{M}$.*

Proof. One direction is clear, so we just prove the other direction. Suppose $f^{-1}(\mathcal{C}) \subseteq \mathcal{M}$. Observe that

$$\{B \in \mathcal{P}(Y) \mid f^{-1}(B) \in \mathcal{M}\}$$

is a σ-algebra which contains $\mathcal{C}$. Indeed, it is a σ-algebra since f^{-1} maps the emptyset set to the emptyset and maps the whole space Y to the whole space X, and since f^{-1} commutes with unions and complements. Furthemore, this σ-algebra contains $\mathcal{C}$ since $f^{-1}(\mathcal{C}) \subseteq \mathcal{M}$. Since $\mathcal{N}$ is the *smallest* σ-algebra which contains $\mathcal{C}$, it follows that

$$\mathcal{N} \subseteq \{B \in \mathcal{P}(Y) \mid f^{-1}(B) \in \mathcal{M}\}.$$

In particular, if $B \in \mathcal{N}$, then $f^{-1}(B) \in \mathcal{M}$. Thus f is measurable. □

Proposition 31.3. *Let $\mathcal{C} = \{(-\infty, c) \mid c \in \mathbb{R}\}$. Then $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$.*

Proof. Let $\mathcal{I}_n$ be the collection of all subintervals of $[n, n+1)$ and let $\mathcal{B}_n = \sigma(\mathcal{I}_n)$. So

$$\mathcal{B}(\mathbb{R}) = \{E \subseteq \mathbb{R} \mid E \cap [n, n+1) \in \mathcal{B}_n \text{ for all } n \in \mathbb{Z}\}.$$

Let $c \in \mathbb{R}$. Then since $(-\infty, c) \cap [n, n+1]$ is a subinterval of $[n, n+1]$ for all $n \in \mathbb{Z}$, it follows that $(-\infty, c) \in \mathcal{B}$ for all $n \in \mathbb{Z}$. Thus $\mathcal{C} \subseteq \mathcal{B}$ which implies $\sigma(\mathcal{C}) \subseteq \mathcal{B}$ (as $\sigma(\mathcal{C})$ is the *smallest* σ-algebra which contains $\mathcal{C}$). Conversely, note that $\sigma(\mathcal{C})$ contains all subintervals of $[n, n+1)$ for all $n \in \mathbb{Z}$. Thus $\sigma(\mathcal{C}) \supseteq \mathcal{B}_n$ for all $n \in \mathbb{Z}$ (as $\mathcal{B}_n$ is the *smallest* σ-algebra which contains all subintervals of $[n, n+1)$. Since $\mathcal{B}(\mathbb{R}) = \sigma\left(\bigcup_{n \in \mathbb{Z}} \mathcal{B}_n\right)$, it follows that $\mathcal{B}(\mathbb{R}) \subseteq \sigma(\mathcal{C})$. $\square$

Corollary 12. *Let $(X, \mathcal{M})$ be a measurable space and let $f \colon X \to \mathbb{R}$ be a function. Then f is measurable with respect to $\mathcal{M}$ and $\mathcal{B}(\mathbb{R})$ if and only if $\{f < c\} \in \mathcal{M}$ for all $c \in \mathbb{R}$.*

Proof. Follows from Proposition (38.3) and Proposition (38.2). $\square$

Remark 50. Here $\{f < c\}$ denotes the set $\{x \in X \mid f(x) < c\}$. In fact, we shall often use this shorthand notation for other sets like this. For instance, we write

$$\begin{aligned}
\{f < c\} &= \{x \in X \mid f(x) < c\} \\
\{f \leq c\} &= \{x \in X \mid f(x) \leq c\} \\
\{f > c\} &= \{x \in X \mid f(x) > c\} \\
\{f \geq c\} &= \{x \in X \mid f(x) \geq c\} \\
\{f = c\} &= \{x \in X \mid f(x) = c\} \\
\{f \neq c\} &= \{x \in X \mid f(x) \neq c\}.
\end{aligned}$$

This is notation is very common throughout the literature.

31.2.1 Combining Measurable Functions to get More Measurable Functions

Proposition 31.4. *Let $f, g \colon X \to \mathbb{R}$ be measurable functions and let $a \in \mathbb{R}$. Then af, $f + g$, f^2, $|f|$, fg, $\max\{f, g\}$, and $\min\{f, g\}$ are all measurable.*

Proof. We first show af is measurable. If $a = 0$, then af is the zero function, which is measurable, so assume $a \neq 0$. Then for any $c \in \mathbb{R}$ we have

$$\{af < c\} = \begin{cases} \{f < c/a\} \in \mathcal{M} & \text{if } a > 0 \\ \{f > c/a\} \in \mathcal{M} & \text{if } a < 0 \end{cases}$$

It follows that af is measurable.

Next we show that $f + g$ is measurable. Observe that for any $c \in \mathbb{R}$, we have

$$\begin{aligned}
x \in \{f + g < c\} &\iff f(x) + g(x) < c \\
&\iff \text{there exists an } r \in \mathbb{Q} \text{ such that } f(x) < r \text{ and } r < c - g(x) \\
&\iff \text{there exists an } r \in \mathbb{Q} \text{ such that } x \in \{f < r\} \cap \{g < c - r\}. \\
&\iff x \in \bigcup_{r \in \mathbb{Q}} \{f < r\} \cap \{g < c - r\}.
\end{aligned}$$

Therefore

$$\{f + g < c\} = \bigcup_{r \in \mathbb{Q}} \{f < r\} \cap \{g < c - r\} \in \mathcal{M}.$$

It follows that $f + g$ is measurable.

Next we show that f^2 is measurable. For any $c \in \mathbb{R}$, we have

$$\{f^2 > c\} = \begin{cases} \{f > \sqrt{c}\} \cup \{f < -\sqrt{c}\} \in \mathcal{M} & c \geq 0 \\ X \in \mathcal{M} & c < 0 \end{cases}$$

It follows that f^2 is measurable.

Next we show that $|f|$ is measurable. For any $c \in \mathbb{R}$, we have

$$\{|f| > c\} = \begin{cases} \{f > c\} \cup \{f < -c\} \in \mathcal{M} & c \geq 0 \\ X \in \mathcal{M} & c < 0 \end{cases}$$

It follows that $|f|$ is measurable.

Finally, note that the remaining functions can be expressed as combinations of the previous ones. Indeed,

$$\begin{aligned} fg &= \frac{1}{4}\left((f+g)^2 - (f-g)^2\right) \\ \max\{f,g\} &= \frac{1}{2}(|f+g| + |f-g|) \\ \min\{f,g\} &= \frac{1}{2}(|f+g| - |f-g|) \end{aligned}$$

It follows that they are all measurable. □

Proposition 31.5. *Let $(f_n \colon X \to \mathbb{R})$ be a sequence of measurable functions. Then $\sup f_n$, $\inf f_n$, $\limsup f_n$, and $\liminf f_n$ are all measurable. In particular, if*

$$\lim_{n\to\infty} f_n(x)$$

exists for all $x \in X$. The corresponding function is also measurable.

Proof. Let $c \in \mathbb{R}$. Then we have

$$\{\sup f_n > c\} = \bigcup_{n \geq 1} \{f_n > c\}.$$

This implies $\sup f_n$ is measurable. Similarly, we have

$$\{\inf f_n < c\} = \bigcup_{n \geq 1} \{f_n < c\}.$$

This implies $\inf f_n$ is measurable. Finally since

$$\limsup f_n = \inf_{N \geq 1} \sup_{n \geq N} f_n \qquad \text{and} \qquad \liminf f_n = \sup_{N \geq 1} \inf_{n \geq N} f_n,$$

we see that both $\limsup f_n$ and $\liminf f_n$ are measurable. □

31.2.2 Criterion for Nonnegative Function to be Measurable

If $f \colon X \to \mathbb{R}$ is a nonnegative function, then we have another criterion for f to be measurable.

Proposition 31.6. *Let $f \colon X \to [0, \infty]$ be a nonnegative function (which possibly takes infinite values). The following are equivalent:*

1. *f is measurable.*
2. *There exists an increasing sequence of nonnegative simple functions $(\varphi_n \colon X \to [0, \infty])$ such that (φ_n) converges pointwise to f.*

Proof. Proposition (31.5) gives us 2 implies 1 (since simple functions are measurable!), so we just need to show 1 implies 2. Suppose f is measurable. For each $n \in \mathbb{N}$ and for each $1 \leq i \leq 2^n$, we define

$$E_n = \{f \geq n\} \text{ and } E_{n,i} = \left\{\frac{i-1}{2^n} \leq f < \frac{i}{2^n}\right\}.$$

Note that $E_n, E_{n,i} \in \mathcal{M}$. For each $n \in \mathbb{N}$ we define $\varphi_n \colon X \to [0, \infty]$ by

$$\varphi_n = n 1_{E_n} + \sum_{i=1}^{2^n} \left(\frac{i-1}{2^n}\right) 1_{E_{n,i}}$$

Therefore each φ_n is a simple function. It's easy to check that (φ_n) is an increasing sequence of nonnegative functions. Let us check that (φ_n) converges pointwise to f. Let $x \in X$. If $f(x) = \infty$, then $\varphi_n(x) = n$ for all $n \in \mathbb{N}$, and thus

$$\begin{aligned}\varphi_n(x) &= n \\ &\to \infty \\ &= f(x)\end{aligned}$$

as $n \to \infty$, so assume $f(x) < \infty$. Let $\varepsilon > 0$ and choose $N \in \mathbb{N}$ such that $2^{-N} < \varepsilon$. Then $n \geq N$ implies

$$\begin{aligned}f(x) - \varphi_n(x) &< 2^{-n} \\ &\leq 2^{-N} \\ &< \varepsilon.\end{aligned}$$

This implies $\varphi_n(x) \to f(x)$ as $n \to \infty$. Since x was arbitrary, we see that (φ_n) converges pointwise to f. □

31.3 The Integral of a Nonnegative Measurable Function

We are now ready to describe how to integrate nonnegative measurable functions.

Definition 31.4. Let $f\colon X \to [0, \infty]$ be a nonnegative measurable function. The **integral of** f is defined to be

$$\int_X f \mathrm{d}\mu := \sup \left\{ \int_X \varphi \mathrm{d}\mu \mid \varphi \text{ is a nonnegative simple function and } \varphi \leq f \right\}.$$

Recall that integrating nonnegative simple functions satisfies three very nice properties; namely $\mathbb{R}_{\geq 0}$-scaling, monotonicity, and additivity. In fact, these three properties continue to hold when integrating nonnegative measurable functions. The first two are easy to show, but additivity requires a little more effort (it will follow from the so-called Monotone Convergence Theorem).

31.3.1 Monotone Convergence Theorem

Theorem 31.2. *(MCT) Let $(f_n\colon X \to [0, \infty])$ be an increasing sequence of nonnegative measurable functions which converges pointwise to a nonnegative function $f\colon X \to [0, \infty]$. Then*

$$\lim_{n \to \infty} \int_X f_n \mathrm{d}\mu = \int_X f \mathrm{d}\mu.$$

Proof. The fact that $f\colon X \to [0, \infty]$ is measurable follows from (Proposition (31.5). Therefore $\int_X f \mathrm{d}\mu$ is defined. Since $f_n \leq f$ for all n, we have

$$\lim_{n \to \infty} \int_X f_n \mathrm{d}\mu \leq \int_X f \mathrm{d}\mu.$$

by monotonicity of integration for nonnegative functions. For the reverse inequality, we just need to show that

$$\lim_{n \to \infty} \int_X f_n \mathrm{d}\mu \geq \int_X \varphi \mathrm{d}\mu \tag{75}$$

for any nonnegative simple function φ such that $\varphi \leq f$, so let φ be such a nonnegative simple function and let $c \in (0, 1)$. For each $n \in \mathbb{N}$, define

$$A_n := \{f_n - c\varphi > 0\}.$$

Then A_n is measurable since $f_n - c\varphi$ is measurable for all n. Also, (A_n) is an ascending sequence of sets such that

$$\bigcup_{n=1}^{\infty} A_n = X$$

since f_n converges pointwise to f. Therefore

$$\begin{aligned}\lim_{n\to\infty}\int_X f_n \mathrm{d}\mu &\geq \lim_{n\to\infty}\int_X f_n 1_{A_n}\mathrm{d}\mu \\ &\geq \lim_{n\to\infty}\int_X c\varphi 1_{A_n}\mathrm{d}\mu \\ &= c\lim_{n\to\infty}\int_X \varphi 1_{A_n}\mathrm{d}\mu \\ &= c\int_X \varphi \mathrm{d}\mu\end{aligned}$$

where we obtained the fourth line from the third line from the fact that the function $\nu\colon \mathcal{M}\to[0,\infty]$ defined by

$$\nu(E) = \int_X \varphi 1_E \mathrm{d}\mu$$

for all $E\in\mathcal{M}$ is a measure. In particular

$$\begin{aligned}\lim_{n\to\infty}\int_X \varphi 1_{A_n}\mathrm{d}\mu &= \lim_{n\to\infty}\nu(A_n) \\ &= \nu\left(\bigcup_{n=1}^{\infty} A_n\right) \\ &= \nu(X) \\ &= \int_X \varphi 1_X \mathrm{d}\mu \\ &= \int_X \varphi \mathrm{d}\mu.\end{aligned}$$

Now we take $c\to 1$ to get (75). □

31.3.2 $\mathbb{R}_{\geq 0}$-Scaling, Monotonicity, and Additivity of Integration for Nonnegative Measurable Functions

Proposition 31.7. *Let $f,g\colon X\to[0,\infty]$ be measurable and let $a\geq 0$. Then we have*

1. *$\mathbb{R}_{\geq 0}$-scaling of integration for nonnegative measurable functions.:*

$$\int_X af\mathrm{d}\mu = a\int_X f\mathrm{d}\mu. \tag{76}$$

2. *Monotonicity of integration for nonnegative measurable functions: if $f\leq g$, then*

$$\int_X f\mathrm{d}\mu \leq \int_X g\mathrm{d}\mu.$$

3. *Additivity of integration for nonnegative measurable functions.:*

$$\int_X (f+g)\mathrm{d}\mu = \int_X f\mathrm{d}\mu + \int_X g\mathrm{d}\mu.$$

Proof. 1) If $a=0$, then (76) is obvious, so assume $a>0$. Let $\varepsilon>0$ and choose a nonnegative simple function $\varphi\colon X\to[0,\infty]$ such that $\varphi\leq f$ and

$$\int_X f\mathrm{d}\mu - \varepsilon < \int_X \varphi\mathrm{d}\mu.$$

Then $a\varphi$ is a nonnegative simple function such that $a\varphi\leq af$. Furthermore, we have

$$\begin{aligned}\int_X af\mathrm{d}\mu &\geq \int_X a\varphi\mathrm{d}\mu \\ &= a\int_X \varphi\mathrm{d}\mu \\ &> a\left(\int_X f\mathrm{d}\mu - \varepsilon\right) \\ &= a\int_X f\mathrm{d}\mu - a\varepsilon.\end{aligned}$$

Taking $\varepsilon \to 0$ gives us

$$\int_X af\mathrm{d}\mu \geq a\int_X f\mathrm{d}\mu.$$

This gives us one inequality.

For the reverse inequality, let $\varepsilon > 0$ and choose a nonnegative simple function $\varphi\colon X \to [0,\infty]$ such that $\varphi \leq af$ and

$$\int_X af\mathrm{d}\mu - \varepsilon < \int_X \varphi\mathrm{d}\mu.$$

Then $a^{-1}\varphi$ is a nonnegative simple function such that $a^{-1}\varphi \leq f$. Furthermore, we have

$$\begin{aligned} a\int_X f\mathrm{d}\mu &\geq a\int_X a^{-1}\varphi\mathrm{d}\mu \\ &= \int_X \varphi\mathrm{d}\mu \\ &> \int_X af\mathrm{d}\mu - \varepsilon \end{aligned}$$

Taking $\varepsilon \to 0$ gives us

$$\int_X af\mathrm{d}\mu \leq a\int_X f\mathrm{d}\mu.$$

Thus we have $\mathbb{R}_{\geq 0}$-scaling of integration for nonnegative measurable functions.

2. Assume $f \leq g$. Let $\varphi\colon X \to [0,1]$ be a nonnegative simple function such that $\varphi \leq f$. Then $\varphi \leq g$ since $f \leq g$. This implies

$$\begin{aligned} \int_X f\mathrm{d}\mu &:= \sup\left\{\int_X \varphi\mathrm{d}\mu \mid \varphi \text{ is a nonnegative simple function and } \varphi \leq f\right\} \\ &\leq \sup\left\{\int_X \varphi\mathrm{d}\mu \mid \varphi \text{ is a nonnegative simple function and } \varphi \leq g\right\} \\ &:= \int_X g\mathrm{d}\mu. \end{aligned}$$

Thus we have monotonicity of integration nonnegative measurable functions.

3. Choose an increasing sequence $(\varphi_n\colon X \to [0,\infty])$ of nonnegative simple functions which converges pointwise to f (we can do this since f is a nonnegative measurable function). Similarly, choose an increasing sequence $(\psi_n\colon X \to [0,\infty])$ of nonnegative simple functions which converges pointwise to g (we can do this since g is a nonnegative measurable function). Then $(\varphi_n + \psi_n)$ is an increasing sequence of nonnegative simple functions which converges pointwise to $f+g$. It follows from MCT that

$$\begin{aligned} \int_X (f+g)\mathrm{d}\mu &= \lim_{n\to\infty}\int_X (\varphi_n + \psi_n)\mathrm{d}\mu \\ &= \lim_{n\to\infty}\int_X \varphi_n\mathrm{d}\mu + \lim_{n\to\infty}\int_X \psi_n\mathrm{d}\mu \\ &= \int_X f\mathrm{d}\mu + \int_X g\mathrm{d}\mu. \end{aligned}$$

Thus we have additivity of integration for nonnegative measurable functions. □

Remark 51. Note that in each proof for the three statements in Proposition (31.7), we needed to use the fact that these three statements hold when integrating nonnegative *simple* functions.

31.3.3 Fatou's Lemma

Proposition 31.8. *(Fatou's Lemma) Let $(f_n\colon X \to [0,\infty])$ be a sequence of nonnegative measurable functions. Then*

$$\int_X \liminf f_n\mathrm{d}\mu \leq \liminf \int_X f_n\mathrm{d}\mu.$$

In particular, if (f_n) converges to a function $f\colon X \to [0,\infty]$ pointwise, then

$$\int_X f\mathrm{d}\mu \leq \liminf \int_X f_n\mathrm{d}\mu. \tag{77}$$

Proof. For each $N \in \mathbb{N}$, set $g_N = \inf_{n \geq N} f_n$. Then (g_N) is an increasing sequence of nonnegative measurable functions which converges pointwise to $\liminf f_n$. Indeed, for any $x \in X$, we have

$$\begin{aligned} \lim_{N\to\infty} g_N(x) &= \lim_{N\to\infty} \inf_{n\geq N} f_n(x) \\ &:= \liminf f_n(x) \end{aligned}$$

It follows from the MCT that

$$\begin{aligned} \int_X \liminf f_n \mathrm{d}\mu &= \lim_{N\to\infty} \int_X g_N \mathrm{d}\mu \\ &= \lim_{N\to\infty} \int_X \inf_{n\geq N} f_n \mathrm{d}\mu \\ &\leq \lim_{N\to\infty} \inf_{n\geq N} \int_X f_n \mathrm{d}\mu \\ &:= \liminf \int_X f_n \mathrm{d}\mu, \end{aligned}$$

where we obtained the third line from the second line from monotonicity of integration (for any $m \geq N$ we have $f_m \geq \inf_{n\geq N} f_n$ which implies $\int_X f_m \mathrm{d}\mu \geq \int_X \inf_{n\geq N} f_n \mathrm{d}\mu$ which implies $\inf_{n\geq N} \int_X f_n \mathrm{d}\mu \geq \int_X \inf_{n\geq N} f_n \mathrm{d}\mu$). The identity (77) follows from the fact that if (f_n) converges to f pointwise, then $f = \liminf f_n$. □

31.4 Integrable Functions

Definition 31.5. Let $(X, \mathcal{M}, \mu)$ be a measure space and let $f\colon X \to [-\infty, \infty]$ be a measurable function. The **positive part** of f, denoted f^+, and the **negative part** of f, denoted f^-, are defined by

$$f^+(x) = \max\{f(x), 0\} \quad \text{and} \quad f^-(x) = -\min\{f(x), 0\}$$

for all $x \in X$. Note that both f^+ and f^- are both nonnegative measurable functions. Furthermore, we have

$$f = f^+ - f^- \quad \text{and} \quad |f| = f^+ + f^-.$$

We say f is **integrable** if

$$\int_X f^+ \mathrm{d}\mu < \infty \quad \text{and} \quad \int_X f^- \mathrm{d}\mu < \infty.$$

Since $|f| = f^+ + f^-$, this is equivalent to saying

$$\int_X |f| \mathrm{d}\mu < \infty.$$

In this case, we define the **integral** of f to be

$$\int_X f \mathrm{d}\mu := \int_X f^+ \mathrm{d}\mu - \int_X f^- \mathrm{d}\mu,$$

31.4.1 $\mathbb{R}$-Scaling, Monotonicity, and Additivity of Integration for Integrable Functions

Proposition 31.9. *Let $f, g\colon X \to \mathbb{R}$ be integrable functions and let $a \in \mathbb{R}$. Then we have*

1. *$\mathbb{R}$-scaling of integration for integrable functions:*

$$\int_X af \mathrm{d}\mu = a \int_X f \mathrm{d}\mu. \tag{78}$$

2. *Additivity of integration for integrable functions:*

$$\int_X (f+g) \mathrm{d}\mu = \int_X f \mathrm{d}\mu + \int_X g \mathrm{d}\mu.$$

3. *Monotonicity of integration for integrable functions: if $f \leq g$, then*

$$\int_X f \mathrm{d}\mu \leq \int_X g \mathrm{d}\mu.$$

Proof. 1. If $a \geq 0$, then $(af)^+ = af^+$ and $(af)^- = af^-$. Therefore

$$\begin{aligned}
\int_X af \mathrm{d}\mu &= \int_X af^+ \mathrm{d}\mu - \int_X af^- \mathrm{d}\mu \\
&= a\int_X f^+ \mathrm{d}\mu - a\int_X f^- \mathrm{d}\mu \\
&= a\left(\int_X f^+ \mathrm{d}\mu - \int_X f^- \mathrm{d}\mu\right) \\
&= a\int_X f \mathrm{d}\mu.
\end{aligned}$$

If $a < 0$, then $(af)^+ = -af^-$ and $(af)^- = -af^+$. Therefore

$$\begin{aligned}
\int_X af \mathrm{d}\mu &= \int_X -af^- \mathrm{d}\mu - \int_X -af^+ \mathrm{d}\mu \\
&= -a\int_X f^- \mathrm{d}\mu + a\int_X f^+ \mathrm{d}\mu \\
&= a\left(\int_X f^+ \mathrm{d}\mu - \int_X f^- \mathrm{d}\mu\right) \\
&= a\int_X (f^+ - f^-) \mathrm{d}\mu \\
&= a\int_X f \mathrm{d}\mu.
\end{aligned}$$

Thus we have $\mathbb{R}$-scaling of integration for integrable functions.

2. Observe that

$$\begin{aligned}
f + g &= f^+ - f^- + g^+ - g^- \\
&= f^+ + g^+ - (f^- + g^-).
\end{aligned}$$

It follows from Proposition (**??**) that

$$\begin{aligned}
\int_X (f+g) \mathrm{d}\mu &= \int_X (f^+ + g^+) \mathrm{d}\mu - \int_X (f^- + g^-) \mathrm{d}\mu \\
&= \int_X f^+ \mathrm{d}\mu + \int_X g^+ \mathrm{d}\mu - \int_X f^- \mathrm{d}\mu - \int_X g^- \mathrm{d}\mu \\
&= \int_X f^+ \mathrm{d}\mu - \int_X f^- \mathrm{d}\mu + \int_X g^+ \mathrm{d}\mu - \int_X g^- \mathrm{d}\mu \\
&= \int_X f \mathrm{d}\mu + \int_X g \mathrm{d}\mu.
\end{aligned}$$

Thus we have additivity of integration for integrable functions.

3. Assume $f \leq g$. Observe that $f \leq g$ implies $g - f \geq 0$. It follows from 1 and 2 that.

$$\begin{aligned}
\int_X g \mathrm{d}\mu - \int_X f \mathrm{d}\mu &= \int_X (g - f) \mathrm{d}\mu \\
&\geq 0.
\end{aligned}$$

This implies $\int_X g \mathrm{d}\mu \geq \int_X f \mathrm{d}\mu$. □

Remark 52. Note that in each proof for the three statements in Proposition (31.9), we needed to use the fact that these three statements hold when integrating nonnegative *measurable* functions.

31.4.2 Lebesgue Dominated Convergence Theorem

Theorem 31.3. *(DCT) Let $g\colon X \to [0,\infty]$ be a nonnegative integrable function. Suppose $(f_n\colon X \to \mathbb{R})$ is a sequence of integrable functions such that*

1. *(f_n) converges pointwise to $f\colon X \to \mathbb{R}$.*
2. *$|f_n| \leq g$ pointwise for all $n \in \mathbb{N}$.*

Then

$$\lim_{n\to\infty} \int_X f_n \mathrm{d}\mu = \int_X f \mathrm{d}\mu.$$

Proof. Since $|f_n| \leq g$ for all $n \in \mathbb{N}$, we have (by taking limits) $|f| \leq g$. Thus f is integrable by monotonicity of integration for integrable functions. Observe that $(g - f_n)$ is a sequence of nonnegative measurable functions. Thus by Fatou's Lemma, we have

$$\begin{aligned}
\int_X g \mathrm{d}\mu - \int_X f \mathrm{d}\mu &= \int_X (g - f) \mathrm{d}\mu \\
&\leq \liminf_{n\to\infty} \int_X (g - f_n) \mathrm{d}\mu \\
&= \int_X g \mathrm{d}\mu - \limsup_{n\to\infty} \int f_n \mathrm{d}\mu.
\end{aligned}$$

In other words, we have

$$\limsup_{n\to\infty} \int_X f_n \mathrm{d}\mu \leq \int_X f \mathrm{d}\mu.$$

Now we apply the same argument with functions $g + f_n$ in place of $g - f_n$, and we obtain

$$\liminf_{n\to\infty} \int_X f_n \mathrm{d}\mu \geq \int_X f \mathrm{d}\mu.$$

□

31.4.3 Chebyshev-Markov Inequality

Proposition 31.10. *(C-M) Let $f\colon X \to \mathbb{R}$ be an integrable function and let $c > 0$. Then*

$$\mu\{|f| \geq c\} \leq \frac{1}{c} \int_X |f| \mathrm{d}\mu.$$

Proof. We have

$$\begin{aligned}
\int_X |f| \mathrm{d}\mu &\geq \int_X |f| 1_{\{|f| \geq c\}} \mathrm{d}\mu \\
&\geq \int_X c 1_{\{|f| \geq c\}} \mathrm{d}\mu \\
&= c\mu\{|f| \geq c\}.
\end{aligned}$$

□

Proposition 31.11. *Let $(X, \mathcal{M}, \mu)$ be a measure space and let f be a measurable function. Then $\mu\{f \neq 0\} = 0$ if and only if $\int_X |f| \mathrm{d}\mu = 0$.*

Proof. We first note that that $\{f \neq 0\} = \{|f| \neq 0\}$. Now suppose $\mu\{f \neq 0\} = 0$. Then we have

$$\begin{aligned}
\int_X |f| \mathrm{d}\mu &= \int_{\{f=0\}} |f| \mathrm{d}\mu + \int_{\{f\neq 0\}} |f| \mathrm{d}\mu \\
&= \int_{\{f\neq 0\}} |f| \mathrm{d}\mu \\
&= \sup\{\int_{\{f\neq 0\}} \varphi \mathrm{d}\mu \mid \varphi \leq |f| \text{ is nonnegative simple function}\} \\
&= \sup\{0 \mid \varphi \leq |f| \text{ is nonnegative simple function}\} \\
&= 0.
\end{aligned}$$

Conversely, suppose $\int_X |f|\mathrm{d}\mu = 0$. Then by Chebyshev-Markov's inequality, we have

$$\begin{aligned}\mu\{f \neq 0\} &= \mu\{|f| \neq 0\} \\ &= \mu\left(\bigcup_{n=1}^{\infty}\{|f| \geq 1/n\}\right) \\ &= \lim_{n\to\infty} \mu(\{|f| \geq 1/n\}) \\ &\leq \lim_{n\to\infty} n\int_X |f|\mathrm{d}\mu \\ &= 0.\end{aligned}$$

□

Definition 31.6. A property is said to hold **μ-almost everywhere** (or more simpwith respect to μ (denoted μ a.e.) if it is true everywhere except on a set of measure zero. If the measure μ is understood from context, then we will simply write "almost everywhere" instead of "μ-almost everywhere".

Corollary 13. *Let $(X, \mathcal{M}, \mu)$ be a measure space and let f and g be integrable functions. Then $f = g$ be almost everywhere if and only if $\int_X |f|\mathrm{d}\mu = \int_X |g|\mathrm{d}\mu$.*

Proof. Suppose $f = g$ almost everywhere. Then also $|f| = |g|$ almost everywhere, and so we have

$$\begin{aligned}\int_X |f|\mathrm{d}\mu &= \int_{\{f=g\}} |f|\mathrm{d}\mu + \int_{\{f\neq g\}} |f|\mathrm{d}\mu \\ &= \int_{\{f=g\}} |f|\mathrm{d}\mu \\ &= \int_{\{f=g\}} |g|\mathrm{d}\mu \\ &= \int_{\{f=g\}} |g|\mathrm{d}\mu + \int_{\{f\neq g\}} |g|\mathrm{d}\mu \\ &= \int_X |g|\mathrm{d}\mu.\end{aligned}$$

Conversely, suppose $\int_X |f|\mathrm{d}\mu = \int_X |g|\mathrm{d}\mu$. Then we have

$$\begin{aligned}0 &= \int_X |f|\mathrm{d}\mu - \int_X |g|\mathrm{d}\mu \\ &= \int_X (|f| - |g|)\mathrm{d}\mu.\end{aligned}$$

It follows that $|f| - |g| = 0$ almost everywhere, which implies $f = g$ almost everywhere. □

31.5 L^1-Spaces

Throughout this subsection, let $(X, \mathcal{M}, \mu)$ be a measure space. The set of all integrable function $f\colon X \to \mathbb{R}$ is denoted $\mathrm{Int}^1(X, \mathcal{M}, \mu)$. By Proposotion (31.9) we see that $\mathrm{Int}^1(X, \mathcal{M}, \mu)$ is an $\mathbb{R}$-vector space. We define a relation $\sim$ on $\mathrm{Int}^1(X, \mathcal{M}, \mu)$ as follows, if $f, g \in \mathrm{Int}^1(X, \mathcal{M}, \mu)$, then we set $f \sim g$ if and only if $f = g$ almost everywhere. One checks that $\sim$ is in fact an equivalence relation. We will denote

$$[\mathrm{Int}^1(X, \mathcal{M}, \mu)] := \mathrm{Int}^1(X, \mathcal{M}, \mu)/\!\sim .$$

Thus $[\mathrm{Int}^1(X, \mathcal{M}, \mu)]$ is the set of all integrable functions from X to $\mathbb{R}$ where two such functions are identified if they agree almost everywhere. Technically speaking, elements in $[\mathrm{Int}^1(X, \mathcal{M}, \mu)]$ are *cosets* of functions. Thus an element in $[\mathrm{Int}^1(X, \mathcal{M}, \mu)]$ should be expressed like $[f]$, where f is an integrable function which represents the coset $[f]$. In practice however, we tend to abuse notation by dropping the square brackets around $[f]$ altogether.

We can give $\mathrm{Int}^1(X, \mathcal{M}, \mu)$ the structure of a pseudo-normed space as follows: We define $\|\cdot\|_1\colon \mathrm{Int}^1(X, \mathcal{M}, \mu) \to [0, \infty)$ by

$$\|f\|_1 := \int_X |f|\mathrm{d}\mu$$

for all $f \in \mathrm{Int}^1(X, \mathcal{M}, \mu)$. Observe that monotonicity of integration implies subadditivity of $\|\cdot\|_1$. Also linearity of integration combined with absolute homogeneity of $|\cdot|$ implies absolute homogeneity of $\|\cdot\|_1$. The reason why $\|\cdot\|_1$ is just a pseudo-norm and not a norm is because it lacks the positive-definiteness property: there exists $f \in \mathrm{Int}^1(X, \mathcal{M}, \mu)$ such that $f \neq 0$ but $\|f\|_1 = 0$. On the other hand, $\|\cdot\|_1$ induces a genuine norm when we pass to the quotient space $[\mathrm{Int}^1(X, \mathcal{M}, \mu)]$! Indeed, we define a norm on $[\mathrm{Int}^1(X, \mathcal{M}, \mu)]$, which again denote by $\|\cdot\|_1$, by

$$\|f\|_1 := \int_X |f| \mathrm{d}\mu \tag{79}$$

for all $f \in [\mathrm{Int}^1(X, \mathcal{M}, \mu)]$. Note that (79) is well-defined by Corollary (13). In fact, Corollary (13) tells us that $[\mathrm{Int}^1(X, \mathcal{M}, \mu)]$ is the normed linear space induced by $\mathrm{Int}^1(X, \mathcal{M}, \mu)$ with respect to the equivalence relation $\sim$ introduced above. We will denote this normed linear space by

$$L^1(X, \mathcal{M}, \mu) := ([\mathrm{Int}^1(X, \mathcal{M}, \mu)], \|\cdot\|_1).$$

We often refer to $\|\cdot\|_1$ as the L^1**-norm**.

31.5.1 L^1-Completeness

We now wish to show that $L^1(X, \mathcal{M}, \mu)$ is not just any normed linear space; it is a Banach space.

Theorem 31.4. *$L^1(X, \mathcal{M}, \mu)$ is a Banach space.*

To prove this we'll use the following criterion to test for completeness in a normed linear space.

Lemma 31.5. *Let $\mathcal{X}$ be a normed linear space. Then $\mathcal{X}$ is a Banach space if and only if every absolutely convergent series in $\mathcal{X}$ is convergent.*

Proof. Suppose first that every absolutely convergent series in $\mathcal{X}$ is convergent. Let (x_n) be a Cauchy sequence in $\mathcal{X}$. To show that (x_n) is convergent, it suffices to show that a subsequence of (x_n) is convergent, by Lemma (39.1). Choose a subsequence $(x_{\pi(n)})$ of (x_n) such that

$$\|x_{\pi(n)} - x_{\pi(n-1)}\| < \frac{1}{2^n}$$

and for all $n \in \mathbb{N}$ (we can do this since (x_n) is Cauchy). Then the series $\sum_{n=1}^{\infty}(x_{\pi(n)} - x_{\pi(n-1)})$ is absolutely convergent since

$$\begin{aligned}\sum_{n=1}^{\infty} \|x_{\pi(n)} - x_{\pi(n-1)}\| &< \sum_{n=1}^{\infty} \frac{1}{2^n} \\ &= 1.\end{aligned}$$

Therefore it must be convergent, say $\sum_{n=1}^{\infty}(x_{\pi(n)} - x_{\pi(n-1)}) \to x$. On the other hand, for each $n \in \mathbb{N}$, we have

$$x_{\pi(n)} - x_{\pi(1)} = \sum_{m=1}^{n} (x_{\pi(m)} - x_{\pi(1)}).$$

In particular, $x_{\pi(n)} \to x - x_{\pi(1)}$ as $n \to \infty$. Thus $(x_{\pi(n)})$ is a convergent subsequence of (x_n).

Conversely, suppose $\mathcal{X}$ is a Banach space and suppose $\sum_{n=1}^{\infty} x_n$ is absolutely convergent. Let $\varepsilon > 0$ and choose $K \in \mathbb{N}$ such that $N \geq M \geq K$ implies

$$\sum_{n=M}^{N} \|x_n\| < \varepsilon.$$

Then $N \geq M \geq K$ implies

$$\begin{aligned}\left\| \sum_{n=1}^{N} x_n - \sum_{n=1}^{M} x_n \right\| &= \left\| \sum_{n=M}^{N} x_n \right\| \\ &\leq \sum_{n=M}^{N} \|x_n\| \\ &< \varepsilon.\end{aligned}$$

It follows that the sequence of partial sums $(\sum_{n=1}^{N} x_n)_N$ is Cauchy. Since $\mathcal{X}$ is a Banach space, it follows that $\sum_{n=1}^{\infty} x_n$ is convergent. □

Now we prove Theorem (31.4).

Proof. By Lemma (39.2), it suffices to show that every absolutely convergent sequence is convergent. Suppose (f_n) is a sequence in $L^1(X, \mathcal{M}, \mu)$ such that $\sum_{n=1}^{\infty} \|f_n\|_1 < \infty$. Denote by

$$G_N = \sum_{n=1}^{N} |f_n| \quad \text{and} \quad G = \sum_{n=1}^{\infty} |f_n|.$$

Observe that (G_N) is an increasing sequence of nonnegative measurable functions which converges pointwise to G. Therefore by MCT we have

$$\begin{aligned}
\|G\|_1 &= \lim_{N\to\infty} \|G_N\|_1 \\
&= \lim_{N\to\infty} \int_X G_N \mathrm{d}\mu \\
&= \lim_{N\to\infty} \int_X \sum_{n=1}^{N} |f_n| \mathrm{d}\mu \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \int_X |f_n| \mathrm{d}\mu \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \|f_n\|_1 \mathrm{d}\mu \\
&= \sum_{n=1}^{\infty} \|f_n\|_1 \\
&< \infty.
\end{aligned}$$

It follows that $G \in L^1(X, \mathcal{M}, \mu)$. In particular, $\{G = \infty\}$ has measure zero, so if we denote $F_N = \sum_{n=1}^{N} f_n$ and if we define $F \colon X \to \mathbb{R}$ by

$$F(x) = \begin{cases} 0 & \text{if } G(x) = \infty \\ \sum_{n=1}^{\infty} f_n(x) & \text{if } G(x) < \infty \end{cases}$$

then we see that (F_N) converges pointwise to F almost everywhere, so the sequence $(F - F_N)$ converges pointwise to the zero function almost everywhere. Note that if $G(x) < \infty$, then $\sum_{n=1}^{\infty} f_n(x)$ converges since it converges absolutely, so the definition of F is well-posed. Also note that $F_N, F \in L^1(X, \mathcal{M}, \mu)$ since $|F_N|, |F| \leq G$ and $G \in L^1(X, \mathcal{M}, \mu)$. Finally, we note that $(F - F_N)$ is dominated by the integrable function $2G$. It follows from the DCT that

$$\begin{aligned}
\lim_{N\to\infty} \|F - F_N\|_1 &= \lim_{N\to\infty} \int_X |F_N - F| \mathrm{d}\mu \\
&= \int_X 0 \mathrm{d}\mu \\
&= 0.
\end{aligned}$$

Therefore every absolutely convergent series is convergent, which implies $L^1(X, \mathcal{M}, \mu)$ is complete. □

What do we need to choose $(X, \mathcal{M}, \mu)$ to be such that $L^1(X, \mathcal{M}, \mu) = \ell^1(\mathbb{N})$. We need $X = \mathbb{N}$, $\mathcal{M} = \mathcal{P}(\mathbb{N})$, and μ to be counting measure. In this case, if $f \colon \mathbb{N} \to [0, \infty)$, then we have

$$\begin{aligned}
\|f\|_1 &= \int_{\mathbb{N}} f \mathrm{d}\mu \\
&= \sum_{n=1}^{\infty} f(n)\mu(\{n\}) \\
&= \sum_{n=1}^{\infty} f(n) \\
&= \sum_{n=1}^{\infty} |f(n)).
\end{aligned}$$

By using $f = f^+ - f^-$ we can easily see that $\|f\|_1 = \sum_{n=1}^{\infty} f(n)$.

31.5.2 Set of all Integrable Simple Functions is a Dense Subspace of $L^1(X, \mathcal{M}, \mu)$

We denote by $\mathrm{S}(X, \mathcal{M}, \mu)$ to be the set of all integrable simple functions where we identify two simple functions if they agree almost everywhere. It is easy to check that $\|\cdot\|_1$ restricts to a norm on $\mathrm{S}(X, \mathcal{M}, \mu)$ making it into a normed linear subspace of $L^1(X, \mathcal{M}, \mu)$. We now want to show that $\mathrm{S}(X, \mathcal{M}, \mu)$ is a dense subspace of $L^1(X, \mathcal{M}, \mu)$.

Proposition 31.12. *$\mathrm{S}(X, \mathcal{M}, \mu)$ is a dense subspace of $L^1(X, \mathcal{M}, \mu)$.*

Proof. Let $f \in L^1(X, \mathcal{M}, \mu)$. Decompose f into its positive and negative parts:

$$f = f^+ - f^-.$$

There exists an increasing sequence (φ_n) of nonnegative simple functions which converges to f^+ pointwise. Similarly, there exists an increasing sequence (ψ_n) of nonnegative simple functions which converges to f^- pointwise. Then $(\varphi_n + \psi_n)$ is an increasing sequence of nonnegative simple functions which converges pointwise to $|f|$. Also note that $(\varphi_n - \psi_n)$ is a sequence of simple functions which converges pointwise to f. Now set $s_n = \varphi_n - \psi_n$. We claim that $\|s_n - f\|_1 \to 0$ as $n \to \infty$. In other words, we claim that

$$\int_X |s_n - f| \mathrm{d}\mu \to 0$$

as $n \to \infty$. To this, we'll use DCT. Observe that for each $n \in \mathbb{N}$ we have

$$\begin{aligned}
|s_n - f| &\leq |s_n| + |f| \\
&= |\varphi_n - \psi_n| + |f| \\
&\leq \varphi_n + \psi_n + |f| \\
&\leq |f| + |f| \\
&\leq 2|f|
\end{aligned}$$

So $2|f|$ is a dominating function, which means we can apply DCT. Therefore

$$\begin{aligned}
\lim_{n\to\infty} \|s_n - f\|_1 &= \lim_{n\to\infty} \int_X |s_n - f| \mathrm{d}\mu \\
&= \int_X 0 \mathrm{d}\mu \\
&= 0.
\end{aligned}$$

□

31.5.3 $\mathrm{C}[a, b]$ is a dense subspace of $L^1[a, b]$

Proposition 31.13. *The space of continuous functions $\mathrm{C}[a, b]$ is a dense subspace of $L^1[a, b]$.*

Proof. By the previous proposition, it is enough to show that any simple function can be approximated arbitrarily well with a continuous function. Let φ be a simple function and write its canonical form as

$$\varphi = \sum_{k=1}^{n} c_k 1_{E_k}$$

where each E_k is a Borel subset of $[a, b]$. We proved earlier in the semester that for each $\varepsilon > 0$ and Borel set E_k there exists A_k in the interval algebra of $[a, b]$ such that $\mathrm{d}_m(E_k, A_k) < \varepsilon$ where m is the Lebesgue measure. In fact, we have

$$\mathrm{m}(E_k \Delta A_k) < \varepsilon \iff \int_a^b |1_{E_k} - 1_{A_k}| \mathrm{dm} < \varepsilon \iff \|1_{E_k} - 1_{A_k}\|_1 < \varepsilon.$$

Since each A_k in the interval algebra is a finite union of intervals, it is enough to show that 1_I, where I is a subinterval of $[a, b]$ can be approximated arbitrarily well by a continuous function. For this we can use trapezoid functions just like in HW1. Thus $L^1[a, b]$ is a completion of $C[a, b]$ with respect to $\|\cdot\|_1$ norm. □

31.6 L^p-Spaces

Throughout this subsection, let $(X, \mathcal{M}, \mu)$ be a measure space. Let $1 < p < \infty$. Just as we defined the notion of an integral function, we can also define the notion of a **p-integral** function. In particular, if f is a measurable function, then we say it is p-integral if

$$\int_X |f|^p \mathrm{d}\mu < \infty.$$

We denote by $\mathrm{Int}^p(X, \mathcal{M}, \mu)$ to be the set of all p-integral functions from X to $\mathbb{R}$. It is straightforward to check that $\mathrm{Int}^p(X, \mathcal{M}, \mu)$ is an $\mathbb{R}$-vector space just like in the case of $\mathrm{Int}^1(X, \mathcal{M}, \mu)$, and just like before, we denote $[\mathrm{Int}^p(X, \mathcal{M}, \mu)]$ to be the quotient space induced by $\sim$. We can equip $[\mathrm{Int}^p(X, \mathcal{M}, \mu)]$ with another norm, which we call the **L^p-norm**. It is defined by

$$\|f\|_p = \left(\int_X |f|^p \mathrm{d}\mu \right)^{\frac{1}{p}}$$

for all $f \in [\mathrm{Int}^p(X, \mathcal{M}, \mu)]$. We will denote this normed linear space by

$$L^p(X, \mathcal{M}, \mu) := ([\mathrm{Int}^p(X, \mathcal{M}, \mu)], \|\cdot\|_p).$$

Absolute homogeneity and positive-definiteness of $\|\cdot\|_p$ are straightforward to check. It turns out that subadditivity of $\|\cdot\|_p$ takes a little more work to show. The idea is to use the so called **Hölder's inequality**. However this inequality itself relies on another inequality called **Young's inequality**, so let's begin with that. We often refer to $\|\cdot\|_1$ as the **L^1-norm**.

31.6.1 Young's Inequality

Lemma 31.6. *Let x and y be nonnegative real numbers and let $0 < \gamma < 1$. Then we have*

$$x^\gamma y^{1-\gamma} \leq \gamma x + (1-\gamma) y. \tag{80}$$

Proof. We may assume that $x, y > 0$ since otherwise it is trivial. Set $t = x/y$ and rewrite (97) as

$$t^\gamma - \gamma t \leq 1 - \gamma. \tag{81}$$

Thus, to show (97) for all $x, y > 0$, we just need to show (98) for all $t > 0$. To see why (98) holds, define $f \colon \mathbb{R}_{>0} \to \mathbb{R}$ by

$$f(t) = t^\gamma - \gamma t$$

for all $t \in \mathbb{R}_{>0}$. Observe that f is a smooth function on $\mathbb{R}_{>0}$, with it's first derivative and second derivative given by

$$f'(t) = \gamma t^{\gamma-1} - \gamma \quad \text{and} \quad f''(t) = \gamma(\gamma-1)t^{\gamma-2}$$

for all $t \in \mathbb{R}_{>0}$ respectively. Observe that

$$\begin{aligned} f'(t) = 0 &\iff \gamma t^{\gamma-1} = \gamma \\ &\iff t^{\gamma-1} = 1 \\ &\iff t = 1, \end{aligned}$$

where the last if and only if follows from the fact that t is a positive real number. Also, we clearly have $f''(t) < 0$ for all $t \in \mathbb{R}_{>0}$. Thus, since f is concave down on all of $\mathbb{R}_{>0}$, and $f'(t) = 0$ if and only if $t = 1$, it follows that f has a global maximum at $t = 1$. In particular, we have

$$\begin{aligned} t^\gamma - \gamma t &= f(t) \\ &\leq f(1) \\ &\leq 1^\gamma - \gamma \cdot 1 \\ &= 1 - \gamma \end{aligned}$$

for all $t \in \mathbb{R}_{>0}$. □

Proposition 31.14. *(Young's inequality) Let a and b be nonnegative real numbers and let $1 \leq p, q < \infty$ such that $1/p + 1/q = 1$. Then we have*

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Proof. Set $\gamma = 1/p$ (so $1 - \gamma = 1/q$), $a = x^\gamma$, and $b = y^{1-\gamma}$. Then Young's Inequality becomes (80), which was proved in Lemma (31.6). □

31.6.2 Hölder's Inequality

Proposition 31.15. *(Hölder's inequality) Let $1 < p, q < \infty$ such that $1/p + 1/q = 1$, let $f \in L^p(X, \mathcal{M}, \mu)$, and let $g \in L^q(X, \mathcal{M}, \mu)$. Then $fg \in L^1(X, \mathcal{M}, \mu)$ and*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q. \tag{82}$$

Proof. If $\|f\|_p = 0$ or $\|g\|_q = 0$, then one of them is equal to 0 almost everywhere. In this case $fg = 0$ almost everywhere. Thus the inquality is trivial in this case, so we may assume that $\|f\|_p \neq 0$ and $\|g\|_q \neq 0$. We will first show the inequality in the special case where $\|f\|_p = 1 = \|g\|_q$. Then the righthand side of (82) is 1, so we need to show $\|fg\|_1 \leq 1$. This follows immediately from Young's inequality:

$$\begin{aligned}
\|fg\|_1 &= \int_X |fg| \mathrm{d}\mu \\
&\leq \int_X \left(\frac{|f|^p}{p} + \frac{|g|^q}{q} \right) \mathrm{d}\mu \\
&= \frac{1}{p} \int_X |f|^p \mathrm{d}\mu + \frac{1}{q} \int_X |g|^q \mathrm{d}\mu \\
&= \frac{1}{p} + \frac{1}{q} \\
&= 1.
\end{aligned}$$

Now we prove the general case. Let $F = f/\|f\|_p$ and $G = g/\|g\|_q$. Then $\|F\|_p = 1 = \|G\|_q$. Applying the special case that we just proved, we have

$$\begin{aligned}
1 &\geq \int_X |FG| \mathrm{d}\mu \\
&= \int_X \left| \frac{f}{\|f\|_p} \frac{g}{\|g\|_q} \right| \mathrm{d}\mu \\
&= \frac{1}{\|f\|_p \|g\|_q} \int_X |fg| \mathrm{d}\mu \\
&= \frac{1}{\|f\|_p \|g\|_q} \|fg\|_1.
\end{aligned}$$

After multiplying both sides by $\|f\|_p \|g\|_q$, we obtain Hölder's inequality. □

31.6.3 Minkowski's Inequality

For historical reasons, subadditivity of $\| \cdot \|_p$ is refered to as **Minkowski's inequality**.

Proposition 31.16. *(Minkowski's inequality) Let $1 \leq p < \infty$ and let $f, g \in L^p(X, \mathcal{M}, \mu)$. Then*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Proof. We proved this for $p = 1$, so let $p > 1$. If $\|f + g\|_p = 0$, then the inequality is obvious, so we can assume

$\|f+g\|_p > 0$. Observe that for q such that $1/p + 1/q = 1$, we have $(p-1)q = p$. Thus we have

$$\begin{aligned}
\|f+g\|_p &= \||f+g|^p\|_1^{1/p} \\
&= \||f+g|^p\|_1^{1-1/q} \\
&= \||f+g|^p\|_1 \||f+g|^p\|_1^{-1/q} \\
&= \||f+g||f+g|^{p-1}\|_1 \||f+g|^p\|_1^{-1/q} \\
&\leq \left(\||f||f+g|^{p-1}\|_1 + \||g||f+g|^{p-1}\|_1\right) \||f+g|^p\|_1^{-1/q} \\
&\leq \left(\|f\|_p\||f+g|^{p-1}\|_q + \|g\|_p\||f+g|^{p-1}\|_q\right) \||f+g|^p\|_1^{-1/q} \\
&= (\|f\|_p + \|g\|_p)\||f+g|^{p-1}\|_q \||f+g|^p\|_1^{-1/q} \\
&= (\|f\|_p + \|g\|_p)\||f+g|^{q(p-1)}\|_1^{1/q} \||f+g|^p\|_1^{-1/q} \\
&= (\|f\|_p + \|g\|_p)\||f+g|^{p}\|_1^{1/q} \||f+g|^p\|_1^{-1/q} \\
&= \|f\|_p + \|g\|_p.
\end{aligned}$$

□

31.7 Types of Convergences

In this subsection, we wish to to discuss various types of convergences of sequences of functions.

Definition 31.7. Let $(X, \mathcal{M}, \mu)$ be a measure space, let $(f_n \colon X \to \mathbb{R})$ be a sequence of functions, and let $f \colon X \to \mathbb{R}$ be a function.

1. We say (f_n) converges **pointwise** to f, denoted $f_n \xrightarrow{\text{pw}} f$, if for all $x \in X$ and for all $\varepsilon > 0$ there exists $N_{x,\varepsilon} \in \mathbb{N}$ (which depends on x and ε) such that $n \geq N_{x,\varepsilon}$ implies
$$|f_n(x) - f(x)| < \varepsilon.$$
2. We say (f_n) converges **uniformly** to f, denoted $f_n \xrightarrow{\text{u}} f$, if for all $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ (which only depends on ε) such that $n \geq N_\varepsilon$ implies
$$|f_n(x) - f(x)| < \varepsilon$$
3. We say (f_n) converges to f **almost uniformly**, denoted $f_n \xrightarrow{\text{au}} f$, if for all $\delta > 0$ there exists a set $E_\delta \subseteq X$ with $\mu(E_\delta) < \delta$ such that $f_n \to f$ uniformly on E_δ^c. In other words, (f_n) converges to f almost uniformly if for all $\varepsilon, \delta > 0$ there exists $E_\delta \subseteq X$ with $\mu(E_\delta) < \delta$ and $N_{\varepsilon,\delta} \in \mathbb{N}$ such that $n \geq N_\varepsilon$ implies
$$|f_n(x) - f(x)| < \varepsilon$$
for all $x \in E_\delta^c$.
4. We say (f_n) converges to f in **measure zero**, denoted $f_n \xrightarrow{\text{m}} f$, if for all $\varepsilon > 0$ we have
$$\mu\{f_n - f \geq \varepsilon\} \to 0$$
as $n \to \infty$. In other words, for all $\varepsilon, \delta > 0$, there exists $N_{\varepsilon,\delta} \in \mathbb{N}$ such that $n \geq N_{\varepsilon,\delta}$ implies
$$\mu\{f_n - f \geq \varepsilon\} < \delta.$$
5. A convergence is said to hold **almost everywhere** if there exists a set $E \in \mathcal{M}$ such that $\mu(E) = 0$ and such that the convergence holds on E^c.

31.7.1 Almost Pointwise is Equivalent to Pointwise Almost Everywhere

It may seem reasonable to make an additonal defintion: we say (f_n) converges **almost pointwise** to f if for all $\delta > 0$ there exists $E_\delta \in \mathcal{M}$ such that $\mu(E_\delta) < \delta$ and (f_n) converges pointwise to f on E_δ^c. It's clear that if (f_n)

converges to f pointwise almost everywhere, then (f_n) converges almost pointwise to f. In fact, the converse holds too! So "almost pointwise" and "pointwise almost everywhere" are equivalent notions. Let us show this.

Proposition 31.17. *Let $(X, \mathcal{M}, \mu)$ be a measure space, let $(f_n \colon X \to \mathbb{R})$ be a sequence of functions, and let $f \colon X \to \mathbb{R}$ be a function. Then (f_n) converges almost pointwise to f if and only if (f_n) converges to f pointwise almost everywhere.*

Proof. We just need to show that almost pointwise implies pointwise almost everywhere, since the converse direction is trivial. Suppose (f_n) converges almost pointwise to f. For each $k \in \mathbb{N}$, choose $E_k \in \mathcal{M}$ such that $\mu(E_k) < 1/k$ and (f_n) converges to f pointwise on E_k^c. Set

$$E = \bigcap_{k=1}^{\infty} E_k.$$

Clearly we have $\mu(E) = 0$. We claim that (f_n) converges to pointwise to f on E^c. Indeed, let $x \in E^c$ and let $\varepsilon > 0$. Since

$$E^c = \bigcup_{k=1}^{\infty} E_k^c,$$

there exists a $k \in \mathbb{N}$ such that $x \in E_k^c$, so we choose such a $k \in \mathbb{N}$. Then since (f_n) converges pointwise to f on E_k^c, we can choose $N \in \mathbb{N}$ such that $n \geq N$ implies

$$|f_n(x) - f(x)| < \varepsilon.$$

It follows that (f_n) converges pointwise to f almost everywhere. Note that N depends on k, ε, and x, but this isn't a problem! □

31.7.2 Uniform Convergence on Finite Measure Space Implies L^p Convergence

Proposition 31.18. *Let $(X, \mathcal{M}, \mu)$ be a finite measure space, let $(f_n \colon X \to \mathbb{R})$ be a sequence of functions, let $f \colon X \to \mathbb{R}$ be a function, and let $0 < p < \infty$. Suppose (f_n) converges to f uniformly on X. Then (f_n) converges to f in the L^p-norm.*

Proof. Let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies

$$|f_n(x) - f(x)| < \frac{\varepsilon^{1/p}}{\mu(X)}$$

for all $x \in X$. Then $n \geq N$ implies

$$\begin{aligned} \|f_n - f\|_p &= \int_X |f_n - f|^p \mathrm{d}\mu \\ &\leq \sup |f_n - f|^p \mu(X) \\ &< \frac{\varepsilon^{1/p}}{\mu(X)} \mu(X) \\ &= \varepsilon. \end{aligned}$$

It follows that (f_n) converges to f in the L^p-norm. □

31.7.3 Convergence in L^p Implies Convergence in Measure Zero

Proposition 31.19. *Let $(X, \mathcal{M}, \mu)$ be a measure space, let $(f_n \colon X \to \mathbb{R})$ be a sequence of functions, let $f \colon X \to \mathbb{R}$ be a function, and let $0 < p < \infty$. Suppose (f_n) converges to f in the L^p-norm. Then (f_n) converges to f in measure zero.*

Proof. Let $\varepsilon, \delta > 0$. Choose $N_{\varepsilon,\delta} \in \mathbb{N}$ such that $n \geq N_{\varepsilon,\delta}$ implies

$$\|f_n - f\|_p < \varepsilon^{1/p} \delta^{1/p}.$$

Then it follows from Chebyshev's inequality that

$$\begin{aligned} \mu\{|f_n - f| \geq \varepsilon^{1/p}\} &\leq \mu\{|f_n - f|^p \geq \varepsilon\} \\ &\leq \frac{1}{\varepsilon} \|f_n - f\|_p^p \\ &< \frac{1}{\varepsilon} \varepsilon \delta \\ &= \delta. \end{aligned}$$

Therefore (f_n) converges to f in meaure zero. □

31.7.4 Convergence in L^p Does Not Imply Convergence Pointwise Almost Everywhere and Vice-Versa

Example 31.1. Let $(X, \mathcal{M}, \mu)$ be a measure space, let $(f_n : X \to \mathbb{R})$ be a sequence of functions, let $f : X \to \mathbb{R}$ be a function, and let $0 < p < \infty$. Clearly pointwise almost everywhere convergence does not imply L^p-convergence. It also turns out that if (f_n) converges to f in the L^p-norm, then it is not necessarily the case that (f_n) converges to f pointwise almost everywhere. Similarly, if (f_n) converges to f pointwise almost everywhere, then it is not necessarily the case that (f_n) converges to f in the L^p-norm. To see this, consider the case where $X = [0, 1]$, $\mu = \mathrm{m}$ is the Lebesgue measure, and (f_n) is the sequence of functions which starts out as

$$\begin{aligned}
f_1 &= 1_{[0,1]} \\
f_2 &= 1_{[0,1/2]} \\
f_3 &= 1_{[1/2,1]} \\
f_4 &= 1_{[0,1/4]} \\
f_5 &= 1_{[1/4,1/2]} \\
f_6 &= 1_{[1/2,3/4]} \\
f_7 &= 1_{[3/4,1]} \\
f_8 &= 1_{[0,1/8]} \\
&\vdots
\end{aligned}$$

and so on. It is easy to check that (f_n) converges to the 0 function in the L^1-norm. However it does not converge pointwise almost everywhere to any function! Indeed, for any $x \in [0, 1]$, we have $f_n(x) = 0$ for infinitely many n and $f_n(x) = 1$ for infinitely many n.

31.7.5 Convergence in Measure Zero Implies a Subsequence Converges Pointwise Almost Everywhere

Lemma 31.7. *Let $(X, \mathcal{M}, \mu)$ be a measure space and let (E_n) be a sequence in $\mathcal{M}$ such that $\sum_{n=1}^{\infty} \mu(E_n) < \infty$. Then*

$$\mu(\limsup E_n) = 0.$$

Proof. Note that the sequence

$$\left(\bigcup_{n \geq N} E_n \right)_{N \in \mathbb{N}}$$

is a descending sequence in N. This together with the fact that $\mu\left(\bigcup_{n=1}^{\infty} E_n\right) < \infty$ implies

$$\begin{aligned}
\mu(\limsup E_n) &= \mu\left(\bigcap_{N=1}^{\infty} \left(\bigcup_{n \geq N} E_n \right) \right) \\
&= \lim_{N \to \infty} \mu\left(\bigcup_{n \geq N} E_n \right) \\
&\leq \lim_{N \to \infty} \sum_{n=N}^{\infty} \mu(E_n) \\
&= 0,
\end{aligned}$$

where the last equality follows since $\sum_{n=1}^{\infty} \mu(E_n) < \infty$. □

Proposition 31.20. *Let $(X, \mathcal{M}, \mu)$ be a measure space, let $(f_n : X \to \mathbb{R})$ be a sequence of functions, and let $f : X \to \mathbb{R}$ be a function. Suppose (f_n) converges to f in measure zero. Then there exists a subsequence $(f_{\pi(n)})$ of (f_n) such that $(f_{\pi(n)})$ converges to f pointwise almost everywhere.*

Proof. For each $m \in \mathbb{N}$ choose $\pi(m) \geq m$ such that $n \geq \pi(m)$ implies

$$\mu\{|f_n - f| \geq 1/2^m\} < 1/2^m. \tag{83}$$

In fact, we don't need to know that (83) holds for *every* $n \geq \pi(m)$, we just need to know that it holds for $n = \pi(m)$! Indeed, for each $m \in \mathbb{N}$, denote

$$E_m = \{|f_{\pi(m)} - f| \geq 1/2^m\}$$

Then $\mu(E_m) < 1/2^m$ since this is a special case of (83) where $n = \pi(m)$. Next, denote

$$E = \limsup E_m = \bigcap_{M \geq 1} \bigcup_{m \geq M} E_m$$

Observe that $\sum_{m=1}^{\infty} \mu(E_m) < 1$. It follows from Lemma (31.7) that $\mu(E) = 0$. We claim that (f_n) converges pointwise to f on E^c. To see this, let $x \in E^c$. Then there exists some M such that $x \notin \bigcup_{m \geq M} E_m$. In other words, there exists some M such that $x \notin E_m$ for any $m \geq M$. In other words, there exists some M such that

$$|f_{\pi(m)}(x) - f(x)| < 1/2^m$$

for all $m \geq M$. Therefore $f_{\pi(m)}(x) \to f(x)$ as $m \to \infty$. This implies $(f_{\pi(m)})$ converges to f pointwise almost everywhere. □

Corollary 14. *Let $(X, \mathcal{M}, \mu)$ be a measure space, let $(f_n \colon X \to \mathbb{R})$ be a sequence of functions, and let $f \colon X \to \mathbb{R}$ be a function. Suppose (f_n) converges to f in the L^p-norm. Then there exists a subsequence $(f_{\pi(n)})$ of (f_n) such that $(f_{\pi(n)})$ converges to f pointwise almost everywhere.*

31.7.6 Convergence Pointwise Almost Everywhere on a Finite Measure Space Implies Almost Uniform Convergence

Proposition 31.21. *Let $(X, \mathcal{M}, \mu)$ be a finite measure space, let $(f_n \colon X \to \mathbb{R})$ be a sequence of functions, let $f \colon X \to \mathbb{R}$ be a function, and let $0 < p < \infty$. Suppose (f_n) converges to f pointwise almost everywhere. Then (f_n) converges to f almost uniformly.*

Proof. Let $\varepsilon, \delta > 0$. We must find an $E_\delta \in \mathcal{M}$ and an $N_{\varepsilon,\delta} \in \mathbb{N}$ such that $\mu(E_\delta) < \delta$ and $n \geq N_{\varepsilon,\delta}$ implies

$$|f_n(x) - f(x)| < \frac{1}{k}$$

for all $x \in E_\delta^c$. Before we begin with the proof, let us state up front what E_δ and $N_{\varepsilon,\delta}$ will be. We will have

$$E_\delta = \bigcup_{k=1}^{\infty} \bigcup_{n \geq \pi(k)} \{|f_n - f| \geq 1/k\}.$$

Here, $\pi(k) \geq k$ will be chosen large enough such that $\mu(E_\delta) < \delta$. So $x \in E_\delta$ if and only if there exists $k \in \mathbb{N}$ and an $n \geq \pi(k)$ such that

$$|f_n(x) - f(x)| \geq \frac{1}{k}$$

and $x \in E_\delta^c$ if and only if for all $k \in \mathbb{N}$ we have

$$|f_n(x) - f(x)| < \frac{1}{k}$$

for all $n \geq \pi(k)$. Thus if we choose $k \in \mathbb{N}$ such that $1/k < \varepsilon$, then we set $N_{\varepsilon,\delta} = \pi(k)$. It will then follow that $n \geq N_{\varepsilon,\delta}$ implies

$$|f_n(x) - f(x)| < \varepsilon$$

for all $x \in E_\delta^c$. This will imply almost uniform convergence. So the main goal of the proof is to find $\pi(k)$ for each k such that $\mu(E_\delta) < \delta$.

For each $k, N \in \mathbb{N}$, let

$$E_N(k) = \bigcup_{n \geq N} \{|f_n - f| \geq 1/k\}.$$

Observe that $(E_N(k))_{N \in \mathbb{N}}$ is a descending sequence of sets in N. Furthermore, we observe that

$$\mu\left(\bigcap_{N=1}^{\infty} E_N(k)\right) = 0.$$

Indeed, $x \in \bigcap_{N=1}^{\infty} E_N(k)$ if and only if there exists a subsequence $(f_{\rho(n)})$ of (f_n) such that

$$|f_{\rho(n)}(x) - f(x)| \geq 1/k$$

for all $n \in \mathbb{N}$ if and only if $f_n(x) \not\to f(x)$. This set has measure zero by assumption. Now, since $\mu(X) < \infty$, we have $\mu(E_1(k)) < \infty$, and thus

$$\begin{aligned}\lim_{N\to\infty} \mu(E_N(k)) &= \mu\left(\bigcap_{N=1}^{\infty} E_N(k)\right) \\ &= 0.\end{aligned}$$

It follows that for each $k \in \mathbb{N}$, we can choose $\pi(k) \in \mathbb{N}$ such that

$$\mu(E_{\pi(k)}(k)) < \frac{\delta}{2^k}.$$

Now as we discuss earlier, we set $E_\delta = \bigcup_{k=1}^{\infty} E_{\pi(k)}(k)$. Then we have

$$\begin{aligned}\mu(E_\delta) &= \mu\left(\bigcup_{k=1}^{\infty} E_{\pi(k)}(k)\right) \\ &\leq \sum_{k=1}^{\infty} \mu\left(E_{\pi(k)}(k)\right) \\ &< \sum_{k=1}^{\infty} \frac{\delta}{2^k} \\ &= \delta.\end{aligned}$$

□

31.7.7 Convergences Diagram

We summarize our findings in the convergence implications diagram below.

$$\begin{tikzcd}
 & f_n \xrightarrow{\mathrm{u}} f \arrow[dl] \arrow[dr, "\mu(X)<\infty"] & \\
f_n \xrightarrow{\mathrm{au}} f \arrow[d, shift left] \arrow[drr] & & f_n \xrightarrow{L^1} f \arrow[d] \\
f_n \xrightarrow{\mathrm{pwae}} f \arrow[u, shift left, "\mu(X)<\infty"] & & f_n \xrightarrow{\mathrm{m}} f \arrow[ll, dashed, "\text{subseq}"'] \arrow[ull, dashed, "\text{subseq}"']
\end{tikzcd}$$

An unlabeled undashed arrow indicates implication. For instance, $f_n \xrightarrow{\mathrm{u}} f$ implies $f_n \xrightarrow{\mathrm{au}} f$. A labeled undashed arrows indicates implication if the labeled condition holds. For instance, if $\mu(X) < \infty$, then $f_n \xrightarrow{\mathrm{u}} f$ implies $f_n \xrightarrow{L^1} f$. Finally a labeled dashed arrow indicaties partial implication. For instance, $f_n \xrightarrow{\mathrm{m}} f$ implies $f_{\pi(n)} \xrightarrow{\mathrm{pwae}} f$ for some subsequence $(f_{\pi(n)})$ of (f_n).

32 Product Measures

Recall the following HW problem:

$$\sum_{n=1}^{\infty} \int_X f_n \mathrm{d}\mu = \int_X \left(\sum_{n=1}^{\infty} f_n\right) \mathrm{d}\mu \tag{84}$$

whenever f_n is a nonnegative measurable function. We can think of $\sum_{n=1}^{\infty}$ as an integral on $\mathbb{N}$ with respect to the counting measure. If we denote the counting measure by σ, we can write (84) as

$$\int_{\mathbb{N}} \int_X f(x,n) \mathrm{d}\mu(x) \mathrm{d}\sigma(n) = \int_X \int_{\mathbb{N}} f(x,n) \mathrm{d}\sigma(n) \mathrm{d}\mu(x).$$

Thus we can exchange the order of integration whenever $f(x,n) \geq 0$.

32.1 Definining the Product σ-Algebra

Throughout the rest of this section, let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be two finite measure spaces. We want to equip $X \times Y$ with a σ-algebra from the σ-algebras $\mathcal{M}$ and $\mathcal{N}$ and we want to equip a measure to this σ-algebra from the measures μ and ν. Clearly any set of the form $A \times B$ where $A \in \mathcal{M}$ and $B \in \mathcal{N}$ shoud be included in this σ-algebra that we want to construct. These sets are called **measurable rectangles**. Let us denote by $\mathcal{R}(\mathcal{M}, \mathcal{N})$ to the collection of all measurable rectangles. If $\mathcal{M}$ and $\mathcal{N}$ are understood from context (as in our case at the moment), then we simply denote this by $\mathcal{R}$ rather than $\mathcal{R}(\mathcal{M}, \mathcal{N})$. Observe that $\mathcal{R}$ forms a semialgebra. Indeed,

1. It is closed under finite intersections: $(A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \cap A_2) \times (B_1 \cap B_2)$
2. Complements can be expressed as finite pairwise disjoint unions: $(A \times B)^c = (X \times B^c) \cup (A^c \times Y)$

If $\mathcal{M}$ and $\mathcal{N}$ are understood from context, then we simply denote this by $\mathcal{R}$ rather than $\mathcal{R}(\mathcal{M}, \mathcal{N})$. Unfortunately $\mathcal{R}$ does not form a σ-algebra. The **product σ-algebra** $\mathcal{M} \otimes \mathcal{N}$ is defined to be the smallest σ-algebra which contains all the measurable rectangles. Thus

$$\mathcal{M} \otimes \mathcal{N} = \sigma(\mathcal{R}).$$

On the other hand, $\mathcal{R}$ forms a semi-algebra:

1. It is closed under finite intersections: $(A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \cap A_2) \times (B_1 \cap B_2)$
2. Complements can be expressed as finite pairwise disjoint unions: $(A \times B)^c = (X \times B^c) \cup (A^c \times B)$

Therefore the collection $\mathcal{A}$ formed by the collection of all finite disjoint unions of members of $\mathcal{R}$ forms an algebra. Whenever we start with a finite premeasure on a semialgebra $\mathcal{R}$, we can always extend it (uniquely!) to a finite premeasure to $\mathcal{A}$, and then extend it (uniquely!) to a finite measure on $\sigma(\mathcal{A})$. With this in mind, we define the **product σ-algebra**, denoted $\mathcal{M} \otimes \mathcal{N}$, to be the smallest σ-algebra which contains all the measurable rectangles. Thus

$$\mathcal{M} \otimes \mathcal{N} = \sigma(\mathcal{A}).$$

We define a finite measure on $\mathcal{M} \otimes \mathcal{N}$, denoted $\mu \otimes \nu \colon \mathcal{M} \otimes \mathcal{N} \to [0, \infty]$, by first defining on $\mathcal{R}$ by

$$(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B)$$

for all $A \times B \in \mathcal{R}$, and then extending it uniquely to all $\mathcal{A}$, and then finally extending it uniquely to all of $\mathcal{M} \otimes \mathcal{N}$. The extension from $\mathcal{A}$ to $\mathcal{M} \otimes \mathcal{N}$ is obtained from Theorem (30.1). How do we extend it from $\mathcal{R}$ to $\mathcal{A}$? We do this as follows: let $A_1 \times B_1, \dots, A_n \times B_n$ be a pairwise disjoint sequence of measurable rectangles. Then we define

$$(\mu \otimes \nu) \left(\bigcup_{i=1}^{n} A_i \times B_i \right) = \sum_{i=1}^{n} \mu(A_i)\nu(B_i). \tag{85}$$

One must be careful, as we need to check that (85) is well-defined. In particular, suppose $A \times B = \bigcup_{i=1}^{n} (A_i \times B_i)$ where $A_1 \times B_1, \dots, A_n \times B_n$ is a pairwise disjoint sequence of measurable rectangles. Then one must show that

$$\mu(A)\nu(B) = \sum_{i=1}^{n} \mu(A_i)\nu(B_i).$$

We leave this as an exercise.

32.2 Sections

Definition 32.1. Let $E \subseteq X \times Y$ and let $f\colon X \times Y \to \mathbb{R}$.

1. For each $x \in X$, we define the x**-section** of E, denoted E_x, to be the set

$$E_x = \{y \in Y \mid (x,y) \in E\}.$$

   We also define the x-**section** of f, denoted f_x, to be the function $f_x\colon Y \to \mathbb{R}$ given by

$$f_x(y) = f(x,y)$$

   for all $y \in Y$. Observe that if $g\colon X \times Y \to \mathbb{R}$ is another function, then we have $f_x + g_x = (f+g)_x$. Indeed, for any $y \in Y$, we have

$$\begin{aligned}(f_x + g_x)(y) &= f_x(y) + g_x(y)\\ &= f(x,y) + g(x,y)\\ &= (f+g)(x,y)\\ &= (f+g)_x(y).\end{aligned}$$

   Similarly, if $a \in \mathbb{R}$, then we have $(af)_x = af_x$. Indeed, for any $y \in Y$, we have

$$\begin{aligned}(af)_x(y) &= (af)(x,y)\\ &= af(x,y)\\ &= af_x(y).\end{aligned}$$

   Thus we can view $(-)_x$ as an $\mathbb{R}$-linear map from the set of functions from $X \times Y \to \mathbb{R}$ to itself. Another important property that we observe which is easy to prove is that $(1_E)_x = 1_{E_x}$.

2. For each $y \in Y$, we define the y**-section** of E, denoted E^y, to be the set

$$E^y = \{x \in X \mid (x,y) \in E\}.$$

   We also define the y-**section** of f, denoted f^y, to be the function $f^y\colon X \to \mathbb{R}$ given by

$$f^y(x) = f(x,y)$$

   for all $x \in X$. Similar to what mentioned above, we can view $(-)^y$ an $\mathbb{R}$-linear map from the set of functions from $X \times Y \to \mathbb{R}$ to itself, and we also have $(1_E)^y = E^y$.

Proposition 32.1.

Proposition 32.2. *The following statements hold:*

1. *For any $E \in \mathcal{M} \otimes \mathcal{N}$, we have $E_x \in \mathcal{N}$ for all $x \in X$. Similarly, for any $E \in \mathcal{M} \otimes \mathcal{N}$, we have $E^y \in \mathcal{M}$ for all $y \in Y$.*

2. *For any $E \in \mathcal{M} \otimes \mathcal{N}$, the function $\nu(E_{(-)})\colon X \to \mathbb{R}$ defined by*

$$\nu(E_{(-)})(x) = \nu(E_x)$$

for all $x \in X$ is $\mathcal{M}$-measurable. Similarly, for any $E \in \mathcal{M} \otimes \mathcal{N}$, the function $\mu(E^{-})\colon Y \to \mathbb{R}$ defined by

$$\mu(E^{(-)})(y) = \mu(E^y)$$

for all $y \in Y$ is $\mathcal{N}$-measurable.

3. *For any $E \in \mathcal{M} \otimes \mathcal{N}$, we have*

$$\int_X \nu(E_{(-)})\mathrm{d}\mu = (\mu \otimes \nu)(E).$$

Similarly, for any $E \in \mathcal{M} \otimes \mathcal{N}$, we have

$$(\mu \otimes \nu)(E) = \int_Y \mu(E^{(-)})\mathrm{d}\nu.$$

Proof. 1. It suffices to show the first part of this statement since the proof of the second part is completely symmetrical to the proof the first part. Let $x \in X$ and define

$$\mathcal{C} = \{E \in \mathcal{M} \otimes \mathcal{N} \mid E_x \in \mathcal{N}\}.$$

We will show that $\mathcal{C}$ is a σ-algebra which contains $\mathcal{R}$, and this will force $\mathcal{C} = \mathcal{M} \otimes \mathcal{N}$ (which implies $E_x \in \mathcal{N}$ for all $E \in \mathcal{M} \otimes \mathcal{N}$ and for all $x \in X$). First, let us show that $\mathcal{C}$ contains R. Let $A \times B \in \mathcal{R}$ (so $A \in \mathcal{M}$ and $B \in \mathcal{N}$). Then

$$(A \times B)_x = \begin{cases} B & \text{if } x \in A \\ \emptyset & \text{if } x \in A^c \end{cases}$$

In either case, we see that $(A \times B)_x \in \mathcal{N}$, and thus $A \times B \in \mathcal{C}$. Thus $\mathcal{C}$ contains $\mathcal{R}$.

Now we will show that $\mathcal{C}$ is a σ-algebra. It is nonempty since $\mathcal{R} \subseteq \mathcal{C}$. Let us show that $\mathcal{C}$ is closed under countable unions. Let (E_n) be a countable sequence of members of $\mathcal{C}$. Observe that

$$\begin{aligned} y \in \left(\bigcup_{n=1}^{\infty} E_n\right)_x &\iff (x,y) \in \bigcup_{n=1}^{\infty} E_n \\ &\iff (x,y) \in E_n \text{ for some } n \\ &\iff y \in (E_n)_x \text{ for some } n \\ &\iff y \in \bigcup_{n=1}^{\infty} (E_n)_x. \end{aligned}$$

Therefore

$$\left(\bigcup_{n=1}^{\infty} E_n\right)_x = \bigcup_{n=1}^{\infty} (E_n)_x \in \mathcal{N}.$$

This shows that $\mathcal{C}$ is closed under countable unions. Now let us show that $\mathcal{C}$ is closed under complements. Let $E \in \mathcal{C}$. Then observe that

$$\begin{aligned} y \in (E^c)_x &\iff (x,y) \in E^c \\ &\iff (x,y) \notin E \\ &\iff y \notin E_x \\ &\iff y \in (E_x)^c. \end{aligned}$$

Therefore

$$(E^c)_x = (E_x)^c \in \mathcal{N}.$$

This shows that $\mathcal{C}$ is closed under complements. Thus we have shown $\mathcal{C}$ is a σ-algebra which contains $\mathcal{R}$.

2. It suffices to show the first part of this statement since the proof of the second part is completely symmetrical to the proof the first part. We proceed as in the proof of part 1. Let

$$\mathcal{C} = \{E \in \mathcal{M} \otimes \mathcal{N} \mid \nu(E_{(-)}) \text{ is } \mathcal{M}\text{-measurable}\}.$$

We will show that $\mathcal{C}$ is a σ-algebra which contains $\mathcal{R}$ (just like in the proof of part 1). First we show it contains $\mathcal{R}$. Let $A \times B \in \mathcal{R}$ and let $c \in \mathbb{R}$. First note that for any $x \in X$, we have

$$\nu((A \times B)_x) = \begin{cases} \nu(B) & \text{if } x \in A \\ 0 & \text{if } x \in A^c \end{cases}$$

In particular, we see that $\nu(A \times B)_{(-)}$ is a simple function, namely the simple function

$$\nu(A \times B)_{(-)} = \nu(B) 1_A.$$

It follows easily from this that $\nu(A \times B)_{(-)}$ is measurable. Thus, since $A \times B \in \mathcal{R}$ is arbitrary, we see that $\mathcal{C}$ contains $\mathcal{R}$.

Now we will show that $\mathcal{C}$ is a σ-algebra. Let us first show that it is closed under complements. Let $E \in \mathcal{C}$. Then observe that for all $x \in X$, we have

$$\begin{aligned} \nu((E^c)_x) &= \nu((E_x)^c) \\ &= \nu(Y) - \nu(E_x), \end{aligned}$$

where we used the fact that ν if finite. Thus, $\nu((E^c)_{(-)}) = \nu(Y) - \nu(E_{(-)})$, which implies $\nu((E^c)_{(-)})$ is a $\mathcal{M}$-measurable function. Thus $E^c \in \mathcal{C}$ and hence $\mathcal{C}$ is closed under complements. Now let us show that $\mathcal{C}$ is closed under ascending unions. Let (E_n) be an ascending sequence in $\mathcal{C}$. Then observe that for all $x \in X$, we have

$$\begin{aligned} \nu\left(\left(\bigcup_{n=1}^{\infty} E_n\right)_x\right) &= \nu\left(\bigcup_{n=1}^{\infty} (E_n)_x\right) \\ &= \lim_{n \to \infty} \nu((E_n)_x) \end{aligned}$$

In particular, the sequence of $\mathcal{M}$-measurable functions $(\nu(E_n)_{(-)})$ converges pointwise to the function $\nu\left((\bigcup_{n=1}^{\infty} E_n)_{(-)}\right)$. Thus $\nu\left((\bigcup_{n=1}^{\infty} E_n)_{(-)}\right)$ is a $\mathcal{M}$-measurable function, which implies $\bigcup_{n=1}^{\infty} E_n \in \mathcal{C}$. Thus we have shown $\mathcal{C}$ is a σ-algebra which contains $\mathcal{R}$.

3. It suffices to show the first part of this statement since the proof of the second part is completely symmetrical to the proof the first part. We proceed as in the proof of parts 1 and 2. Let

$$\mathcal{C} = \{E \in \mathcal{M} \otimes \mathcal{N} \mid \int_X \nu(E_{(-)}) = (\mu \otimes \nu)(E)\}.$$

We will show that $\mathcal{C}$ is a σ-algebra which contains $\mathcal{R}$ (just like in the proofs of part 1 and 2). First we show it contains $\mathcal{R}$. Let $A \times B \in \mathcal{R}$. Then we have

$$\begin{aligned} \int_X \nu((A \times B)_{(-)}) \mathrm{d}\mu &= \int_X \nu(B) 1_A \mathrm{d}\mu \\ &= \mu(A)\nu(B) \\ &= (\mu \otimes \nu)(A \times B). \end{aligned}$$

Thus, since $A \times B \in \mathcal{R}$ is arbitrary, we see that $\mathcal{C}$ contains $\mathcal{R}$.

Now we will show that $\mathcal{C}$ is a σ-algebra. Let us first show that it is closed under complements. Let $E \in \mathcal{C}$. Then we have

$$\begin{aligned} \int_X \nu((E^c)_{(-)}) \mathrm{d}\mu &= \int_X (\nu(Y) - \nu(E_{(-)})) \mathrm{d}\mu \\ &= \int_X \nu(Y) \mathrm{d}\mu - \int_X \nu(E_{(-)}) \mathrm{d}\mu \\ &= \mu(X)\nu(Y) - (\mu \otimes \nu)(E) \\ &= (\mu \otimes \nu)(X \times Y) - (\mu \otimes \nu)(E) \\ &= (\mu \otimes \nu)(E^c). \end{aligned}$$

Thus $E^c \in \mathcal{C}$, and hence $\mathcal{C}$ is closed under complements. Here, we need to remind the the reader that the manipulations we did above crucially depend on the fact that both $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are finite measure spaces. For instance, to get from the second line to the first line, we needed to use the fact that $\nu(E_-)$ is integrable (and not just $\mathcal{M}$-measurable). The fact $\nu(E_-)$ is integrable follows from the fact that both $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are finite measure spaces.

Now let us show that $\mathcal{C}$ is closed under ascending unions. Let (E_n) be an ascending sequence in $\mathcal{C}$. Then we have

$$\begin{aligned}\int_X \nu\left(\left(\bigcup_{n=1}^{\infty} E_n\right)_{(-)}\right) \mathrm{d}\mu &= \int_X \lim_{n\to\infty} \nu((E_n)_{(-)}) \mathrm{d}\mu \\ &= \lim_{n\to\infty} \int_X \nu((E_n)_{(-)}) \mathrm{d}\mu \\ &= \lim_{n\to\infty} (\mu \otimes \nu)(E_n) \\ &= (\mu \otimes \nu)\left(\bigcup_{n=1}^{\infty} E_n\right),\end{aligned}$$

where we used MCT to get from the first line to the second line (since $(\nu((E_n)_{(-)}))$ is an increasing sequence of nonnegative measurable functions which converges pointwise to $\nu\left((\bigcup_{n=1}^{\infty} E_n)_{(-)}\right)$). Thus we have shown $\mathcal{C}$ is a σ-algebra which contains $\mathcal{R}$. □

32.3 Tonelli's Theorem and Fubini's Theorem

Theorem 32.1. *The following statements hold.*

1. *(Tonelli) Let $f\colon X \times Y \to [0, \infty]$ be a nonnegative $\mathcal{M} \otimes \mathcal{N}$-measurable function. Define the function $\int_Y f_{(-)} \mathrm{d}\nu \colon X \to [0, \infty]$ by*

$$\left(\int_Y f_{(-)} \mathrm{d}\nu\right)(x) = \int_Y f_x \mathrm{d}\nu$$

for all $x \in X$. Then $\int_Y f_{(-)} \mathrm{d}\nu$ is $\mathcal{M}$-measurable. Furthermore,

$$\int_X \left(\int_Y f_{(-)} \mathrm{d}\nu\right) \mathrm{d}\mu = \int_{X\times Y} f \mathrm{d}(\mu \otimes \nu)$$

Similarly, define the function $\int_X f^{(-)} \mathrm{d}\mu \colon Y \to [0, \infty]$ by

$$\left(\int_X f^{(-)} \mathrm{d}\mu\right)(y) = \int_X f^y \mathrm{d}\mu$$

for all $y \in Y$. Then $\int_X f^{(-)} \mathrm{d}\mu$ is $\mathcal{N}$-measurable. Furthermore,

$$\int_{X\times Y} f \mathrm{d}(\mu \otimes \nu) = \int_Y \left(\int_X f^{(-)} \mathrm{d}\mu\right) \mathrm{d}\nu.$$

2. *(Fubini) Let $f\colon X \times Y \to \mathbb{R}$ be a $\mathcal{M} \otimes \mathcal{N}$-integrable function. Then the function $f_{(-)}$ is $\mathcal{N}$-integrable. Furthemore,*

$$\int_X \left(\int_Y f_{(-)} \mathrm{d}\nu\right) \mathrm{d}\mu = \int_{X\times Y} f \mathrm{d}(\mu \otimes \nu).$$

Similarly, for almost every $y \in Y$, the function $f^{(-)}$ is $\mathcal{M}$-integrable. Furthermore,

$$\int_{X\times Y} f \mathrm{d}(\mu \otimes \nu) = \int_Y \left(\int_X f^{(-)} \mathrm{d}\mu\right) \mathrm{d}\nu$$

Proof. 1. (Tonelli) It suffices to show the first part of this statement since the proof of the second part is completely symmetrical to the proof the first part. First we prove this for nonnegative simple functions. Let $\varphi\colon X \times Y \to$

$[0,\infty]$ be a nonnegative simple function, and suppose its canonical form is given by $\varphi = \sum_{i=1}^n c_i 1_{E_i}$. Then note that

$$\begin{aligned}
\int_Y \varphi_{(-)} \mathrm{d}\nu &= \int_Y \left(\sum_{i=1}^n c_i 1_{E_i} \right)_{(-)} \mathrm{d}\nu \\
&= \int_Y \sum_{i=1}^n c_i (1_{E_i})_{(-)} \mathrm{d}\nu \\
&= \int_Y \sum_{i=1}^n c_i 1_{(E_i)_{(-)}} \mathrm{d}\nu \\
&= \sum_{i=1}^n c_i \left(\int_Y 1_{(E_i)_{(-)}} \mathrm{d}\nu \right) \\
&= \sum_{i=1}^n c_i \nu((E_i)_{(-)}).
\end{aligned}$$

Thus $\int_Y \varphi_{(-)} \mathrm{d}\nu$ is a sum of $\mathcal{M}$-measurable functions, and is hence $\mathcal{M}$-measurable. Furthermore, we have

$$\begin{aligned}
\int_X \left(\int_Y \varphi_{(-)} \mathrm{d}\nu \right) \mathrm{d}\mu &= \int_X \sum_{i=1}^n c_i \nu((E_i)_{(-)}) \mathrm{d}\mu \\
&= \sum_{i=1}^n c_i \left(\int_X \nu((E_i)_{(-)}) \mathrm{d}\mu \right) \\
&= \sum_{i=1}^n c_i (\mu \otimes \nu)(E_i) \\
&= \int_{X \times Y} \varphi \mathrm{d}(\mu \otimes \nu).
\end{aligned}$$

Now we prove it for more generally, let $f \colon X \times Y \to [0,\infty]$ be a nonnegative $\mathcal{M} \otimes \mathcal{N}$-measurable function. Choose an increasing sequence $(\varphi_n \colon X \times Y \to [0,\infty])$ of nonngetative simple functions such that (φ_n) converges to f pointwise. Then note that for each $x \in X$, we have

$$\begin{aligned}
\int_Y f_x \mathrm{d}\nu &= \int_Y \lim_{n\to\infty} ((\varphi_n)_x) \mathrm{d}\nu \\
&= \lim_{n\to\infty} \int_Y ((\varphi_n)_x) \mathrm{d}\nu
\end{aligned}$$

where we used MCT to get from the first line to the second line. Since $x \in X$ was arbitary, it follows that the sequence $(\int_Y \varphi_{(-)} \mathrm{d}\nu)$ of nonnegative $\mathcal{M}$-measurable converges pointwise to the function $\int_Y f_{(-)} \mathrm{d}\nu$. Hence $\int_Y f_{(-)} \mathrm{d}\nu$ is a nonnegative $\mathcal{M}$-measurable function. Furthermore, we have

$$\begin{aligned}
\int_X \left(\int_Y f_{(-)} \mathrm{d}\nu \right) \mathrm{d}\mu &= \int_X \left(\lim_{n\to\infty} \int_Y (\varphi_n)_{(-)} \mathrm{d}\nu \right) \mathrm{d}\mu \\
&= \lim_{n\to\infty} \left(\int_X \left(\int_Y (\varphi_n)_{(-)} \mathrm{d}\nu \right) \mathrm{d}\mu \right) \\
&= \lim_{n\to\infty} \int_{X \times Y} \varphi_n \mathrm{d}(\mu \otimes \nu) \\
&= \int_{X \times Y} \lim_{n\to\infty} \varphi_n \mathrm{d}(\mu \otimes \nu) \\
&= \int_{X \times Y} f \mathrm{d}(\mu \otimes \nu),
\end{aligned}$$

where used MCT twice.

2. (Fubini) It suffices to show the first part of this statement since the proof of the second part is completely symmetrical to the proof the first part. To prove Fubini, we write $f = f^+ - f^-$ and apply Tonelli to both f^+ and

f^-. We have

$$\begin{aligned}\int_{X\times Y} f\mathrm{d}(\mu\otimes\nu) &= \int_{X\times Y} f^+\mathrm{d}(\mu\otimes\nu) - \int_{X\times Y} f^-\mathrm{d}(\mu\otimes\nu).\\ &= \int_X \left(\int_Y (f^+)_{(-)}\mathrm{d}\nu\right)\mathrm{d}\mu - \int_X\left(\int_Y (f^-)_{(-)}\mathrm{d}\nu\right)\mathrm{d}\mu\\ &= \int_X\left(\int_Y (f^+)_{(-)}\mathrm{d}\nu - \int_Y (f^-)_{(-)}\mathrm{d}\nu\right)\mathrm{d}\mu\\ &= \int_X\left(\int_Y f_{(-)}\mathrm{d}\nu\right)\mathrm{d}\mu.\end{aligned}$$

We need to justify some of the steps that we used above. The first line is simply the definition of $\int_{X\times Y} f\mathrm{d}(\mu\otimes\nu)$. We obtained the second line from the first line from Tonelli. To get the third line from the second line, we needed to use the fact that $\int_Y f_{(-)}\mathrm{d}\nu$ is $\mathcal{M}$-integrable. To see why $\int_Y f_{(-)}\mathrm{d}\nu$ is $\mathcal{M}$-integrable, it suffices to show that $(\int_Y f_{(-)}\mathrm{d}\nu)^+$ (the positive part of $\int_Y f_{(-)}\mathrm{d}\nu$) and $(\int_Y f_{(-)}\mathrm{d}\nu)^-$ (the negative part of $\int_Y f_{(-)}\mathrm{d}\nu$) are $\mathcal{M}$-integrable. To see why $(\int_Y f_{(-)}\mathrm{d}\nu)^+$ is $\mathcal{M}$-integrable, observed that

$$\begin{aligned}\int_X\left(\int_Y f_{(-)}\mathrm{d}\nu\right)^+\mathrm{d}\mu &\leq \int_X\left(\int_Y (f_{(-)})^+\mathrm{d}\nu\right)\mathrm{d}\mu\\ &= \int_X\left(\int_Y (f^+)_{(-)}\mathrm{d}\nu\right)\mathrm{d}\mu\\ &= \int_{X\times Y} f^+\mathrm{d}(\mu\otimes\nu)\\ &< \infty.\end{aligned}$$

A similar proof shows that $(\int_Y f_{(-)}\mathrm{d}\nu)^-$ is $\mathcal{M}$-integrable. Thus $\int_Y f_{(-)}\mathrm{d}\nu$ is $\mathcal{M}$-integrable. The last step that we need to justify, is how we obtained fourth line from the third line. To see why this holds, observe that for each $x\in X$, we have

$$\begin{aligned}\int_Y (f^+)_x\mathrm{d}\nu - \int_Y (f^-)_x\mathrm{d}\nu &= \int_Y (f_x)^+\mathrm{d}\nu - \int_Y (f_x)^-\mathrm{d}\nu\\ &= \int_Y f_x\mathrm{d}\nu,\end{aligned}$$

where we applied the definition of $\int_Y f_x\mathrm{d}\nu$ to obtain the second from the first line. Note that this makes sense because f_x is ν-integrable almost everywhere. Indeed, since

$$\int_X\left(\int_Y (f_{(-)})^+\mathrm{d}\nu\right)\mathrm{d}\mu < \infty,$$

it follows that $\int_Y (f_x)^+\mathrm{d}\nu < \infty$ for almost all $x\in X$. Similarly, $\int_Y (f_x)^-\mathrm{d}\nu < \infty$ for almost all $y\in Y$. Thus $\int_Y f_{(-)}\mathrm{d}\nu$ is $\mathcal{N}$-integrable almost everywhere. □

Example 32.1. Assume you need to compute

$$\int_X\left(\int_Y f(x,y)\mathrm{d}\nu(y)\right)\mathrm{d}\mu(x).$$

Assume you know how to compute

$$\int_Y\left(\int_X f(x,y)\mathrm{d}\mu(x)\right)\mathrm{d}\nu(y).$$

Assume f is *not* nonnegative. To apply Fubini, we need $f\in L^1(X\times Y)$. Since $|f|$ is a nonnegative measurable function, we can apply Tonelli to $|f|$ to get

$$\int_{X\times Y}|f|\mathrm{d}(\mu\otimes\nu) = \int_Y\left(\int_X |f(x,y)|\mathrm{d}\mu(x)\right)\mathrm{d}\nu(y).$$

If we can compute the last itterated integral and show it is $<\infty$, then $f\in L^1(X\times Y)$. So we can use Fubini.

33 Signed Measures

We'll discuss only *finite* signed measures.

Definition 33.1. Let $(X, \mathcal{M})$ be a measurable space. A function $\nu \colon \mathcal{M} \to \mathbb{R}$ is said to be a **finite signed measure** if $\nu(\emptyset) = 0$ and if ν is countably additive: if (E_n) is sequence of pairwise disjoint sets in $\mathcal{M}$, then

$$\nu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \nu(E_n)$$

where the series converges absolutely.

Example 33.1. If μ_1 and μ_2 are usual (positive) finite measures on $\mathcal{M}$, then $\nu = \mu_1 - \mu_2$ is a signed measure.

Example 33.2. If μ is a measure on $(X, \mathcal{M})$ and $f \in L^1(X, \mathcal{M}, \mu)$, then

$$\nu(E) = \int_X 1_E f \mathrm{d}\mu$$

is a signed measure.

Proposition 33.1. *Let ν be a finite signed measure on $(X, \mathcal{M})$ and let (E_n) be a sequence in $\mathcal{M}$, Then*

1. *if (E_n) is an ascending sequence, then $\nu\left(\bigcup_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \nu(E_n)$.*
2. *if (E_n) is a descending sequence, then $\nu\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \nu(E_n)$.*

Proof. Same proof as for usual measures. □

Definition 33.2. Let $E \in \mathcal{M}$. We say

1. E is a ν**-positive set** if $\nu(F) \geq 0$ for all $F \subseteq E$ with $F \in \mathcal{M}$.
2. E is ν**-negative set** if $\nu(F) \leq 0$ for all $F \subseteq E$ with $F \in \mathcal{M}$.
3. E is a ν**-null set** if $\nu(F) = 0$ for all $F \subseteq E$ with $F \in \mathcal{M}$.

Example 33.3. Let $a, b \in X$. We define

$$\delta_a(E) = \begin{cases} 0 & \text{if } a \notin E \\ 1 & \text{if } a \in E \end{cases}$$

We also define $\nu = \delta_a - \delta_b$. Then ν is signed measure. If $a \in E$ and $b \notin E$, then $\nu(E) = 1$. If $a \notin E$ and $b \in E$, then $\nu(E) = -1$. If $a, b \in E$, then $\nu(E) = 0$.

Lemma 33.1. *If $(P_n) \subseteq \mathcal{M}$ is a sequence of ν-positive sets, then $\bigcup_{n=1}^{\infty} P_n$ is also a ν-positive set.*

Proof. By disjointifying if necessary, we may assume that the sequence (P_n) is pairwise disjoint. Let $E \in \mathcal{M}$ such that $E \subseteq \bigcup_{n=1}^{\infty} P_n$. Then

$$\begin{aligned}
\nu(E) &= \nu\left(E \cap \left(\bigcup_{n=1}^{\infty} P_n\right)\right) \\
&= \nu\left(\bigcup_{n=1}^{\infty} (E \cap P_n)\right) \\
&= \sum_{n=1}^{\infty} \nu(E \cap P_n) \\
&\geq 0.
\end{aligned}$$

□

33.0.1 Hahn Decomposition Theorem

Theorem 33.2. *(Hahn decomposition theorem) If ν is a signed measure on $(X, \mathcal{M})$, then there exists $P \in \mathcal{M}$ ν-positive and $Q \in \mathcal{M}$ ν-negative such that $X = P \cup Q$ and $P \cap Q = \emptyset$. If P', Q' is another such pair, then $P \Delta P'$ and $Q \Delta Q'$ are both ν-null sets.*

Proof. Let $m = \sup\{\nu(E) \mid E \text{ is } \nu\text{-positive}\}$. Then for all $n \in \mathbb{N}$, there exists a ν-positive set P_n such that

$$m \geq \nu(P_n) > m - \frac{1}{n}.$$

Then $\lim_{n\to\infty} \nu(P_n) = m$. Take $P = \bigcup_{n=1}^{\infty} P_n$. We showed last time P is also ν-positive and

$$\begin{aligned} \nu(P) &= \nu\left(\bigcup_{n=1}^{\infty} P_n\right) \\ &= \lim_{n\to\infty} \nu(P_n) \\ &= m. \end{aligned}$$

Set $N = P^c$. We need to show N is ν-negative. First, we make two observations. First, if $E \subseteq N$ is a ν-positive set, then E must be a ν-null set. Indeed, assume $\nu(E) > 0$. Then

$$\begin{aligned} \nu(E \cup P) &= \nu(E) + \nu(P) \\ &= \nu(P) \\ &= m \end{aligned}$$

and $E \cup P$ is ν-positive, but this is a contradiction. This establishes our first claim. The next claim we make is that if $A \subseteq N$ and $\nu(A) > 0$, then there exists $B \subseteq A$ with $\nu(B) > \nu(A)$. Indeed, we just proved that A cannot be ν-positive. Therefore there exists $C \subseteq A$ such that $\nu(C) < 0$. Take $B = A \backslash C$. Then

$$\begin{aligned} \nu(B) &= \nu(A) - \nu(C) \\ &> \nu(A). \end{aligned}$$

This establishes our second claim.

Now we construct two sequence $(n_k) \subseteq \mathbb{N}$ and (A_k) of subsets of N in the following way: Set

$$n_1 = \min\{k \in \mathbb{N} \mid \exists B \subseteq N \text{ with } \nu(B) > 1/k\}.$$

Next we define A_1 to be one such subset B, so $A_1 \subseteq N$ such that $\nu(A_1) > 1/n_1$. Next, we define

$$n_2 = \min\{k \in \mathbb{N} \mid \exists B \subseteq A_1 \text{ with } \nu(B) > \nu(A_1) + 1/k\}.$$

Next we define A_2 to be one such subset B, so $A_2 \subseteq A_1$ such that $\nu(A_2) > \nu(A_1) + 1/n_2$. Continuing in this way, we obtain (n_k) and (A_k). Let $A = \bigcap_{n=1}^{\infty} A_n$. Then

$$\begin{aligned} \nu(A) &= \lim_{k\to\infty} \nu(A_k) \\ &\geq \sum_{k=1}^{\infty} \frac{1}{n_k}. \end{aligned}$$

Since $\nu(A)$ is finite, we see that this series must be convergent. Thus $1/n^k \to 0$ as $k \to \infty$. Then there exists $B \subseteq A$ with the property that $\nu(B) > \nu(A)$. Let $N \in \mathbb{N}$ such that $1/N < \nu(B) - \nu(A)$. There exists j such that

$$\frac{1}{n_j} < \frac{1}{N} < \nu(B) - \nu(A).$$

That is, $\nu(B) > \nu(A) + 1/n_j$. Also

$$\begin{aligned} B &\subseteq A \\ &= \bigcap_{k=1}^{\infty} A_k \\ &\subseteq A_j. \end{aligned}$$

So our assumption that N is not negative is wrong. Thus N is positive. The uniqueness part is simpler and left as an exercise. $\square$

33.0.2 Mutually Singular Signed Measures

Definition 33.3. The signed measures ν_1, ν_2 on $(X, \mathcal{M})$ are said to be **mutually singular**, denoted $\nu_1 \perp \nu_2$, if there exists $E, F \in \mathcal{M}$ such that $E \cap F = \emptyset$, $E \cup F = X$, E is a ν_1-null set and F is a ν_2-null set.

33.0.3 Jordan Decomposition Theorem

Theorem 33.3. *(Jordan decomposotion theorem) If ν is a signed measure on $(X, \mathcal{M})$, then there exists a unique pair of measures ν^+ and ν^- such that $\nu = \nu^+ - \nu^-$ and $\nu^+ \perp \nu^-$.*

Proof. Let $X = P \cup N$ be a Hahn decomposition of ν. Define

$$\nu^+(E) = \nu(E \cap P) \quad \text{and} \quad \nu^-(E) = -\nu(E \cap N)$$

for all $E \in \mathcal{M}$. It is straightforward to check that ν^+ and ν^- are both measures on $(X, \mathcal{M})$. Also,

$$\begin{aligned} \nu(E) &= \nu(E \cap P) + \nu(E \cap N) \\ &= \nu^+(E) - \nu^-(E). \end{aligned}$$

Thus $\nu = \nu^+ - \nu^-$. Also P is a ν^--null set. Indeed, if $F \subseteq P$, then

$$\begin{aligned} \nu^-(F) &= -\nu(F \cap N) \\ &= -\nu(\emptyset) \\ &= 0. \end{aligned}$$

Similarly, N is a ν^+-null set. It only remains to show uniqueness of such a pair. So suppose $\nu = \mu^+ - \mu^-$ for another such pair (μ^+, μ^-) with $\mu^+ \perp \mu^-$. Since $\mu^+ \perp \mu^-$, there exists $E, F \in X$ such that $X = E \cup F$, $E \cap F = \emptyset$, E is a μ^--null set, and F is a μ^+-null set. Then $X = E \cup F$ is another Hahn decomposition (since E is a μ-positive set and F is a ν-negative set). Therefore $P \Delta E$ and $Q \Delta F$ are both ν-null sets. Therefore if $A \in \mathcal{M}$, we have

$$\begin{aligned} \mu^+(A) &= \mu^+(A \cap E) + \mu^+(A \cap F) \\ &= \nu(A \cap E) + 0 \\ &= \nu(A \cap P) \\ &= \nu^+(A), \end{aligned}$$

so $\mu^+ = \nu^+$. Similarly, $\mu^- = \nu^-$. □

33.1 Banach Space of Signed Measures

Let $(X, \mathcal{M})$ be a measurable space. We denote by $M(X, \mathcal{M})$ to be the set of all finite signed measures on $\mathcal{M}$. First let us give $M(X, \mathcal{M})$ the structure of an $\mathbb{R}$-vector space as follows: Let $a \in \mathbb{R}$ and let $\mu, \nu \in M(X, \mathcal{M})$. We define addition $\mu + \nu \colon \mathcal{M} \to \mathbb{R}$ and scalar multiplication $a\mu \colon \mathcal{M} \to \mathbb{R}$ by

$$\begin{aligned} (\mu + \nu)(E) &= \mu(E) + \nu(E) \\ (a\mu)(E) &= a\mu(E) \end{aligned}$$

for all $E \in \mathcal{M}$. It is straightforward to check that addition and scalar multiplication defined in this way gives $M(X, \mathcal{M})$ the structure of an $\mathbb{R}$-vector space.

Next let us give $M(X, \mathcal{M})$ the structure of a normed linear space as follows: Let $\mu \in M(X, \mathcal{M})$ and let $\mu = \mu^+ - \mu^-$ be the Jordan decomposition of μ. We define the norm $\| \cdot \| \colon M(X, \mathcal{M}) \to \mathbb{R}_{\geq 0}$ by

$$\|\mu\| = \mu^+(X) + \mu^-(X).$$

Again it is staightforward to check that this gives $M(X, \mathcal{M})$ the structure of a normed linear space.

33.2 Absolute Continuity For Signed Measures

Definition 33.4. Let ν be a signed measure on $(X, \mathcal{M})$ and let μ be a usual measure on $(X, \mathcal{M})$. We say ν is **absolutely continuous** with respect to μ if for all $E \in \mathcal{M}$, we have $\mu(E) = 0$ implies $\nu(E) = 0$. We denote this by $\nu \ll \mu$.

It's esay to prove that

$$\begin{aligned} \nu \ll \mu &\iff \nu^+ \ll \mu \text{ and } \nu^- \ll \mu \\ &\iff |\nu| \ll \mu. \end{aligned}$$

Proposition 33.2. *$\nu \ll \mu$ if and only if for all $\varepsilon > 0$ there exists $\delta > 0$ such that for all $E \in \mathcal{M}$ with $\mu(E) < \delta$, we have $\nu(E) < \delta$.*

Proof. ($\Longleftarrow$) is easy. ($\Longrightarrow$) assume $\nu \ll \mu$ but doesn't hold. Then there exists $\varepsilon > 0$ fo rall $n \in \mathbb{N}$ ther exists $E_n \in \mathcal{M}$ with $\mu(E_n) < 1/2^n$ buth $|\nu|(E_n) \geq \varepsilon$. On the other hand, since $|\nu|(E_n) \geq \varepsilon$, we have $|\nu|(F) \geq 0$ which contradicts

$$|\nu| \ll \mu \iff \nu \ll \mu.$$

□

33.2.1 Radon-Nikodym

Theorem 33.4. *(Radon-Nikodym) $\nu \ll \mu$ implies there exists $h \in L^1(\mu)$ such that*

$$\nu(E) = \int_X 1_E h \mathrm{d}\mu.$$

Proof. Assume first that ν is the usual measure. Consider the Hilbert space $L^2(\mu + \nu)$. The map $\ell \colon L^2(\mu+\nu) \to \mathbb{R}$ defined by

$$\ell(f) = \int_X f \mathrm{d}\mu$$

for all $f \in L^2(\mu + \nu)$ is a linear functional and

$$\begin{aligned} |\ell(f)| &= \left| \int_X f \mathrm{d}\mu \right| \\ &\leq \int_X |f| \mathrm{d}\mu \\ &\leq \sqrt{\int_X |f|^2 \mathrm{d}(\mu+\nu)} \sqrt{\int_X 1^2 \mathrm{d}(\mu+\nu)} \\ &= \|f\|_2 (\mu(X) + \nu(X)). \end{aligned}$$

So by the Riesz representation theorem, there exists $g \in L^2(\mu + \nu)$ such that

$$\ell(f) = \langle f, g \rangle = \int_X fg \mathrm{d}(\mu + \nu).$$

So

$$\int_X f \mathrm{d}\mu = \int_X fg \mathrm{d}(\mu+\nu)$$

which is equivalent to

$$\int_X f(1-g) \mathrm{d}\mu = \int_X fg \mathrm{d}\nu.$$

We claim that $0 < g \leq 1$ for μ almost everywhere. Define $F \in \mathcal{M}$ by

$$F = \{x \in X \mid g(x) \leq 0\}.$$

Take $f = 1_F$. Then

$$\begin{aligned} \int_X 1_F (1-g) \mathrm{d}\mu &= \int 1_F g \mathrm{d}\nu \\ &\leq 0 \end{aligned}$$

which implies $\mu(0) \leq 0$ which implies $\mu(F) = 0$ which implies $0 < g$ for μ a.e. Similarly, by considering the set

$$G = \{x \in X \mid g(x) > 1\},$$

we can get $\mu(G) = 0$ which implies $g \leq 1$ for μ a.e.

Now we can define $h = (1-g)/g$. For $E \in \mathcal{M}$, pick $f = 1_E \cdot (1/g)$ plug in

$$\int_X 1_E \frac{1}{g}(1-g)\mathrm{d}\mu = \int 1_E \mathrm{d}\nu$$

this implies

$$\int 1_E h \mathrm{d}\mu = \int 1_E \mathrm{d}\nu = \nu(E)$$

Thus

$$\nu(E) = \int_X 1_E h \mathrm{d}\mu.$$

For general signed measures ν, we write $\nu = \nu^+ - \nu^-$ and we use $\nu \ll \mu$ to imply $\nu^+ \ll \mu$ and $\nu^- \ll \mu$ to imlpies

$$\nu^+(E) = \int 1_E h_1 \mathrm{d}\mu \quad \text{and } \nu^-(E) = \int 1_E h_2 \mathrm{d}\mu$$

which implies

$$\nu(E) = \nu^+(E) - \nu^-(E) = \int 1_E(h_1 - h_2)\mathrm{d}\mu.$$

So we are done. □

Homework Problems

34 Homework 1

Throughout this homework, let X be a set and let $\mathcal{P}(X)$ denote the set of all subsets of X.

34.1 Charactertistic Function Identities

Proposition 34.1. *Let $A, B \in \mathcal{P}(X)$. Then*

1. $1_A = 1_B$ *if and only if* $A = B$;
2. $1_{A \cap B} = 1_A 1_B$;
3. $1_{A \cup B} = 1_A + 1_B - 1_A 1_B$;
4. $1_{A^c} = 1 - 1_A$;
5. $1_{A \setminus B} = 1_A - 1_B$ *if and only if* $B \subseteq A$;
6. $1_A + 1_B \equiv 1_{A \Delta B} \mod 2$.

Proof.
1. Suppose $1_A = 1_B$ and let $x \in A$. Then

$$\begin{aligned} 1 &= 1_A(x) \\ &= 1_B(x) \end{aligned}$$

implies $x \in B$. Thus $A \subseteq B$. Similarly, if $x \in B$, then

$$\begin{aligned} 1 &= 1_B(x) \\ &= 1_A(x) \end{aligned}$$

implies $x \in A$. Thus $B \subseteq A$.

Conversely, suppose $A = B$ and let $x \in X$. If $x \in A$, then $x \in B$, hence

$$\begin{aligned} 1_A(x) &= 1 \\ &= 1_B(x). \end{aligned}$$

If $x \notin A$, then $x \notin B$, hence

$$\begin{aligned} 1_A(x) &= 0 \\ &= 1_B(x). \end{aligned}$$

Therefore the indicator functions 1_A and 1_B agree on all of X, and hence must be equal to each other.

2. Let $x \in X$. If $x \in A \cap B$, then $x \in A$ and $x \in B$, and thus we have

$$\begin{aligned} 1_{A\cap B}(x) &= 1 \\ &= 1 \cdot 1 \\ &= 1_A(x)1_B(x). \end{aligned}$$

If $x \notin A \cap B$, then either $x \notin A$ or $x \notin B$. Without loss of generality, say $x \notin A$. Then we have

$$\begin{aligned} 1_{A\cap B}(x) &= 0 \\ &= 0 \cdot 1_B(x) \\ &= 1_A(x)1_B(x). \end{aligned}$$

Therefore the functions $1_{A\cap B}$ and $1_A 1_B$ agree on all of X, and hence must be equal to each other.

3. Let $x \in X$. If $x \in A \cup B$, then either $x \in A$ or $x \in B$. Without loss of generality, say $x \in A$. Then we have

$$\begin{aligned} 1_{A\cup B}(x) &= 1 \\ &= 1 + 1_B(x) - 1_B(x) \\ &= 1 + 1_B(x) - 1 \cdot 1_B(x) \\ &= 1_A(x) + 1_B(x) - 1_A(x)1_B(x). \end{aligned}$$

If $x \notin A \cup B$, then $x \notin A$ and $x \notin B$. Therefore we have

$$\begin{aligned} 1_{A\cup B}(x) &= 0 \\ &= 0 + 0 - 0 \cdot 0 \\ &= 1_A(x) + 1_B(x) - 1_A(x)1_B(x). \end{aligned}$$

Thus the functions $1_{A\cup B}$ and $1_A + 1_B - 1_A 1_B$ agree on all of X, and hence must be equal to each other.

4. Let $x \in X$. If $x \in A$, then $x \notin A^c$, hence

$$\begin{aligned} 1_{A^c}(x) &= 0 \\ &= 1 - 1 \\ &= 1 - 1_A(x). \end{aligned}$$

If $x \notin A$, then $x \in A^c$, hence

$$\begin{aligned} 1_{A^c}(x) &= 1 \\ &= 1 - 0 \\ &= 1 - 1_A(x). \end{aligned}$$

Therefore the functions 1_{A^c} and $1 - 1_A$ agree on all of X, and hence must be equal to each other.

5. Suppose $B \subseteq A$ and let $x \in X$. If $x \in A \backslash B$, then $x \in A$ and $x \notin B$, hence

$$\begin{aligned}
1_{A\backslash B}(x) &= 1 \\
&= 1 - 0 \\
&= 1_A(x) - 1_B(x).
\end{aligned}$$

If $x \in B$, then $x \in A$ but $x \notin A \backslash B$, hence

$$\begin{aligned}
1_{A\backslash B}(x) &= 0 \\
&= 1 - 1 \\
&= 1_A(x) - 1_B(x).
\end{aligned}$$

If $x \notin A$, then $x \notin B$ and $x \notin A \backslash B$, hence

$$\begin{aligned}
1_{A\backslash B}(x) &= 0 \\
&= 0 - 0 \\
&= 1_A(x) - 1_B(x).
\end{aligned}$$

Therefore the functions $1_{A\backslash B}$ and $1_A - 1_B$ agree on all of X, and hence must be equal to each other.
For converse direction, we prove the contrapositive statement. Suppose $B \not\subseteq A$. Choose $b \in B$ such that $b \notin A$. Then

$$\begin{aligned}
1_{A\backslash B}(b) &= 0 \\
&\neq -1 \\
&= 0 - 1 \\
&= 1_A(b) - 1_B(b).
\end{aligned}$$

Therefore $1_{A\backslash B} \neq 1_A - 1_B$.

5. Let $x \in X$. If $x \in A$, then $x \notin A^c$, hence

$$\begin{aligned}
1_{A^c}(x) &= 0 \\
&= 1 - 1 \\
&= 1 - 1_A(x).
\end{aligned}$$

If $x \notin A$, then $x \in A^c$, hence

$$\begin{aligned}
1_{A^c}(x) &= 1 \\
&= 1 - 0 \\
&= 1 - 1_A(x).
\end{aligned}$$

Therefore the functions 1_{A^c} and $1 - 1_A$ agree on all of X, and hence must be equal to each other.

6. We have

$$\begin{aligned}
1_{A\Delta B} &= 1_{(A\backslash B)\cup(B\backslash A)} \\
&= 1_{A\backslash B} + 1_{B\backslash A} - 1_{A\backslash B}1_{B\backslash A} \\
&= 1_{A\backslash A\cap B} + 1_{B\backslash A\cap B} - 1_{(A\backslash B)\cap(B\backslash A)} \\
&= 1_A - 1_{A\cap B} + 1_B - 1_{A\cap B} - 1_{\varnothing} \\
&= 1_A + 1_B - 2 \cdot 1_{A\cap B} \\
&\equiv 1_A + 1_B \bmod 2.
\end{aligned}$$

$\square$

34.2 Cauchy Sequence in $(C[a,b], \|\cdot\|_1)$ Converging Pointwise and in L^1 to Indicator Function

Proposition 34.2. *Let I be a subinterval of $[a,b]$. Then there exists a Cauchy sequence (f_n) in $(C[a,b], \|\cdot\|_1)$ such that (f_n) converges pointwise to 1_I on $[a,b]$ and moreover*

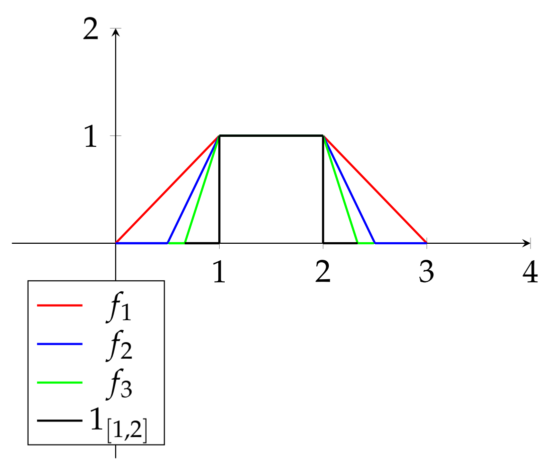
$$\lim_{n\to\infty} \|f_n\|_1 = \text{length}(I).$$

Proof. If $I = \emptyset$, then we take $f_n = 0$ for all $n \in \mathbb{N}$. Thus assume I is a nonempty subinterval of $[a,b]$. We consider two cases; namely $I = (c,d)$ and $I = [c,d]$. The other cases ($I = (c,d]$ and $I = [c,d)$) will easily be seen to be a mixture of these two cases.

Case 1: Suppose $I = [c,d]$. For each $n \in \mathbb{N}$, define

$$f_n(x) = \begin{cases} 0 & \text{if } a \le x < c - \left(\frac{c-a}{n}\right) \\ \frac{n}{c-a}(x-c)+1 & \text{if } c - \left(\frac{c-a}{n}\right) \le x < c \\ 1 & \text{if } c \le x \le d \\ \frac{n}{d-b}(x-d)+1 & \text{if } d < x \le d + \left(\frac{b-d}{n}\right) \\ 0 & \text{if } d + \left(\frac{b-d}{n}\right) < x \le b \end{cases}$$

The image below gives the graphs for f_1, f_2, and f_3 in the case where $[a,b] = [0,3]$ and $[c,d] = [1,2]$.



For each $n \in \mathbb{N}$, the function f_n is continuous since each of its segments is continuous and are equal on their boundaries.

Let us check that (f_n) converges pointwise to 1_I: If $x \in [a,c)$, then we choose $N \in \mathbb{N}$ such that

$$x \le c - \left(\frac{c-a}{N}\right).$$

Then $f_n(x) = 0$ for all $n \ge N$. Thus

$$\lim_{n\to\infty} f_n(x) = 0 = 1_I(x).$$

Similarly, if $x \in (d,b]$, then we choose $N \in \mathbb{N}$ such that

$$x \ge d + \left(\frac{b-d}{N}\right).$$

Then $f_n(x) = 0$ for all $n \ge N$. Thus

$$\lim_{n\to\infty} f_n(x) = 0 = 1_I(x).$$

Finally, if $x \in [c,d]$, then $f_n(x) = 1$ for all $n \in \mathbb{N}$ by definition and thus

$$\lim_{n\to\infty} f_n(x) = 0 = 1_I(x).$$

Let us check that (f_n) is Cauchy in $(C[a,b], \|\cdot\|_1)$. Let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that

$$\frac{c-a+b-d}{n} < \varepsilon$$

for all $n \geq N$. Then $n \geq m \geq N$ implies

$$
\begin{aligned}
\|f_n - f_m\|_1 &= \int_a^b |f_n(x) - f_m(x)| \mathrm{d}x \\
&= \int_a^b (f_n(x) - f_m(x)) \mathrm{d}x \\
&= \int_{c-\left(\frac{c-a}{m}\right)}^{c} (f_n(x) - f_m(x)) \mathrm{d}x + \int_d^{d+\left(\frac{b-d}{m}\right)} (f_n(x) - f_m(x)) \mathrm{d}x \\
&\leq \int_{c-\left(\frac{c-a}{m}\right)}^{c} \mathrm{d}x + \int_d^{d+\left(\frac{b-d}{m}\right)} \mathrm{d}x \\
&= \frac{c-a}{m} + \frac{b-d}{m} \\
&= \frac{c-a+b-d}{m} \\
&< \varepsilon.
\end{aligned}
$$

Thus the sequence (f_n) is Cauchy in $(C[a,b], \|\cdot\|_1)$.

Finally, we check that $\|f_n\|_1 \to \text{length}(I)$ as $n \to \infty$. We have

$$
\begin{aligned}
d - c &\leq \|f_n\|_1 \\
&= \int_a^b |f_n(x)| \mathrm{d}x \\
&= \int_a^b f_n(x) \mathrm{d}x \\
&= \int_{c-\left(\frac{c-a}{n}\right)}^{c} f_n(x) \mathrm{d}x + \int_c^d \mathrm{d}x + \int_d^{d+\left(\frac{b-d}{n}\right)} f_n(x) \mathrm{d}x \\
&\leq \int_{c-\left(\frac{c-a}{n}\right)}^{c} \mathrm{d}x + \int_c^d \mathrm{d}x + \int_d^{d+\left(\frac{b-d}{n}\right)} \mathrm{d}x \\
&= \frac{c-a}{n} + d - c + \frac{b-d}{n} \\
&\to d - c.
\end{aligned}
$$

Thus for each $n \in \mathbb{N}$, we have

$$
d - c \leq \|f_n\|_1 \leq d - c + \frac{c-a+b-d}{n}. \tag{86}
$$

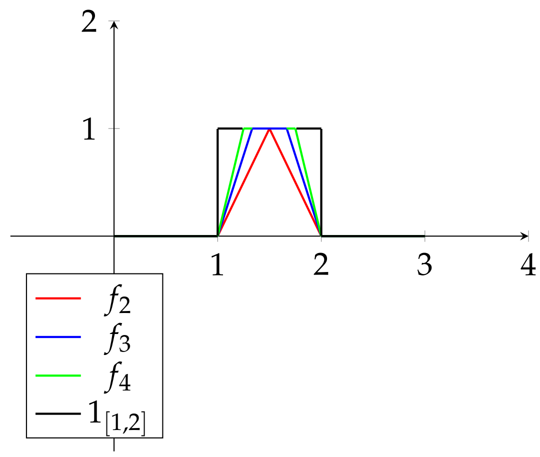
By taking $n \to \infty$ in (86), we see that

$$
\begin{aligned}
\lim_{n\to\infty} \|f_n\|_1 &= d - c \\
&= \text{length}(I).
\end{aligned}
$$

Case 2: Suppose $I = (c,d)$. For each $n \geq 2$, define

$$
f_n(x) = \begin{cases}
0 & \text{if } a \leq x \leq c \\
\frac{n}{d-c}(x-c) & \text{if } c < x \leq c + \left(\frac{d-c}{n}\right) \\
1 & \text{if } c + \left(\frac{d-c}{n}\right) \leq x \leq d - \left(\frac{d-c}{n}\right) \\
\frac{n}{c-d}(x-d) & \text{if } d - \left(\frac{d-c}{n}\right) \leq x \leq d \\
0 & \text{if } d \leq x \leq b
\end{cases}
$$

The image below gives the graphs for f_2, f_3, and f_4 in the case where $[a,b] = [0,3]$ and $(c,d) = (1,2)$.



That (f_n) is a Cauchy sequence of continuous funtions in $(C[a,b], \|\cdot\|_1)$ which converges pointwise to 1_I and $\|f_n\|_1 \to \text{length}(I)$ as $n \to \infty$ follows from similar arguments used in case 1. $\square$

34.3 Algebra of Subsets of X Closed under Symmetric Differenences and Relative Compliments

Proposition 34.3. *Let $\mathcal{A}$ be an algebra of subsets of X. Then*

1. *$\mathcal{A}$ is closed under finite unions: if $A, B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$.*
2. *$\mathcal{A}$ is closed under relative compliments: if $A, B \in \mathcal{A}$, then $A \backslash B \in \mathcal{A}$.*
3. *$\mathcal{A}$ is closed under symmetric differences: if $A, B \in \mathcal{A}$, then $A \Delta B \in \mathcal{A}$.*

Proof.
1. Let $A, B \in \mathcal{A}$. Then

$$\begin{aligned} A \cup B &= ((A \cup B)^c)^c \\ &= (A^c \cap B^c)^c \\ &\in \mathcal{A}. \end{aligned}$$

2. Let $A, B \in \mathcal{A}$. Then

$$\begin{aligned} A \backslash B &= A \cap B^c \\ &\in \mathcal{A}. \end{aligned}$$

3. Let $A, B \in \mathcal{A}$. Then it follows from 1 and 2 that

$$\begin{aligned} A \Delta B &= (A \backslash B) \cup (B \backslash A) \\ &\in \mathcal{A}. \end{aligned}$$

$\square$

34.4 Collection of all Subintervals of $[a,b]$ Forms a Semialgebra

Definition 34.1. A nonempty collection $\mathcal{E}$ of subsets of X is said to be a **semialgebra** of sets if it satisfies the following properties:

1. $\emptyset \in \mathcal{E}$;
2. If $E, F \in \mathcal{E}$, then $E \cap F \in \mathcal{E}$;
3. If $E \in \mathcal{E}$, then E^c is a finite disjoint union of sets from $\mathcal{E}$.

Problem 4.a

Proposition 34.4. *The collection of all subintervals of $[a, b]$ forms a semialgebra of sets.*

Proof. Let $\mathcal{I}$ denote the collection of all subintervals of $[a, b]$. We have $\emptyset \in \mathcal{I}$ since $\emptyset = (c, c)$ for any $c \in [a, b]$.

Now we show $\mathcal{I}$ is closed under finite intersections. Let I_1 and I_2 be subintervals of $[a, b]$. Taking the closure of I_1 and I_2 gives us closed intervals, say

$$\overline{I}_1 = [c_1, d_1] \quad \text{and} \quad \overline{I}_2 = [c_2, d_2].$$

Assume without loss of generality that $c_1 \leq c_2$. If $d_1 < c_2$, then $I_1 \cap I_2 = \emptyset$, so assume that $d_1 \geq c_2$. If $d_1 \geq d_2$, then $I_1 \cap I_2 = I_2$, so assume that $d_1 < d_2$. If $c_1 = c_2$, then $I_1 \cap I_2 = I_1$, so assume that $c_2 > c_1$. So we have reduced the case to where

$$c_1 < c_2 \leq d_1 < d_2.$$

With these assumptions in mind, we now consider four cases:

Case 1: If $I_1 = [c_1, d_1]$ or $I_1 = (c_1, d_1]$ and $I_2 = [c_2, d_2]$ or $I_2 = [c_2, d_2)$, then $I_1 \cap I_2 = [c_2, d_1]$.

Case 2: If $I_1 = [c_1, d_1)$ or $I_1 = (c_1, d_1)$ and $I_2 = [c_2, d_2]$ or $I_2 = [c_2, d_2)$, then $I_1 \cap I_2 = [c_2, d_1)$.

Case 3: If $I_1 = [c_1, d_1]$ or $I_1 = (c_1, d_1)$ and $I_2 = (c_2, d_2]$ or $I_2 = (c_2, d_2)$, then $I_1 \cap I_2 = (c_2, d_1)$.

Case 4: If $I_1 = [c_1, d_1)$ or $I_1 = (c_1, d_1)$ and $I_2 = (c_2, d_2]$ or $I_2 = (c_2, d_2)$, then $I_1 \cap I_2 = (c_2, d_1)$.

In all cases, we see that $I_1 \cap I_2$ is a subinterval of $[a, b]$.

Now we show that compliments can be expressed as finite disjoint unions. Let I be a subinterval of $[a, b]$ and write $\overline{I} = [c, d]$. We consider four cases:

Case 1: If $I = [c, d]$, then $I^c = [a, c) \cup (d, b]$.

Case 2: If $I = (c, d]$, then $I^c = [a, c] \cup (d, b]$.

Case 3: If $I = [c, d)$, then $I^c = [a, c) \cup [d, b]$.

Case 4: If $I = (c, d)$, then $I^c = [a, c] \cup [d, b]$.

Thus in all cases, we can express I^c as a disjoint union of intervals since $a \leq c \leq d \leq b$.

□

Problem 4.b

Proposition 34.5. *Let $\mathcal{I}$ be the collection of all subintervals of $\mathbb{R} \cup \{\infty\}$ of the form $(a, b]$. Then $\mathcal{I}$ forms a semialgebra of sets.*

Proof. We have $\emptyset \in \mathcal{I}$ since $\emptyset = (c, c]$ for any $c \in \mathbb{R} \cup \{\infty\}$.

Now we show $\mathcal{I}$ is closed under finite intersections. Let $I_1 = (c_1, d_1]$ and $I_2 = (c_2, d_2]$. Assume without loss of generality that $c_1 \leq c_2$. Then

$$I_1 \cap I_2 = \begin{cases} (c_2, d_1] & \text{if } c_2 \leq d_1 \\ \emptyset & \text{else} \end{cases}$$

Now we show that compliments can be expressed as finite disjoint unions. Let $I = (c, d]$. Then

$$I^c = (-\infty, c] \cup (d, \infty],$$

where the union is disjoint since $c \leq d$. □

Problem 4.c

Proposition 34.6. *Let $\mathcal{E}$ be a semialgebra of sets. Then the collection $\mathcal{A}$ consisting of all sets which are finite disjoint union of sets in $\mathcal{E}$ forms an algebra of sets.*

Proof. We have $\emptyset \in \mathcal{A}$ since $\emptyset \in \mathcal{E}$.

Next we show that $\mathcal{A}$ is closed under finite intersections. Let $A, A' \in \mathcal{A}$. Express A and A' as a disjoint union of members of $\mathcal{E}$, say

$$A = E_1 \cup \cdots \cup E_n \quad \text{and} \quad A' = E'_1 \cup \cdots \cup E'_{n'}.$$

Then we have

$$\begin{aligned}
A \cap A' &= \left(\bigcup_{i=1}^{n} E_i \right) \cap \left(\bigcup_{i'=1}^{n'} E'_{i'} \right) \\
&= \bigcup_{i'=1}^{n'} \left(\left(\bigcup_{i=1}^{n} E_i \right) \cap E'_{i'} \right) \\
&= \bigcup_{i'=1}^{n'} \left(\bigcup_{i=1}^{n} E_i \cap E'_{i'} \right) \\
&= \bigcup_{\substack{1 \leq i \leq n \\ 1 \leq i' \leq n'}} E_i \cap E'_{i'}
\end{aligned}$$

where the union is disjoint since the E_i and $E'_{i'}$ are disjoint from one another.

Lastly we show that $\mathcal{A}$ is closed under compliments. Let $A \in \mathcal{A}$. Express A as a disjoint union of members of $\mathcal{E}$, say

$$A = E_1 \cup \cdots \cup E_n.$$

Then we have

$$\begin{aligned}
A^c &= (E_1 \cup \cdots \cup E_n)^c \\
&= E_1^c \cap \cdots \cap E_n^c.
\end{aligned}$$

Since the E_i^c belong to $\mathcal{A}$ and $\mathcal{A}$ is closed under finite intersections, it follows that $A^c \in \mathcal{A}$. □

34.5 Collection of Subsets of $\mathbb{Z}$ Forms Algebra Under Certain Conditions

Proposition 34.7. *Let $\mathcal{A}$ be a collection of subsets of $\mathbb{Z}$ such that*

1. *X is a member of $\mathcal{A}$;*
2. *$\mathcal{A}$ is closed under relative compliments: $A \backslash B \in \mathcal{A}$ for all $A, B \in \mathcal{A}$.*

Then $\mathcal{A}$ is an algebra.

Proof. We have $\emptyset \in \mathcal{A}$ since $\emptyset = X \backslash X \in \mathcal{A}$. Clearly $\mathcal{A}$ is closed under compliments since it is closed under relative compliments, so we just need to show that $\mathcal{A}$ is closed under finite intersections. Let $A, B \in \mathcal{A}$. Then

$$\begin{aligned}
A \cap B &= A \cap (B^c)^c \\
&= A \backslash B^c \\
&\in \mathcal{A}.
\end{aligned}$$

□

34.6 Finite Complement Algebra

Proposition 34.8. *Let $\mathcal{A}$ be the collection of subsets of X which satisfies the property that if $A \in \mathcal{A}$ then either A or A^c is finite. Then $\mathcal{A}$ forms an algebra.*

Proof. We have $\varnothing \in \mathcal{A}$ since $\varnothing$ is finite. Clearly $\mathcal{A}$ is closed under compliments since $A \in \mathcal{A}$ implies either A or A^c is finite which implies $A^c \in \mathcal{A}$. It remains to show that $\mathcal{A}$ is closed under finite intersections. Let $A, B \in \mathcal{A}$ and suppose that $A \cap B$ is infinite. We must show that $(A \cap B)^c = A^c \cup B^c$ is finite. In other words, we need to show that both A^c and B^c are finite. Assume for a contradiction that A^c is infinite. Then A must be finite since $A \in \mathcal{A}$. But this implies $A \cap B$ is finite, which is a contradiction. Thus A^c must be finite. Similarly, we can prove by contradiction that B^c is finite too. $\square$

34.7 Ascending Sequence of Algebras is an Algebra

Proposition 34.9. *Let $(\mathcal{A}_n)$ be an ascending sequence of algebras over X, that is, $\mathcal{A}_n$ is an algebra of subsets of X and $\mathcal{A}_n \subseteq \mathcal{A}_{n+1}$ for all $n \in \mathbb{N}$. Then*

$$\mathcal{A} := \bigcup_{n \in \mathbb{N}} \mathcal{A}_n$$

is an algebra.

Proof. We have $\varnothing \in \mathcal{A}$ since $\varnothing \in A_1 \subseteq \mathcal{A}$. Next we show that $\mathcal{A}$ is closed under finite intersections. Let $A, B \in \mathcal{A}$. Then $A \in A_i$ and $B \in \mathcal{A}_j$ for some $i, j \in \mathbb{N}$. Without loss of generality, assume that $i \leq j$. Then $A \in \mathcal{A}_i \subseteq \mathcal{A}_j$. Thus $A \cap B \in \mathcal{A}_j \subseteq \mathcal{A}$. Lastly we show that $\mathcal{A}$ is closed under compliments. Let $A \in \mathcal{A}$. Then $A \in A_i$ for some $i \in \mathbb{N}$. Thus $A^c \in \mathcal{A}_i \subseteq \mathcal{A}$. $\square$

Remark 53. The ascending condition is not necessary. Indeed, consider $X = \{a, b, c\}$ and

$$\begin{aligned} \mathcal{A} &= \{\varnothing, X, \{a\}, \{b, c\}\} \\ \mathcal{B} &= \{\varnothing, X, \{b\}, \{a, c\}\} \\ \mathcal{C} &= \{\varnothing, X, \{c\}, \{a, b\}\} \end{aligned}$$

Then $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are algebras over X, and $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C} = \mathcal{P}(X)$ is an algebra over X, but none of the $\mathcal{A}$, $\mathcal{B}$, or $\mathcal{C}$ contain one another.

35 Homework 2

Throughout this homework, let $(X, \mathcal{M}, \mu)$ be a measure space. We say that $(X, \mathcal{M}, \mu)$ is a **finite** measure space if $\mu(X) < \infty$. Observe that in this case, we have $\mu(A) < \infty$ for all $A \in \mathcal{M}$, by monotonicity of μ.

35.1 Countable Subadditivity of Finite Measure

Proposition 35.1. *Let (E_n) be a sequence of sets in $\mathcal{M}$. Then*

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \mu(E_n)$$

Proof. Disjointify[16] (E_n) into the sequence (D_n); set $D_1 := E_1$ and

$$D_n := E_n \backslash \left(\bigcup_{i=1}^{n-1} E_n\right)$$

for all $n > 1$. Then we have

$$\begin{aligned} \mu\left(\bigcup_{n=1}^{\infty} E_n\right) &= \mu\left(\bigcup_{n=1}^{\infty} D_n\right) \\ &= \sum_{n=1}^{\infty} \mu(D_n) \\ &\leq \sum_{n=1}^{\infty} \mu(E_n), \end{aligned}$$

where we use countable additivity of μ to get from the first line to the second line and where we used monotonicity of μ to get from the second line to the third line. $\square$

[16] See Appendix for details on disjointification.

35.2 Inverse Image of σ-Algebra is σ-Algebra

Proposition 35.2. *Let $(Y, \mathcal{N}, \nu)$ be a measure space and suppose $f\colon X \to Y$ is a function. Then $(X, f^{-1}(\mathcal{N}), f^{-1}\mu)$ is a measure space, where*

$$f^{-1}(\mathcal{N}) = \{f^{-1}(B) \subseteq X \mid B \in \mathcal{N}\}$$

and where $f^{-1}\mu\colon f^{-1}(\mathcal{N}) \to [0, \infty]$ is defined by

$$(f^{-1}\mu)(A) = \mu^*(f(A)).$$

for all $A \in f^{-1}(\mathcal{N})$.

Proof. We first show that $f^{-1}(\mathcal{N})$ is a σ-algebra. This follows from the fact that f^{-1} commutes with unions and compliments:

$$f^{-1}\left(\bigcup_{j\in J} B_j\right) = \bigcup_{j\in J} f^{-1}\left(B_j\right) \quad \text{and} \quad f^{-1}\left(Y \backslash B\right) = f^{-1}(Y) \backslash f^{-1}(B)$$

for all subsets B and B_j of Y for all $j \in J$. Indeed, we have

$$\begin{aligned} x \in \bigcup_{j\in J} f^{-1}\left(B_j\right) &\iff x \in f^{-1}(B_j) \text{ for some } j \in J \\ &\iff f(x) \in B_j \text{ for some } j \in J \\ &\iff f(x) \in \bigcup_{j\in J} B_j \\ &\iff x \in f^{-1}\left(\bigcup_{j\in J} B_j\right) \end{aligned}$$

and we have

$$\begin{aligned} x \in f^{-1}\left(Y \backslash B\right) &\iff f(x) \in Y \backslash B \\ &\iff f(x) \in Y \text{ and } f(a) \notin B \\ &\iff x \in f^{-1}(Y) \text{ and } a \notin f^{-1}(B) \\ &\iff x \in f^{-1}(Y) \backslash f^{-1}(B). \end{aligned}$$

Now we show that the function $f^*\mu$ is a measure. We have

$$\begin{aligned} (f^{-1}\mu)(\emptyset) &= \inf\{\mu(B) \mid f(\emptyset) \subset B\} \\ &= \inf\{\mu(B) \mid \emptyset \subset B\} \\ &\leq \mu(\emptyset) \\ &= 0. \end{aligned}$$

Next, let (A_n) be a sequence of members of $f^{-1}(\mathcal{N})$. Then

$$\begin{aligned} (f^{-1}\mu)\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mu^*\left(f\left(\bigcup_{n=1}^{\infty} A_n\right)\right) \\ &= \mu^*\left(\bigcup_{n=1}^{\infty} f(A_n)\right) \\ &\leq \sum_{n=1}^{\infty} \mu^*(f(A_n)) \\ &= \sum_{n=1}^{\infty} (f^{-1}\mu)(A_n). \end{aligned}$$

Thus $f^{-1}\mu$ is countably subadditive.

Finally, let A and A' be two members in $f^{-1}(\mathcal{N})$ such that $A \cap A' = \emptyset$. Then

$$\begin{aligned}(f^{-1}\mu)(A \cup A') &= \mu^*(f(A) \cup f(A')) \\ &= \\ &= \end{aligned}$$

. $\square$

We first show that

$$\sum_{n=1}^{\infty} \inf \{\mu(B_n) \mid f(A_n) \subset B_n\} \leq \inf \left\{\mu(B) \mid \bigcup_{n=1}^{\infty} f(A_n) \subset B\right\}.$$

Let $\varepsilon > 0$. Choose

35.3 Locally Measurable Sets

Definition 35.1. A set $E \subseteq X$ is called **locally measurable** if $E \cap A \in \mathcal{M}$ for all $A \in \mathcal{M}$ with finite measure.

Proposition 35.3. *Suppose $(X, \mathcal{M}, \mu)$ is a finite measure space. Let $\widetilde{\mathcal{M}}$ be the collection of all locally measurable subsets of X. Then $\mathcal{M} = \widetilde{\mathcal{M}}$.*

Proof. Let first show that $\mathcal{M} \subseteq \widetilde{\mathcal{M}}$. Let $E \in \mathcal{M}$. Then $E \cap A \in \mathcal{M}$ for all $A \in \mathcal{M}$ since $\mathcal{M}$ is closed under finite intersections. In particular, this implies $E \in \widetilde{\mathcal{M}}$. Thus $\mathcal{M} \subseteq \widetilde{\mathcal{M}}$.

Now we show the reverse inclusion $\mathcal{M} \supseteq \widetilde{\mathcal{M}}$. Let $E \in \widetilde{\mathcal{M}}$. Since $\mu(X) < \infty$ and E is locally measurable, we have

$$\begin{aligned} E &= E \cap X \\ &\in \mathcal{M}. \end{aligned}$$

Thus $\mathcal{M} \supseteq \widetilde{\mathcal{M}}$. $\square$

35.4 If $\mu(A) < \infty$, then $\mu(B \backslash A) = \mu(B) - \mu(A)$

Lemma 35.1. *Let $A, B \in \mathcal{M}$ such that $A \subseteq B$. If $\mu(A) < \infty$, then*

$$\mu(B \backslash A) = \mu(B) - \mu(A).$$

Proof. By finite additivity of μ, we have

$$\begin{aligned} \mu(B) &= \mu((B \backslash A) \cup A) \\ &= \mu(B \backslash A) + \mu(A). \end{aligned}$$

If moreover $\mu(A) < \infty$, then we may subtract $\mu(A)$ from both sides to obtain

$$\mu(B \backslash A) = \mu(B) - \mu(A).$$

$\square$

Problem 4.a

Proposition 35.4. *Suppose $(X, \mathcal{M}, \mu)$ is a finite measure space. Then*

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B).$$

for all $A, B \in \mathcal{M}$.

Proof. Let $A, B \in \mathcal{M}$. Then by finite additivity of μ, we have

$$\begin{aligned} \mu(A \cup B) &= \mu(A \cup (B \backslash A)) \\ &= \mu(A) + \mu(B \backslash A) \\ &= \mu(A) + \mu(B \backslash (A \cap B)) \\ &= \mu(A) + \mu(B) - \mu(A \cap B), \end{aligned}$$

where the last equality follows Lemma (35.1) since $(X, \mathcal{M}, \mu)$ is a finite measure space. $\square$

35.5 Countable Additivity of μ for "Almost Pairwise Disjoint" Sets

Proposition 35.5. *Let (A_n) be a sequence of "almost pairwise disjoint" members of $\mathcal{M}$, in the sense that $\mu(A_i \cap A_j) = 0$ whenever $i \neq j$. Then*

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

Proof. First note that countable subadditivity of μ implies

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n),$$

so it suffices to show the reverse inequality. Before doing so, we first prove by induction on $N \geq 1$, that

$$\mu\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} \mu(A_n). \tag{87}$$

The base case $N = 1$ holds trivially. Assume that we have shown (87) holds for some $N > 1$. Then

$$\begin{aligned}
\mu\left(\bigcup_{n=1}^{N+1} A_n\right) &= \mu\left(\left(\bigcup_{n=1}^{N} A_n\right) \cup A_{N+1}\right) \\
&= \mu\left(\bigcup_{n=1}^{N} A_n\right) + \mu(A_{N+1}) - \mu\left(\left(\bigcup_{n=1}^{N} A_n\right) \cap A_{N+1}\right) \\
&= \sum_{n=1}^{N} \mu\left(A_N\right) + \mu(A_{N+1}) - \mu\left(\bigcup_{n=1}^{N} (A_n \cap A_{N+1})\right) \\
&\geq \sum_{n=1}^{N+1} \mu(A_n) - \sum_{n=1}^{N} \mu(A_n \cap A_{N+1}) \\
&= \sum_{n=1}^{N+1} \mu(A_n) - \sum_{n=1}^{N} 0 \\
&= \sum_{n=1}^{N+1} \mu(A_n),
\end{aligned}$$

where we used the induction hypothesis to get from the second line to the third line, and where we used finite subadditivity of μ to get from the third line to the fourth line. We already have

$$\mu\left(\bigcup_{n=1}^{N+1} A_n\right) \leq \sum_{n=1}^{N+1} \mu(A_n)$$

by finite subadditivity of μ, and so it follows that

$$\mu\left(\bigcup_{n=1}^{N+1} A_n\right) = \sum_{n=1}^{N+1} \mu(A_n).$$

Therefore (87) holds for all $N \in \mathbb{N}$ induction.

Now we prove the reverse inequality: for each $N \in \mathbb{N}$, we have

$$\begin{aligned}
\sum_{n=1}^{N} \mu(A_n) &= \mu\left(\bigcup_{n=1}^{N} A_n\right) \\
&\subseteq \mu\left(\bigcup_{n=1}^{\infty} A_n\right)
\end{aligned}$$

by monotonicity of μ. By taking $N \to \infty$, we see that

$$\sum_{n=1}^{\infty} \mu(A_n) \leq \mu\left(\bigcup_{n=1}^{\infty} A_n\right).$$

□

35.6 Nonuniqueness of Extension of Algebra to σ-Algebra

Proposition 35.6. *Let $\mathcal{A}$ be the collection of all finite unions of sets of the form $(a,b] \cap \mathbb{Q}$ where $-\infty \le a \le b \le \infty$. Then*

1. *$\mathcal{A}$ is an algebra of subsets of $\mathbb{Q}$;*
2. *$\sigma(\mathcal{A}) = \mathcal{P}(\mathbb{Q})$ where $\mathcal{P}(\mathbb{Q})$ is the collection of all subsets of $\mathbb{Q}$;*
3. *the function $\mu \colon \mathcal{A} \to [0,\infty]$ defined by $\mu(\emptyset) = 0$ and $\mu(A) = \infty$ for all nonempty $A \in \mathcal{A}$ is a measure on $\mathcal{A}$;*
4. *there is more than one measure on $\sigma(\mathcal{A})$ whose restriction to $\mathcal{A}$ is μ;*

Proof.
1. In Homework 1, it was shown that $(\mathbb{R} \cup \{\infty\}, \mathcal{T})$ was a semialgebra, where $\mathcal{T}$ consisted of all subintervals of $\mathbb{R} \cup \{\infty\}$ of the form $(a,b]$ where $-\infty \le a \le b \le \infty$. If we let $\iota \colon \mathbb{Q} \to \mathbb{R} \cup \{\infty\}$ denote the inclusion map, then we see that $\iota^{-1}(\mathcal{T}) = \mathcal{S}$, where $\mathcal{S}$ denotes the collection of all subintervals of $\mathbb{Q}$ of the form $(a,b] \cap \mathbb{Q}$. It follows easily from Proposition (35.2) that $\mathcal{S}$ is a semialgebra of subsets of $\mathbb{Q}$.[17]
Therefore the set of all finite disjoint unions of members of $\mathcal{S}$ forms an algebra, and as any finite union of members of $\mathcal{S}$ can be expressed as a finite disjoint union of members of $\mathcal{S}$ (since $\mathcal{S}$ is a semialgebra), we see that $\mathcal{A}$ is an algebra.

2. Clearly $\mathcal{P}(\mathbb{Q}) \supseteq \sigma(\mathcal{A})$. Let us prove the reverse inclusion. We first observe that $\{r\} \in \sigma(\mathcal{A})$ for all $r \in \mathbb{Q}$. Indeed, if $r \in \mathbb{Q}$, then we have

$$\{r\} = \bigcap_{n \in \mathbb{N}} (r - 1/n, r] \cap \mathbb{Q} \in \sigma(\mathcal{A})$$

Now let $S \in \mathcal{P}(\mathbb{Q})$. Then since S is countable, we have

$$S = \bigcup_{s \in S} \{s\} \in \sigma(\mathcal{A}).$$

3. We have $\mu(\emptyset) = 0$ by definition. Let (A_n) be a sequence of pairwise disjoint members of $\mathcal{A}$ whose union also belongs to $\mathcal{A}$. If $\bigcup_{n=1}^{\infty} A_n \neq \emptyset$, then $A_n \neq \emptyset$ for some $n \in \mathbb{N}$, and thus

$$\begin{aligned} \mu\left(\bigcup_{n=1}^{\infty} A_n\right) &= \infty \\ &= \mu(A_n) \\ &= \sum_{n=1}^{\infty} \mu(A_n). \end{aligned}$$

Similarly, if $\bigcup_{n=1}^{\infty} A_n = \emptyset$, then $A_n = \emptyset$ for all $n \in \mathbb{N}$, and thus

$$\begin{aligned} \mu\left(\bigcup_{n=1}^{\infty} A_n\right) &= 0 \\ &= \sum_{n=1}^{\infty} 0 \\ &= \sum_{n=1}^{\infty} \mu(A_n). \end{aligned}$$

In both cases, we have

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

[17] Technically we showed that the inverse image of a σ-algebra is a σ-algebra. However the same reasoning used in that proof shows that the inverse image of a semialgebra is a semialgebra: namely f^{-1} commutes with complements and unions.

4. We define $\mu_1 \colon \mathcal{P}(\mathbb{Q}) \to [0,\infty]$ and $\mu_2 \colon \mathcal{P}(\mathbb{Q}) \to [0,\infty]$ by

$$\mu_1(A) = \begin{cases} |A| & \text{if } A \text{ is finite} \\ \infty & \text{else} \end{cases} \quad \text{and} \quad \mu_2(A) = \begin{cases} 0 & \text{if } A = \emptyset \\ \infty & \text{if } A \neq \emptyset \end{cases}$$

for all $A \in \mathcal{P}(\mathbb{Q})$. Both μ_1 and μ_2 restrict to μ as functions since every member of $\mathcal{A}$ is infinite. They are also both distinct as functions since, for example, $\mu_1(\{x\}) = 1$ and $\mu_2(\{x\}) = \infty$ for any $x \in \mathbb{Q}$. Thus it suffices to show that they are measures. That μ_2 is a measure follows from a similar argument as in the case of μ, so we just show that μ_1 is a measure. We have $\mu_1(\emptyset) = 0$ since $|\emptyset| = 0$. Next we show it is finitely additive. Let A and B be members of $\mathcal{P}(\mathbb{Q})$ such that $A \cap B = \emptyset$. If $A = \emptyset$, then

$$\begin{aligned} \mu_1(A \cup B) &= \mu_1(\emptyset \cup B) \\ &= \mu_1(B) \\ &= 0 + \mu_1(B) \\ &= \mu_1(\emptyset) + \mu_1(B) \\ &= \mu_1(A) + \mu_1(B). \end{aligned}$$

Similarly, if $B = \emptyset$, then $\mu_1(A \cup B) = \mu_1(A) \cup \mu_1(B)$. So assume neither A nor B is the emptyset. Write them as

$$A = \{x_1, \dots x_m\} \quad \text{and} \quad B = \{y_1, \dots, y_n\}.$$

Then

$$A \cup B = \{x_1, \dots, x_m, y_1, \dots, y_n\},$$

and so

$$\begin{aligned} \mu_1(A \cup B) &= m + n \\ &= \mu_1(A) + \mu_1(B). \end{aligned}$$

It follows that μ_1 is finitely additive.

Now, let (A_n) be a sequence of pairwise disjoint members of $\mathcal{P}(\mathbb{Q})$. Suppose that $A_n \neq \emptyset$ for only finitely many n, say $n_1, \dots, n_k$. Then it follows from finite additivity of μ_1 and the fact that $\mu(\emptyset) = 0$ that

$$\begin{aligned} \mu_1\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mu_1\left(\bigcup_{i=1}^{k} A_{n_i}\right) \\ &= \sum_{i=1}^{k} \mu(A_{n_i}) \\ &= \sum_{n=1}^{\infty} \mu(A_n). \end{aligned}$$

Now suppose that $A_n \neq \emptyset$ for infinitely many n. By taking a subsequence of (A_n) if necessary, we may assume that $A_n \neq \emptyset$ for all n. Then $\bigcup_{n=1}^{\infty} A_n$ is infinite, and so

$$\begin{aligned} \mu_1\left(\bigcup_{n=1}^{\infty} A_n\right) &= \infty \\ &\geq \sum_{n=1}^{\infty} \mu_1(A_n) \\ &\geq \sum_{n=1}^{\infty} 1 \\ &= \infty. \end{aligned}$$

It follows that

$$\mu_1\left(\bigcup_{n=1}^{\infty} A_n\right) = \infty = \sum_{n=1}^{\infty} \mu_1(A_n).$$

Therefore μ_1 and μ_2 are distinct measures which restrict to μ. $\square$

Remark 54. Note that the extension theorem does not apply here as μ is not a finite measure.

35.7 Symmetric Difference Identities

Proposition 35.7. *Let $A, B \in \mathcal{P}(X)$. Then the following properties hold*

1. $A\Delta A = \emptyset$;
2. $(A\Delta B)\Delta C = A\Delta(B\Delta C)$;
3. $(A\Delta B)\Delta(B\Delta C) = A\Delta C$;
4. $(A\Delta B)\Delta(C\Delta D) = (A\Delta C)\Delta(B\Delta D)$;
5. $|1_A - 1_B| = 1_{A\Delta B}$.

Proof.
1. We have

$$\begin{aligned} A\Delta A &= (A\backslash A) \cup (A\backslash A) \\ &= \emptyset \cup \emptyset \\ &= \emptyset. \end{aligned}$$

2. We have

$$\begin{aligned} (A\Delta B)\Delta C &= ((A\Delta B) \cup C) \cap ((A\Delta B) \cap C)^c \\ &= ((A\Delta B) \cup C) \cap ((A\Delta B)^c \cup C^c) \\ &= (((A \cup B) \cap (A \cap B)^c) \cup C)) \cap ((A \cap B^c) \cup (A^c \cap B))^c \cup C^c) \\ &= (((A \cup B) \cap (A^c \cup B^c)) \cup C)) \cap (((A \cap B^c)^c \cap (A^c \cap B)^c) \cup C^c) \\ &= (A \cup B \cup C) \cap (A^c \cup B^c \cup C) \cap ((A^c \cup B) \cap (A \cup B^c)) \cup C^c) \\ &= (A \cup B \cup C) \cap (A^c \cup B^c \cup C) \cap (A^c \cup B \cup C^c) \cap (A \cup B^c \cup C^c) \\ &= (B \cup C \cup A) \cap (B^c \cup C^c \cup A) \cap (B^c \cup C \cup A^c) \cap (B \cup C^c \cup A^c) \\ &= ((B \cup C \cup A) \cap (B^c \cup C^c \cup A)) \cap ((B^c \cup C) \cap (B \cup C^c)) \cup A^c) \\ &= ((B \cup C \cup A) \cap (B^c \cup C^c \cup A)) \cap (((B \cap C^c)^c \cap (B^c \cap C)^c) \cup A^c) \\ &= (((B \cup C) \cap (B \cap C)^c) \cup A)) \cap ((B \cap C^c) \cup (B^c \cap C))^c \cup A^c) \\ &= ((B\Delta C) \cup A) \cap ((B\Delta C)^c \cup A^c) \\ &= ((B\Delta C) \cup A) \cap ((B\Delta C) \cap A)^c \\ &= (B\Delta C)\Delta A \\ &= A\Delta(B\Delta C) \end{aligned}$$

3. We have

$$\begin{aligned} (A\Delta B)\Delta(B\Delta C) &= A\Delta B\Delta B\Delta C \\ &= A\Delta\emptyset\Delta C \\ &= A\Delta C. \end{aligned}$$

4. We have

$$\begin{aligned} (A\Delta B)\Delta(C\Delta D) &= A\Delta B\Delta C\Delta D \\ &= A\Delta C\Delta B\Delta D \\ &= (A\Delta C)\Delta(B\Delta D) \end{aligned}$$

5. Let $x \in X$. If $x \notin A \cup B$, then

$$\begin{aligned}|1_A(x) - 1_B(x)| &= |0-0| \\ &= 0 \\ &= 0-0 \\ &= 1_{A\cup B}(x) - 1_{A\cap B}(x) \\ &= 1_{A\Delta B}(x).\end{aligned}$$

If $x \in A\backslash B$, then

$$\begin{aligned}|1_A(x) - 1_B(x)| &= |1-0| \\ &= 1 \\ &= 1-0 \\ &= 1_{A\cup B}(x) - 1_{A\cap B}(x) \\ &= 1_{A\Delta B}(x).\end{aligned}$$

If $x \in B\backslash A$, then

$$\begin{aligned}|1_A(x) - 1_B(x)| &= |0-1| \\ &= 1 \\ &= 1-0 \\ &= 1_{A\cup B}(x) - 1_{A\cap B}(x) \\ &= 1_{A\Delta B}(x).\end{aligned}$$

If $x \in A \cap B$, then

$$\begin{aligned}|1_A(x) - 1_B(x)| &= |1-1| \\ &= 0 \\ &= 1-1 \\ &= 1_{A\cup B}(x) - 1_{A\cap B}(x) \\ &= 1_{A\Delta B}(x).\end{aligned}$$

Thus $|1_A(x) - 1_B(x)| = 1_{A\Delta B}(x)$ for all $x \in X$ and hence $|1_A - 1_B| = 1_{A\Delta B}$. □

35.8 More Symmetric Difference Identities

Proposition 35.8. *Let (A_n) and (B_n) be two sequences of sets. Then*

$$\left(\bigcup_{m=1}^{\infty} A_m\right) \Delta \left(\bigcup_{n=1}^{\infty} B_n\right) \subseteq \bigcup_{n=1}^{\infty} (A_n \Delta B_n) \quad \text{and} \quad \left(\bigcap_{m=1}^{\infty} A_m\right) \Delta \left(\bigcap_{n=1}^{\infty} B_n\right) \subseteq \bigcup_{n=1}^{\infty} (A_n \Delta B_n).$$

Proof. We have

$$\begin{aligned}\left(\bigcup_{m=1}^{\infty} A_m\right) \Delta \left(\bigcup_{n=1}^{\infty} B_n\right) &= \left(\left(\bigcup_{m=1}^{\infty} A_m\right) \cup \left(\bigcup_{n=1}^{\infty} B_n\right)\right) \Big\backslash \left(\left(\bigcup_{m=1}^{\infty} A_m\right) \cap \left(\bigcup_{n=1}^{\infty} B_n\right)\right) \\ &= \left(\bigcup_{n=1}^{\infty} (A_n \cup B_n)\right) \Big\backslash \left(\bigcup_{n=1}^{\infty} \bigcup_{m=1}^{\infty} (A_m \cap B_n)\right) \\ &\subseteq \left(\bigcup_{n=1}^{\infty} (A_n \cup B_n)\right) \Big\backslash \left(\bigcup_{n=1}^{\infty} (A_n \cap B_n)\right) \\ &\subseteq \bigcup_{n=1}^{\infty} (A_n \cup B_n) \backslash (A_n \cap B_n) \\ &= \bigcup_{n=1}^{\infty} (A_n \Delta B_n).\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\left(\bigcap_{m=1}^{\infty} A_m\right) \Delta \left(\bigcap_{n=1}^{\infty} B_n\right) &= \left(\bigcap_{m=1}^{\infty} (A_m^c)^c\right) \Delta \left(\bigcap_{n=1}^{\infty} (B_n^c)^c\right) \\
&= \left(\bigcup_{m=1}^{\infty} A_m^c\right)^c \Delta \left(\bigcup_{n=1}^{\infty} B_n^c\right)^c \\
&= \left(\bigcup_{m=1}^{\infty} A_m^c\right) \Delta \left(\bigcup_{n=1}^{\infty} B_n^c\right) \\
&\subseteq \bigcup_{n=1}^{\infty} (A_n^c \Delta B_n^c) \\
&= \bigcup_{n=1}^{\infty} (A_n \Delta B_n).
\end{aligned}$$

□

Appendix

Disjointification

Proposition 35.9. *Let $\mathcal{A}$ be an algebra of subsets of X and let (A_n) be a sequence of sets in $\mathcal{A}$. Then there exists a sequence (D_n) of sets in $\mathcal{A}$ such that*

1. *$D_n \subseteq A_n$ for all $n \in \mathbb{N}$.*
2. *$D_m \cap D_n = \emptyset$ for all $m, n \in \mathbb{N}$ such that $m \neq n$.*
3. *$\bigcup_{m=1}^{n} D_m = \bigcup_{m=1}^{n} A_m$ for all $n \in \mathbb{N}$.*

*We say the sequence (D_n) is the **disjointification** of the sequence (A_n) or that we **disjointify** the sequence (A_n) to the sequence (D_n).*

Proof. Set $D_1 := A_1$ and

$$D_n := A_n \setminus \left(\bigcup_{m=1}^{n-1} A_m\right)$$

for all $n > 1$. It is clear that $D_n \in \mathcal{A}$ and that $D_n \subseteq A_n$ for all $n \in \mathbb{N}$. Let us show that $D_m \cap D_n = \emptyset$ whenever $m \neq n$. Without loss of generality, we may assume that $m < n$. Then since $D_m \subseteq A_m$ and $D_n \cap A_m = \emptyset$, we have $D_m \cap D_n = \emptyset$. It remains to show

$$\bigcup_{m=1}^{n} D_m = \bigcup_{m=1}^{n} A_m$$

for all $n \in \mathbb{N}$. Let $n \in \mathbb{N}$. Since $D_m \subseteq A_m$ for all $m \leq n$, we have

$$\bigcup_{m=1}^{n} D_m \subseteq \bigcup_{m=1}^{n} A_m.$$

To show the reverse inclusion, let $x \in \bigcup_{m=1}^{n} A_m$. Then $x \in A_m$ for some $m = 1, \ldots, n$. Choose m to be the smallest natural number such that $x \in A_m$. Then x belongs to A_m but does not belong to $A_1, \ldots, A_{m-1}$. In other words,

$$x \in D_m \subseteq \bigcup_{k=1}^{n} D_k.$$

This implies the reverse inclusion

$$\bigcup_{m=1}^{n} D_m \supseteq \bigcup_{m=1}^{n} A_m.$$

□

36 Homework 3

Throughout this homework, let X be a set and let $\mathcal{P}(X)$ denote the power set of X.

36.1 Limsup and Liminf (of sets) Identities

Proposition 36.1. *Let (A_n) be a sequence in $\mathcal{P}(X)$. Then*

1. $(\liminf A_n)^c = \limsup A_n^c$;
2. $\liminf A_n = \{x \in X \mid \sum_{n=1}^\infty 1_{A_n^c}(x) < \infty\} = \{x \in X \mid x \in A_n$ *for all but finitely many* $n\}$.
3. $\limsup A_n = \{x \in X \mid \sum_{n=1}^\infty 1_{A_n}(x) = \infty\} = \{x \in X \mid x \in A_{\pi(n)}$ *for all n some subsequence* $(A_{\pi(n)})$ *of* $(A_n)\}$.
4. $\liminf A_n \subseteq \limsup A_n$;
5. $1_{\liminf A_n} = \liminf 1_{A_n}$ *and* $1_{\limsup A_n} = \limsup A_n$.

Proof. 1. We have

$$\begin{aligned}
(\liminf A_n)^c &= \left(\bigcup_{N=1}^\infty \left(\bigcap_{n\geq N} A_n\right)\right)^c \\
&= \bigcap_{N=1}^\infty \left(\left(\bigcap_{n\geq N} A_n\right)^c\right) \\
&= \bigcap_{N=1}^\infty \left(\bigcup_{n\geq N} A_n^c\right) \\
&= \limsup A_n^c.
\end{aligned}$$

2. First note that

$$\begin{aligned}
x \in \liminf A_n &\iff x \in \bigcup_{N=1}^\infty \left(\bigcap_{n\geq N} A_n\right) \\
&\iff x \in \bigcap_{n\geq N} A_n \text{ for some } N \in \mathbb{N} \\
&\iff x \in A_n \text{ for all } n \geq N \text{ for some } N \in \mathbb{N}.
\end{aligned}$$

Now if $x \in A_n$ for all $n \geq N$ for some $N \in \mathbb{N}$, then clearly $x \in A_n$ for all but finitely many n. Conversely, let $x \in A_n$ for all but finitely many n. Set $N = \max\{n \mid x \notin A_n\}$. Then $x \in A_n$ for all $n \geq N$. Thus

$$\liminf A_n = \{x \in X \mid x \in A_n \text{ for all but finitely many } n\}.$$

Similarly, if $x \in X$ such that

$$\sum_{n=1}^\infty 1_{A_n^c}(x) < \infty,$$

then $1_{A_n^c}(x) = 1$ for only finitely many n. In other words, $x \in A_n$ for all but finitely many n. Convsersely, if $x \in A_n$ for all but finitely many n, then $x \in A_n^c$ for only finitely many n, and thus

$$\sum_{n=1}^\infty 1_{A_n^c}(x) < \infty.$$

Therefore

$$\left\{x \in X \mid \sum_{n=1}^\infty 1_{A_n^c}(x) < \infty\right\} = \{x \in X \mid x \in A_n \text{ for all but finitely many } n\}.$$

3. First note that

$$\begin{aligned} x \in \limsup A_n &\iff x \in \bigcap_{N=1}^{\infty} \left(\bigcup_{n \geq N} A_n \right) \\ &\iff x \in \bigcup_{n \geq N} A_n \text{ for all } N \in \mathbb{N} \\ &\iff x \in A_n \text{ for some } n \geq N \text{ for all } N \in \mathbb{N}. \end{aligned}$$

In other words, $x \in \limsup A_n$ if and only if for each $n \in \mathbb{N}$ we can find a $\pi(n) \geq n$ such that $x \in A_{\pi(n)}$, or equivalently, if and only if $x \in A_{\pi(n)}$ for all $n \in \mathbb{N}$ where $(A_{\pi(n)})$ is a subsequence of (A_n). Thus

$$\limsup A_n = \{x \in X \mid x \in A_{\pi(n)} \text{ for some subsequence } (A_{\pi(n)}) \text{ of } (A_n)\}.$$

Similarly, suppose $x \in A_{\pi(n)}$ for all $n \in \mathbb{N}$ where $(A_{\pi(n)})$ is a subsequence of (A_n). Then

$$\begin{aligned} \sum_{n=1}^{\infty} 1_{A_n}(x) &\geq \sum_{n=1}^{\infty} 1_{A_{\pi(n)}}(x) \\ &= \infty. \end{aligned}$$

Conversely, if

$$\sum_{n=1}^{\infty} 1_{A_n}(x) = \infty,$$

then $x \in A_n$ for infinitely many n. Thus there is a subsequence $(A_{\pi(n)})$ of (A_n) such that $x \in A_{\pi(n)}$ for all $n \in \mathbb{N}$. Therefore

$$\left\{ x \in X \mid \sum_{n=1}^{\infty} 1_{A_n}(x) = \infty \right\} = \{x \in X \mid x \in A_{\pi(n)} \text{ for some subsequence } (A_{\pi(n)}) \text{ of } (A_n)\}.$$

4. We have

$$\begin{aligned} x \in \liminf A_n &\iff x \in A_n \text{ for all } n \geq N \text{ for some } N \\ &\implies x \in A_n \text{ for infinitely many } n \\ &\iff x \in \limsup A_n. \end{aligned}$$

Thus

$$\liminf A_n \subseteq \limsup A_n.$$

5. We first show $1_{\liminf A_n} = \liminf 1_{A_n}$. Let $x \in X$. First assume that $x \in \liminf A_n$. Then $x \in A_n$ for all $n \geq N$ for some $N \in \mathbb{N}$. Then

$$\begin{aligned} 1 &\geq \liminf(1_{A_n}(x)) \\ &= \lim_{M \to \infty} \inf\{1_{A_m}(x) \mid m \geq M\} \\ &\geq \inf\{1_{A_n}(x) \mid n \geq N\} \\ &= \inf\{1 \mid n \geq N\} \\ &= 1 \end{aligned}$$

implies

$$\begin{aligned} 1_{\liminf A_n}(x) &= 1 \\ &= \liminf(1_{A_n}(x)) \\ &= (\liminf 1_{A_n})(x). \end{aligned}$$

Now assume that $x \notin \liminf A_n$. Then $x \notin A_n$ for infinitely many n. In particular, for each $N \in \mathbb{N}$, there exists a $\pi(N) \geq N$ such that $x \notin A_{\pi(N)}$. Then

$$\begin{aligned} 0 &\leq \liminf(1_{A_n}(x)) \\ &= \lim_{N\to\infty} \inf\{1_{A_n}(x) \mid n \geq N\} \\ &= \lim_{N\to\infty} 0 \\ &= 0 \end{aligned}$$

implies

$$\begin{aligned} 1_{\liminf A_n}(x) &= 0 \\ &= \liminf(1_{A_n}(x)) \\ &= (\liminf 1_{A_n})(x). \end{aligned}$$

Thus all cases we have $1_{\liminf A_n}(x) = (\liminf 1_{A_n})(x)$, and therefore

$$1_{\liminf A_n} = \liminf 1_{A_n}.$$

Now we will show $1_{\limsup A_n} = \limsup 1_{A_n}$. Let $x \in X$. First assume that $x \notin \limsup A_n$. Then $x \notin A_n$ for all $n \geq N$ for some $N \in \mathbb{N}$. Then

$$\begin{aligned} 0 &\leq \limsup(1_{A_n}(x)) \\ &= \lim_{M\to\infty} \sup\{1_{A_m}(x) \mid m \geq M\} \\ &\leq \sup\{1_{A_n}(x) \mid n \geq N\} \\ &= \sup\{0 \mid n \geq N\} \\ &= 0 \end{aligned}$$

implies

$$\begin{aligned} 1_{\limsup A_n}(x) &= 0 \\ &= \limsup(1_{A_n}(x)) \\ &= (\limsup 1_{A_n})(x). \end{aligned}$$

Now assume that $x \in \limsup A_n$. Then $x \in A_n$ for infinitely many n. In particular, for each $N \in \mathbb{N}$, there exists a $\pi(N) \geq N$ such that $x \in A_{\pi(N)}$. Then

$$\begin{aligned} 1 &\geq \limsup(1_{A_n}(x)) \\ &= \lim_{N\to\infty} \sup\{1_{A_n}(x) \mid n \geq N\} \\ &\geq \lim_{N\to\infty} 1 \\ &= 1 \end{aligned}$$

implies

$$\begin{aligned} 1_{\limsup A_n}(x) &= 0 \\ &= \limsup(1_{A_n}(x)) \\ &= (\limsup 1_{A_n})(x). \end{aligned}$$

Thus all cases we have $1_{\limsup A_n}(x) = (\limsup 1_{A_n})(x)$, and therefore

$$1_{\limsup A_n} = \limsup 1_{A_n}.$$

$\square$

36.2 Limsup, Liminf, and Symmetric Difference Identities

Problem 2.a

Proposition 36.2. *Let (A_n) be a sequence in $\mathcal{P}(X)$. Then*

$$\bigcup_{n=1}^{\infty}(A_n \Delta A_{n+1}) = \bigcup_{n=1}^{\infty} A_n \backslash \bigcap_{m=1}^{\infty} A_m.$$

Proof. Suppose $x \in \bigcup_{n=1}^{\infty}(A_n \Delta A_{n+1})$. Choose $n \in \mathbb{N}$ such that $x \in A_n \Delta A_{n+1}$. Thus either $x \in A_n \backslash A_{n+1}$ or $x \in A_{n+1} \backslash A_n$. Without loss of generality, say $x \in A_n \backslash A_{n+1}$. Then since $x \in A_n$, we see that $x \in \bigcup_{n=1}^{\infty} A_n$ and since $x \notin A_{n+1}$, we see that $x \notin \bigcap_{n=1}^{\infty} A_n$. Therefore $x \in \bigcup_{n=1}^{\infty} A_n \backslash \bigcap_{m=1}^{\infty} A_m$. This implies

$$\bigcup_{n=1}^{\infty}(A_n \Delta A_{n+1}) \subseteq \bigcup_{n=1}^{\infty} A_n \backslash \bigcap_{m=1}^{\infty} A_m.$$

Conversely, suppose $x \in \bigcup_{n=1}^{\infty} A_n \backslash \bigcap_{m=1}^{\infty} A_m$. Since $x \in \bigcup_{n=1}^{\infty} A_n$, there exists some $n \in \mathbb{N}$ such that $x \in A_n$. Since $x \notin \bigcap_{m=1}^{\infty} A_m$, there exists some $k \in \mathbb{N}$ such that $x \notin A_k$. Assume without loss of generality that $k < n$. Choose m to be the least natural number number such that $x \in A_m$, $x \notin A_{m-1}$, and $k < m \leq n$. Clearly this number exists since $x \notin A_k$ and $x \in A_n$. Then $x \in A_m \Delta A_{m-1}$, which implies $x \in \bigcup_{n=1}^{\infty}(A_n \Delta A_{n+1})$. Thus

$$\bigcup_{n=1}^{\infty}(A_n \Delta A_{n+1}) \supseteq \bigcup_{n=1}^{\infty} A_n \backslash \bigcap_{m=1}^{\infty} A_m.$$

□

Problem 2.b

Proposition 36.3. *Let (A_n) be a sequence in $\mathcal{P}(X)$. Then*

$$\limsup A_n \backslash \liminf A_n = \limsup (A_n \Delta A_{n+1}).$$

Proof. Suppose $x \in \limsup A_n \backslash \liminf A_n$. Then the sets

$$\{n \in \mathbb{N} \mid x \in A_n\} \quad \text{and} \quad \{n \in \mathbb{N} \mid x \notin A_n\}$$

are both infinite. We claim this implies that the set

$$\begin{aligned}\{n \in \mathbb{N} \mid x \in A_n \Delta A_{n+1}\} &= \{n \in \mathbb{N} \mid x \in (A_n \backslash A_{n+1}) \cup (A_{n+1} \backslash A_n)\} \\ &= \{n \in \mathbb{N} \mid x \in A_n \backslash A_{n+1}\} \cup \{n \in \mathbb{N} \mid x \in A_{n+1} \backslash A_n\}\end{aligned}$$

is infinite. To see this, we first assume without loss of generality that $x \in A_1$. Choose the least $\pi(1) > 1$ such that $x \notin A_{\pi(1)}$ and $x \in A_{\pi(1)-1}$. Observe that $\pi(1)$ exists since otherwise $\{n \in \mathbb{N} \mid x \notin A_n\}$ would be finite. Next, choose $\pi(2) > \pi(1)$ such that $x \in A_{\pi(2)}$ and $x \notin A_{\pi(2)-1}$. We again observe that $\pi(2)$ exists since otherwise $\{n \in \mathbb{N} \mid x \in A_n\}$ would be finite. Continuing in this manner, we obtain a strictly increasing sequence $(\pi(n))$ of natural numbers with

$$x \in A_{\pi(2n)} \backslash A_{\pi(2n)-1} \quad \text{and} \quad x \in A_{\pi(2n-1)-1} \backslash A_{\pi(2n-1)}$$

for all $n \geq 1$. In particular, both sets

$$\{n \in \mathbb{N} \mid x \in A_n \backslash A_{n+1}\} \quad \text{and} \quad \{n \in \mathbb{N} \mid x \in A_{n+1} \backslash A_n\}$$

are infinite. Thus $\{n \in \mathbb{N} \mid x \in A_n \Delta A_{n+1}\}$ is infinite, which implies $x \in \limsup (A_n \Delta A_{n+1})$. Therefore

$$\limsup A_n \backslash \liminf A_n \subseteq \limsup (A_n \Delta A_{n+1}).$$

Conversely, suppose $x \in \limsup (A_n \Delta A_{n+1})$. Then the set

$$\begin{aligned}\{n \in \mathbb{N} \mid x \in A_n \Delta A_{n+1}\} &= \{n \in \mathbb{N} \mid x \in (A_n \backslash A_{n+1}) \cup (A_{n+1} \backslash A_n)\} \\ &= \{n \in \mathbb{N} \mid x \in A_n \backslash A_{n+1}\} \cup \{n \in \mathbb{N} \mid x \in A_{n+1} \backslash A_n\}\end{aligned}$$

is infinite. This implies one of

$$\{n \in \mathbb{N} \mid x \in A_n \backslash A_{n+1}\} \quad \text{or} \quad \{n \in \mathbb{N} \mid x \in A_{n+1} \backslash A_n\}$$

is infinite. Without loss of generality, suppose $\{n \in \mathbb{N} \mid x \in A_n \backslash A_{n+1}\}$ is infinite. Thus there exists a strictly increasing sequence $(\pi(n))$ of natural numbers with $x \in A_{\pi(n)}$ and $x \notin A_{\pi(n)+1}$. In particular, both sets

$$\{n \in \mathbb{N} \mid x \in A_n\} \quad \text{and} \quad \{n \in \mathbb{N} \mid x \notin A_n\}$$

are infinite. Equivalently, we have $x \in \limsup A_n \backslash \liminf A_n$. Therefore

$$\limsup A_n \backslash \liminf A_n \supseteq \limsup(A_n \Delta A_{n+1}).$$

□

36.3 Measure of Intersection of a Descending Sequence of Sets

Proposition 36.4. *Let $(X, \mathcal{M}, \mu)$ be a measure space and let (E_n) be a descending sequence in $\mathcal{M}$ such that $\mu(E_1) < \infty$. Then*

$$\mu\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \mu(E_n) \tag{88}$$

Proof. The sequence $(E_1 \backslash E_n)_{n\in\mathbb{N}}$ is an ascending sequence in $\mathcal{M}$, hence

$$\begin{aligned}
\mu(E_1) - \lim_{n\to\infty} \mu(E_n) &= \lim_{n\to\infty} \mu(E_1) - \lim_{n\to\infty} \mu(E_n) \\
&= \lim_{n\to\infty} (\mu(E_1) - \mu(E_n)) \\
&= \lim_{n\to\infty} \mu(E_1 \backslash E_n) \\
&= \mu\left(\bigcup_{n=1}^{\infty} (E_1 \backslash E_n)\right) \\
&= \mu\left(E_1 \backslash \left(\bigcap_{n=1}^{\infty} E_n\right)\right) \\
&= \mu(E_1) - \mu\left(\bigcap_{n=1}^{\infty} E_n\right),
\end{aligned}$$

where we used the fact that $\mu(E_1) < \infty$ to get from the second line to the third line and also from fifth line to the sixth line. Also since $\mu(E_1) < \infty$, we can subtract $\mu(E_1)$ from both sides to obtain (88). □

36.4 Measure of Liminf is Less Than or Equal to Liminf of Measure

Proposition 36.5. *Let $(X, \mathcal{M}, \mu)$ be a measure space and let (E_n) be a sequence in $\mathcal{M}$. Then*

$$\mu(\liminf E_n) \leq \liminf \mu(E_n)$$

Proof. Note that the sequence

$$\left(\bigcap_{n\geq N} E_n\right)_{N\in\mathbb{N}}$$

is an ascending sequence in N. Therefore we have

$$\begin{aligned}
\mu(\liminf E_n) &= \mu\left(\bigcup_{N=1}^{\infty}\left(\bigcap_{n\geq N} E_n\right)\right) \\
&= \liminf \mu\left(\bigcap_{n\geq N} E_n\right) \\
&\leq \lim_{N\to\infty} \inf\{\mu(E_n) \mid n \geq N\} \\
&= \liminf \mu(E_n),
\end{aligned}$$

where we obtained the third line from the second line since

$$\mu\left(\bigcap_{n\geq N} E_n\right) \leq \mu(E_n)$$

for all $n \geq N$ by monotonicity of μ. □

36.5 Measure of Limsup is Greater Than or Equal to Limsup of Measure (Assuming Some Finiteness Condition)

Proposition 36.6. *Let $(X, \mathcal{M}, \mu)$ be a measure space and let (E_n) be a sequence in $\mathcal{M}$ such that $\mu\left(\bigcup_{n=1}^{\infty} E_n\right) < \infty$. Then*

$$\mu\left(\limsup E_n\right) \geq \limsup \mu(E_n)$$

Proof. Note that the sequence

$$\left(\bigcup_{n\geq N} E_n\right)_{N\in\mathbb{N}}$$

is a descending sequence in N. This together with the fact that $\mu\left(\bigcup_{n=1}^{\infty} E_n\right) < \infty$ implies

$$\begin{aligned}
\mu\left(\limsup E_n\right) &= \mu\left(\bigcap_{N=1}^{\infty}\left(\bigcup_{n\geq N} E_n\right)\right) \\
&= \lim_{N\to\infty} \mu\left(\bigcup_{n\geq N} E_n\right) \\
&\geq \lim_{N\to\infty} \sup\left\{\mu(E_n) \mid n \geq N\right\} \\
&= \limsup \mu(E_n),
\end{aligned}$$

where we obtained the third line from the second line since

$$\mu\left(\bigcup_{n\geq N} E_n\right) \geq \mu(E_n)$$

for all $n \geq N$ by monotonicity of μ. □

36.6 Assuming Some Finiteness Condition, Measure of Limsup is Zero

Proposition 36.7. *Let $(X, \mathcal{M}, \mu)$ be a measure space and let (E_n) be a sequence in $\mathcal{M}$ such that $\sum_{n=1}^{\infty} \mu(E_n) < \infty$. Then*

$$\mu\left(\limsup E_n\right) = 0.$$

Proof. Note that the sequence

$$\left(\bigcup_{n\geq N} E_n\right)_{N\in\mathbb{N}}$$

is a descending sequence in N. This together with the fact that $\mu\left(\bigcup_{n=1}^{\infty} E_n\right) < \infty$ implies

$$\begin{aligned}
\mu\left(\limsup E_n\right) &= \mu\left(\bigcap_{N=1}^{\infty}\left(\bigcup_{n\geq N} E_n\right)\right) \\
&= \lim_{N\to\infty} \mu\left(\bigcup_{n\geq N} E_n\right) \\
&\leq \lim_{N\to\infty} \sum_{n=N}^{\infty} \mu(E_n) \\
&= 0,
\end{aligned}$$

where the last equality follows since $\sum_{n=1}^{\infty} \mu(E_n) < \infty$. □

36.7 Our Measure Equivalence Relation

Let $\mathcal{A}$ be an algebra of subsets of X and let μ be a finite measure on $\mathcal{A}$. Let μ^* be the outer measure on X induced by μ. Define a relation $\sim$ on $\mathcal{P}(X)$ as follows: if $A, B \in \mathcal{P}(X)$, then

$$A \sim B \text{ if and only if } \mu^*(A\Delta B) = 0.$$

We also define the pseudometric d_μ on $\mathcal{P}(X)$ by

$$\mathrm{d}_\mu(A, B) = \mu^*(A\Delta B)$$

for all $A, B \in \mathcal{P}(X)$.

Problem 4.a

Proposition 36.8. *The relation $\sim$ is an equivalence relation.*

Proof. We first check reflexivity. Let $A \in \mathcal{P}(X)$. Then

$$\begin{aligned} \mu^*(A\Delta A) &= \mu^*(\emptyset) \\ &= 0 \end{aligned}$$

implies $A \sim A$. Next we check symmetry. Let $A, B \in \mathcal{P}(X)$ and suppose $A \sim B$. Then

$$\begin{aligned} \mu^*(B\Delta A) &= \mu^*(A\Delta B) \\ &= 0 \end{aligned}$$

implies $B \sim A$. Finally we check transitivity. Let $A, B, C \in \mathcal{P}(X)$ and suppose $A \sim B$ and $B \sim C$. Then

$$\begin{aligned} \mu^*(A\Delta C) &= \mu^*(A\Delta B\Delta B\Delta C) \\ &\leq \mu^*((A\Delta B) \cup (B\Delta C)) \\ &\leq \mu^*(A\Delta B) + \mu^*(B\Delta C) \\ &= 0 + 0 \\ &= 0 \end{aligned}$$

implies $A \sim C$. □

Problem 4.b

Proposition 36.9. *Let $A, B \in \mathcal{P}(X)$. If $A \sim B$, then $\mu^*(A) = \mu^*(B)$. The converse need not be true.*

Proof. Suppose that $A \sim B$. Then $\mu^*(A\Delta B) = 0$ implies

$$\begin{aligned} \mu^*(A) &= \mu^*(A) + \mu^*(A\Delta B) \\ &\geq \mu^*(A \cup (A\Delta B)) \\ &\geq \mu^*(A\Delta A\Delta B) \\ &= \mu^*(B). \end{aligned}$$

Similarly,

$$\begin{aligned} \mu^*(B) &= \mu^*(B) + \mu^*(B\Delta A) \\ &\geq \mu^*(B \cup (B\Delta A)) \\ &\geq \mu^*(B\Delta B\Delta A) \\ &= \mu^*(A). \end{aligned}$$

Thus $\mu^*(A) = \mu^*(B)$.

To see that the converse does not hold, consider the case where $X = \{a, b\}$ and μ is counting measure on this set. Then on the one hand, we have

$$\mu(\{a\}) = 1 = \mu(\{b\}),$$

but on the other hand, we have

$$\begin{aligned}\mu(\{a\}\Delta\{b\}) &= \mu(\{a,b\}) \\ &= 2 \\ &\neq 0.\end{aligned}$$

□

36.8 μ^*-Measurable Forms σ-Algebra

Let $\mathcal{A}$ be an algebra of subsets of X and let μ be a finite measure on $\mathcal{A}$. Let μ^* be the outer measure on X induced by μ. A set E is said to be μ^*-measurable if

$$\mu^*(S) \geq \mu^*(S \cap E) + \mu^*(S \backslash E)$$

for all $S \in \mathcal{P}(X)$. Note that by countable subadditivity of μ^*, this implies

$$\mu^*(S) = \mu^*(S \cap E) + \mu^*(S \backslash E).$$

Denote by $\mathcal{M}$ to be the collection of all μ^*-measurable sets.

Problem 6.a

Proposition 36.10. *Let $A \in \mathcal{A}$. Then A is μ^*-measurable.*

Proof. Let $S \in \mathcal{P}(X)$. Assume for a contradiction that

$$\mu^*(S) < \mu^*(S \cap A) + \mu^*(S \backslash A).$$

Choose $\varepsilon > 0$ such that

$$\mu^*(S) < \mu^*(S \cap A) + \mu^*(S \backslash A) - \varepsilon.$$

Choose $B \in \mathcal{A}$ such that $S \subseteq B$ and

$$\mu(B) \leq \mu^*(S) + \varepsilon.$$

Then

$$\begin{aligned}\mu^*(S) &\geq \mu(B) - \varepsilon \\ &= \mu\left((B \cap A) \cup (B \backslash A)\right) - \varepsilon \\ &= \mu(B \cap A) + \mu(B \backslash A) - \varepsilon \\ &\geq \mu^*(S \cap A) + \mu^*(S \backslash A) - \varepsilon.\end{aligned}$$

This is a contradiction. □

Problem 6.b

Proposition 36.11. *$\mathcal{M}$ is a σ-algebra.*

Proof. We prove this in several steps:

Step 1: We first show $\mathcal{M}$ is an algebra. First we show it is closed under finite unions. Let $A, B \in \mathcal{M}$ and let $S \in \mathcal{P}(X)$. Then

$$\begin{aligned}\mu^*(S) &= \mu^*\left(S \cap A\right) + \mu^*\left(S \backslash A\right) \\ &= \mu^*\left(S \cap A\right) + \mu^*\left((S \backslash A) \cap B\right) + \mu^*\left((S \backslash A) \backslash B\right) \\ &\geq \mu^*\left((S \cap A) \cup ((S \backslash A) \cap B)\right) + \mu^*\left((S \backslash A) \backslash B\right) \\ &= \mu^*\left(S \cap (A \cup B)\right) + \mu^*\left(S \backslash (A \cup B)\right)\end{aligned}$$

Therefore $A \cap B \in \mathcal{M}$.

Next we shows it is closed under complements. Let $A \in \mathcal{M}$ and let $S \in \mathcal{P}(X)$. Then

$$
\begin{aligned}
\mu^*(S) &\geq \mu^*(S \cap A) + \mu^*(S \backslash A) \\
&= \mu^*\left(S \backslash (X \backslash A)\right) + \mu^*(S \backslash A) \\
&= \mu^*\left(S \backslash (X \backslash A)\right) + \mu^*\left(S \cap (X \backslash A)\right).
\end{aligned}
$$

Therefore $X \backslash A \in \mathcal{M}$.

Step 2: We show μ^* is finitely additive on $\mathcal{M}$. In fact, we claim that for any $S \in \mathcal{P}(X)$ and pairwise disjoint $A_1, \ldots, A_n \in \mathcal{M}$, we have

$$
\mu^*\left(S \cap \left(\bigcup_{m=1}^{n} A_m\right)\right) = \sum_{m=1}^{n} \mu^*\left(S \cap A_m\right). \tag{89}
$$

We prove (89) by induction on n. The equality holds trivially for $n = 1$. For the induction step, assume that it holds for some $n \geq 1$. Let S be a subset of X and let$A_1, \ldots, A_{n+1}$ be a finite sequence of members in $\mathcal{M}$. Then

$$
\begin{aligned}
\mu^*\left(S \cap \left(\bigcup_{m=1}^{n+1} A_m\right)\right) &\geq \mu^*\left(S \cap \left(\bigcup_{m=1}^{n+1} A_m\right) \cap A_{n+1}\right) + \mu^*\left(S \cap \left(\bigcup_{m=1}^{n+1} A_m\right) \cap (X \backslash A_{n+1})\right) \\
&= \mu^*\left(S \cap A_{n+1}\right) + \mu^*\left(S \cap \left(\bigcup_{m=1}^{n} A_m\right)\right) \\
&= \mu^*\left(S \cap A_{n+1}\right) + \sum_{m=1}^{n} \mu^*(S \cap A_m) \\
&= \sum_{m=1}^{n+1} \mu^*(S \cap A_m).
\end{aligned}
$$

This establishes (89). Setting $S = X$ in (89) gives us finite additivity of μ^* on $\mathcal{M}$.

Step 3: We prove that $\mathcal{M}$ is a σ-algebra. Since $\mathcal{M}$ was already shown to be an algebra, it suffices to show that $\mathcal{M}$ is closed under countable unions. Let (A_n) be a sequence in $\mathcal{M}$. Disjointify the sequence (A_n) to the sequence (D_n): set $D_1 = A_1$ and $D_n = A_n \backslash \bigcup_{m=1}^{n-1} A_m$ for all $n > 1$. Note that (D_n) is a sequence in $\mathcal{M}$ since $\mathcal{M}$ is algebra. Let $S \in \mathcal{P}(X)$ and $n \in \mathbb{N}$. Observe that

$$
\begin{aligned}
\mu^*(S) &\geq \mu^*\left(S \cap \left(\bigcup_{m=1}^{n} D_m\right)\right) + \mu^*\left(S \backslash \left(\bigcup_{m=1}^{n} D_m\right)\right) \\
&\geq \mu^*\left(S \cap \left(\bigcup_{m=1}^{n} D_m\right)\right) + \mu^*\left(S \backslash \left(\bigcup_{n \in \mathbb{N}} A_n\right)\right) \\
&= \sum_{m=1}^{n} \mu^*\left(S \cap D_m\right) + \mu^*\left(S \backslash \left(\bigcup_{n \in \mathbb{N}} A_n\right)\right),
\end{aligned}
$$

where we applied finite-additivity of μ^* to the first term on the right-hand side and we applied monotonicity of μ^* to the second term on the right-hand side. Taking the limit as $n \to \infty$. We obtain

$$
\begin{aligned}
\mu^*(S) &\geq \sum_{m=1}^{\infty} \mu^*\left(S \cap D_m\right) + \mu^*\left(S \backslash \left(\bigcup_{n \in \mathbb{N}} A_n\right)\right) \\
&\geq \mu^*\left(\bigcup_{n \in \mathbb{N}} (S \cap D_m)\right) + \mu^*\left(S \backslash \left(\bigcup_{n \in \mathbb{N}} A_n\right)\right) \\
&= \mu^*\left(S \cap \bigcup_{n \in \mathbb{N}} D_m\right) + \mu^*\left(S \backslash \left(\bigcup_{n \in \mathbb{N}} A_n\right)\right) \\
&= \mu^*\left(S \cap \left(\bigcup_{n \in \mathbb{N}} A_m\right)\right) + \mu^*\left(S \backslash \left(\bigcup_{n \in \mathbb{N}} A_n\right)\right),
\end{aligned}
$$

where we applied countable subadditivity of μ^* to the first expression on the right-hand side. Thus $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{M}$.

□

Problem 6.c

Proposition 36.12. *We have $\sigma(\mathcal{A}) \subseteq \mathcal{M}$.*

Proof. By problem 6.a and 6.b, we see that $\mathcal{M}$ is a σ-algebra which contains $\mathcal{A}$. Since $\sigma(\mathcal{A})$ is the *smallest* σ-algebra which contains $\mathcal{A}$, we must have $\sigma(\mathcal{A}) \subseteq \mathcal{M}$. $\square$

Problem 6.d

Proposition 36.13. *The outer measure μ^* restricted to $\mathcal{M}$ is a measure.*

Proof. In Proposition (36.11), we showed that μ^* is finitely additive on $\mathcal{M}$. We already know that μ^* is already countably subadditive on $\mathcal{M}$. Therefore μ^* is countably additive on $\mathcal{M}$ since

$$\text{finte additivity + countable subadditivity} \;=\; \text{countable additivity.}$$

To see this, let (A_n) be a sequence of pairwise disjoint members of $\mathcal{M}$. By countable subadditivity of μ^*, we have

$$\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu^*(A_n).$$

For the reverse inequality, notat that for each $N \in \mathbb{N}$, finite additivity of μ^* imlpies

$$\begin{aligned}\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) &\geq \mu^*\left(\bigcup_{n=1}^{N} A_n\right) \\ &= \sum_{n=1}^{N} \mu^*(A_n).\end{aligned}$$

Taking $N \to \infty$ gives us

$$\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) \geq \sum_{n=1}^{\infty} \mu^*(A_n).$$

$\square$

Problem 6.e

Proposition 36.14. *Let $E \in \mathcal{M}$ such that $\mu^*(E) = 0$, and let $F \in \mathcal{P}(X)$ such that $F \subseteq E$. Then $F \in \mathcal{M}$.*

Proof. Let $S \in \mathcal{P}(X)$. Then

$$\begin{aligned}\mu^*(S) &\geq \mu^*(S \backslash F) \\ &= \mu^*(S \cap F) + \mu^*(S \backslash F),\end{aligned}$$

where we used the fact that $\mu^*(S \cap F) = 0$ since $S \cap F \subseteq E$ and $\mu^*(E) = 0$. $\square$

More generally:

Proposition 36.15. *Let $E \in \mathcal{P}(X)$ such that $\mu^*(E) = 0$. Then $E \in \mathcal{M}$.*

Proof. Let $S \in \mathcal{P}(X)$. First note that

$$\begin{aligned}0 &= \mu^*(E) \\ &\geq \mu^*(S \cap E)\end{aligned}$$

implies $\mu^*(S \cap E) = 0$ by monotonicity of μ^*. Therefore

$$\begin{aligned}\mu^*(S) &\geq \mu^*(S \backslash E) \\ &= \mu^*(S \cap E) + \mu^*(S \backslash E).\end{aligned}$$

This implies $E \in \mathcal{M}$. $\square$

Throughout this homework, let $(X, \mathcal{M}, \mu)$ be a measure space.

37 Homework 4

37.1 Characteristic Function Identities

Proposition 37.1. *Let $A, B \in \mathcal{P}(X)$. Then*

1. $1_{A\cap B} = 1_A 1_B$;
2. $1_{A\cup B} = 1_A + 1_B - 1_A 1_B$;
3. $1_{A^c} = 1 - 1_A$;

Proof. 1. Let $x \in X$. If $x \in A \cap B$, then $x \in A$ and $x \in B$, and thus we have

$$\begin{aligned} 1_{A\cap B}(x) &= 1 \\ &= 1 \cdot 1 \\ &= 1_A(x) 1_B(x). \end{aligned}$$

If $x \notin A \cap B$, then either $x \notin A$ or $x \notin B$. Without loss of generality, say $x \notin A$. Then we have

$$\begin{aligned} 1_{A\cap B}(x) &= 0 \\ &= 0 \cdot 1_B(x) \\ &= 1_A(x) 1_B(x). \end{aligned}$$

Therefore the functions $1_{A\cap B}$ and $1_A 1_B$ agree on all of X, and hence must be equal to each other.

2. Let $x \in X$. If $x \in A \cup B$, then either $x \in A$ or $x \in B$. Without loss of generality, say $x \in A$. Then we have

$$\begin{aligned} 1_{A\cup B}(x) &= 1 \\ &= 1 + 1_B(x) - 1_B(x) \\ &= 1 + 1_B(x) - 1 \cdot 1_B(x) \\ &= 1_A(x) + 1_B(x) - 1_A(x) 1_B(x). \end{aligned}$$

If $x \notin A \cup B$, then $x \notin A$ and $x \notin B$. Therefore we have

$$\begin{aligned} 1_{A\cup B}(x) &= 0 \\ &= 0 + 0 - 0 \cdot 0 \\ &= 1_A(x) + 1_B(x) - 1_A(x) 1_B(x). \end{aligned}$$

Thus the functions $1_{A\cup B}$ and $1_A + 1_B - 1_A 1_B$ agree on all of X, and hence must be equal to each other.

3. Let $x \in X$. If $x \in A$, then $x \notin A^c$, hence

$$\begin{aligned} 1_{A^c}(x) &= 0 \\ &= 1 - 1 \\ &= 1 - 1_A(x). \end{aligned}$$

If $x \notin A$, then $x \in A^c$, hence

$$\begin{aligned} 1_{A^c}(x) &= 1 \\ &= 1 - 0 \\ &= 1 - 1_A(x). \end{aligned}$$

Therefore the functions 1_{A^c} and $1 - 1_A$ agree on all of X, and hence must be equal to each other. $\square$

Proposition 37.2. *A product of simple functions is a simple function.*

Proof. Let $\varphi = \sum_{i=1}^{m} a_i 1_{A_i}$ and $\psi = \sum_{j=1}^{n} b_j 1_{B_j}$ be two simple functions. Then

$$\begin{aligned}\varphi \cdot \psi &= \left(\sum_{i=1}^{m} a_i 1_{A_i}\right)\left(\sum_{j=1}^{n} b_j 1_{B_j}\right)\\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} (a_i 1_{A_i})(b_j 1_{B_j})\\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j (1_{A_i} 1_{B_j})\\ &= \sum_{\substack{1\le i\le m\\ 1\le j\le n}} a_i b_j 1_{A_i\cap B_j}.\end{aligned}$$

Since $A_i \cap B_j \in \mathcal{M}$ for all i, j, it follows that $\varphi \cdot \psi$ is simple. □

37.2 E is Measurable if and only if 1_E is Measurable

Proposition 37.3. *Let $E \in \mathcal{P}(X)$. Then $E \in \mathcal{M}$ if and only if 1_E is measurable.*

Proof. Suppose 1_E is measurable. Then

$$E^c = \{x \in X \mid 1_E(x) < 1\}$$

is measurable. This implies E is measurable.

Conversely, suppose E is measurable. Let $c \in \mathbb{R}$. We have three cases

$$\{x \in X \mid 1_E(x) < c\} = \begin{cases} X & \text{if } 1 < c \\ E^c & \text{if } 0 < c \le 1 \\ \emptyset & \text{if } c \le 0 \end{cases}$$

In all three cases, we see that $1_E^{-1}(-\infty, c) \in \mathcal{M}$. This implies 1_E is is measurable. □

37.3 Defining Measure from Integral Weighted by Nonnegative Simple Function

Proposition 37.4. *Let $\phi\colon X \to [0,\infty)$ be a nonnegative simple function. Define a function $\nu\colon \mathcal{M} \to [0,\infty]$ by*

$$\nu(E) = \int_X \phi 1_E \mathrm{d}\mu$$

for all $E \in \mathcal{M}$. Then ν is a measure on $(X, \mathcal{M})$.

Proof. First note that

$$\begin{aligned}\nu(\emptyset) &= \int_X \phi 1_\emptyset \mathrm{d}\mu\\ &= \int_X \phi \cdot 0 \cdot \mathrm{d}\mu\\ &= \int_X 0 \cdot \mathrm{d}\mu\\ &= 0.\end{aligned}$$

Now we show that ν is finitely additive. Let $(E_n)_{n=1}^N$ be a finite sequence of pairwise disjoint sets in $\mathcal{M}$. Then

$$\begin{aligned}
\nu\left(\bigcup_{n=1}^N E_n\right) &= \int_X \phi 1_{\bigcup_{n=1}^N E_n} \mathrm{d}\mu \\
&= \int_X \phi \sum_{n=1}^N 1_{E_n} \mathrm{d}\mu \\
&= \sum_{n=1}^N \int_X \phi 1_{E_n} \mathrm{d}\mu \\
&= \sum_{n=1}^N \nu(E_n),
\end{aligned}$$

where we used the fact that each $\phi 1_{E_n}$ is a nonnegative simple function in order to commute the finite sum with the integral. Thus it follows that ν is finitely additive. It remains to show that ν is countably subadditive. Let (E_n) be a sequence of sets in $\mathcal{M}$. We want to show that

$$\int_X \phi 1_{\bigcup_{n=1}^\infty E_n} \mathrm{d}\mu \leq \sum_{n=1}^\infty \int_X \phi 1_{E_n} \mathrm{d}\mu. \tag{90}$$

To do this, we will show that the sum on the righthand side in (90) is greater than or equal to all integrals of the form $\int \varphi \mathrm{d}\mu$ where $\varphi \colon X \to [0,\infty]$ is a simple function such that $\varphi \leq \phi 1_{\bigcup_{n=1}^\infty E_n}$. Then the inequality (90) will follow from the fact that the integral on the lefthand side in (90) is the supremum of this set. So let $\varphi \colon X \to [0,\infty]$ be a simple function such that $\varphi \leq \phi 1_{\bigcup_{n=1}^\infty E_n}$. Write φ and ϕ in terms of their canonical forms, say

$$\varphi = \sum_{i=1}^k a_i 1_{A_i} \quad \text{and} \quad \phi = \sum_{j=1}^m b_j 1_{B_j}.$$

So $a_i \neq a_{i'}$ and $A_i \cap A_{i'} = \emptyset$ whenever $i \neq i'$ and $b_j \neq b_{j'}$ and $B_j \cap B_{j'} = \emptyset$ whenever $j \neq j'$. Observe that the canonical representation of $\phi 1_{\bigcup_{n=1}^\infty E_n}$ is given by

$$\begin{aligned}
\phi 1_{\bigcup_{n=1}^\infty E_n} &= \left(\sum_{j=1}^m b_j 1_{B_j}\right) 1_{\bigcup_{n=1}^\infty E_n} \\
&= \sum_{j=1}^m b_j 1_{B_j} 1_{\bigcup_{n=1}^\infty E_n} \\
&= \sum_{j=1}^m b_j 1_{\bigcup_{n=1}^\infty B_j \cap E_n},
\end{aligned}$$

where this representation is the canonical representation since $b_j \neq b_{j'}$ and

$$\left(\bigcup_{n=1}^\infty B_j \cap E_n\right) \cap \left(\bigcup_{n=1}^\infty B_{j'} \cap E_n\right) = \emptyset$$

whenever $j \neq j'$ (since $B_j \cap B_{j'} = \emptyset$). Therefore we have

$$\begin{aligned}
\int_X \varphi \mathrm{d}\mu &\leq \int_X \phi 1_{\bigcup_{n=1}^\infty E_n} \mathrm{d}\mu \\
&= \sum_{j=1}^m b_j \mu\left(\bigcup_{n=1}^\infty B_j \cap E_n\right) \\
&\leq \sum_{j=1}^m b_j \sum_{n=1}^\infty \mu(B_j \cap E_n) \\
&= \sum_{n=1}^\infty \sum_{j=1}^m b_j \mu(B_j \cap E_n) \\
&= \sum_{n=1}^\infty \sum_{j=1}^m \int_X b_j 1_{B_j \cap E_n} \mathrm{d}\mu \\
&= \sum_{n=1}^\infty \int_X \sum_{j=1}^m b_j 1_{B_j \cap E_n} \mathrm{d}\mu \\
&= \sum_{n=1}^\infty \int_X \sum_{j=1}^m b_j \left(1_{B_j} 1_{E_n}\right) \mathrm{d}\mu \\
&= \sum_{n=1}^\infty \int_X \left(\sum_{j=1}^m b_j 1_{B_j}\right) 1_{E_n} \mathrm{d}\mu \\
&= \sum_{n=1}^\infty \int_X \phi 1_{E_n} \mathrm{d}\mu,
\end{aligned}$$

where we used monotonicity of integration in the first line and where we used countable subadditivity of μ to get from the second line to the third line.

□

37.4 Equivalent Criterion for Function to be Measurable

Proposition 37.5. *Let $f: X \to \mathbb{R}$ be a function. The following are equivalent;*

1. $f^{-1}(-\infty, c) \in \mathcal{M}$ *for all* $c \in \mathbb{R}$.
2. $f^{-1}[c, \infty) \in \mathcal{M}$ *for all* $c \in \mathbb{R}$.
3. $f^{-1}(c, \infty) \in \mathcal{M}$ *for all* $c \in \mathbb{R}$.
4. $f^{-1}(-\infty, c] \in \mathcal{M}$ *for all* $c \in \mathbb{R}$.

Proof.

($1 \implies 2$) Let $c \in \mathbb{R}$. Then

$$\begin{aligned}
\left(f^{-1}[c, \infty)\right)^c &= f^{-1}(-\infty, c) \\
&\in \mathcal{M}.
\end{aligned}$$

It follows that $f^{-1}[c, \infty) \in \mathcal{M}$ for all $c \in \mathbb{R}$.

($2 \implies 3$). Let $c \in \mathbb{R}$. Then

$$\begin{aligned}
f^{-1}(c, \infty) &= \bigcup_{n=1}^\infty f^{-1}[c + 1/n, \infty) \\
&\in \mathcal{M}.
\end{aligned}$$

It follows that $f^{-1}(c, \infty) \in \mathcal{M}$ for all $c \in \mathbb{R}$.

$(3 \implies 4)$. Let $c \in \mathbb{R}$. Then

$$\begin{aligned}\left(f^{-1}(-\infty, c]\right)^c &= f^{-1}(c, \infty) \\ &\in \mathcal{M}.\end{aligned}$$

It follows that $f^{-1}(-\infty, c] \in \mathcal{M}$ for all $c \in \mathbb{R}$.

$(4 \implies 1)$. Let $c \in \mathbb{R}$. Then

$$\begin{aligned}f^{-1}(-\infty, c) &= \bigcap_{n=1}^{\infty} f^{-1}(-\infty, c + 1/n] \\ &\in \mathcal{M}.\end{aligned}$$

It follows that $f^{-1}(-\infty, c) \in \mathcal{M}$ for all $c \in \mathbb{R}$. □

37.5 Simple Functions are Measurable

Proposition 37.6. *Let $\phi \colon X \to \mathbb{R}$ be a simple function. Then $\phi^{-1}(-\infty, c) \in \mathcal{M}$ for all $c \in \mathbb{R}$.*

Proof. Let $c \in \mathbb{R}$ and express ϕ in terms of its canonical representation, say

$$\phi = \sum_{i=1}^{n} a_i 1_{A_i}.$$

Then

$$\begin{aligned}\phi^{-1}(-\infty, c) &= \bigcup_{i \mid a_i < c} A_i \\ &\in \mathcal{M}.\end{aligned}$$

It follows that $\phi^{-1}(-\infty, c) \in \mathcal{M}$ for all $c \in \mathbb{R}$. □

37.6 Pointwise Convergence of Measurable Functions is Measurable

Proposition 37.7. *Let $(f_n \colon X \to \mathbb{R})$ be a sequence of functions which converges pointwise to a function $f \colon X \to \mathbb{R}$. Then*

$$f^{-1}(-\infty, c] = \bigcap_{k=1}^{\infty} \liminf_{n \to \infty} f_n^{-1}(-\infty, c + 1/k).$$

Proof. We have

$$\begin{aligned}x \in \bigcap_{k=1}^{\infty} \liminf_{n \to \infty} f_n^{-1}(-\infty, c + 1/k) &\iff x \in \liminf_{n \to \infty} f_n^{-1}(-\infty, c + 1/k) \text{ for all } k \\ &\iff x \in f_{\pi_k(n)}^{-1}(-\infty, c + 1/k) \text{ for all } k \text{ for some subsequence } (\pi_k(n))_{n \in \mathbb{N}} \text{ of } (n)_{n \in \mathbb{N}} \\ &\iff f_{\pi_k(n)}(x) < c + 1/k \text{ for all } k \text{ for some subsequence } (\pi_k(n))_{n \in \mathbb{N}} \text{ of } (n)_{n \in \mathbb{N}} \\ &\iff f(x) \leq c + 1/k \text{ for all } k \\ &\iff f(x) \leq c \\ &\iff x \in f^{-1}(-\infty, c]\end{aligned}$$

where we obtained the fourth line from the third line since a subsequence of a convergent sequence must converge to the same limit (so $f_{\pi_k(n)}(x) \to f(x)$ as $n \to \infty$). □

Problem 5.c

Proposition 37.8. *Let $(\phi_n : X \to \mathbb{R})$ be a sequence of simple functions which converges pointwise to a function $f : X \to \mathbb{R}$. Then $f^{-1}(-\infty, c) \in \mathcal{M}$ for all $c \in \mathbb{R}$.*

Proof. Let $c \in \mathbb{R}$. Then by 5.a and 5.b, and the fact that $\mathcal{M}$ is closed under taking intersections and liminf, we have

$$\begin{aligned} f^{-1}(-\infty, c] &= \bigcap_{k=1}^{\infty} \liminf_{n\to\infty} \phi_n^{-1}(-\infty, c + 1/k) \\ &\in \mathcal{M}. \end{aligned}$$

Thus $f^{-1}(-\infty, c] \in \mathcal{M}$ for all $c \in \mathbb{R}$. It follows from Proposition (37.5) that $f^{-1}(-\infty, c) \in \mathcal{M}$ for all $c \in \mathbb{R}$. □

37.7 Max, Min, Addition, and Scalar Multiplication of Measurable Functions is Measurable

Proposition 37.9. *Let $f, g : X \to [0, \infty]$ be two nonnegative measurable functions and let $a \geq 0$. Then af, $f + g$, $\max\{f, g\}$, and $\min\{f, g\}$ are all measurable functions.*

Proof. Choose an increasing sequence of nonnegative simple functions (φ_n) which converges pointwise to f and choose an increasing sequence of nonnegative simple functions (ψ_n) which converges pointwise to g. For each $x \in X$, we have

$$\begin{aligned} \lim_{n\to\infty}(c\varphi_n(x)) &= c \lim_{n\to\infty} \varphi_n(x) \\ &= cf(x), \end{aligned}$$

and

$$\begin{aligned} \lim_{n\to\infty}(\varphi_n(x) + \psi_n(x)) &= \lim_{n\to\infty} \varphi_n(x) + \lim_{n\to\infty} \psi_n(x) \\ &= f(x) + g(x), \end{aligned}$$

and

$$\begin{aligned} \lim_{n\to\infty}(\varphi_n(x)\psi_n(x)) &= \lim_{n\to\infty} \varphi_n(x) \lim_{n\to\infty} \psi_n(x) \\ &= f(x)g(x). \end{aligned}$$

It follows that $(c\varphi_n)$, $(\varphi_n + \psi_n)$, and $(\varphi_n\psi_n)$ are increasing functions of simple functions which converges pointwise to cf, $f + g$, and fg respectively. Therefore cf, $f + g$, and fg are measurable functions.

It remains to show that $\max\{f, g\}$ and $\min\{f, g\}$ are measurable. First note that $\max\{\varphi, \psi\}$ and $\min\{\varphi, \psi\}$ are simple functions for any two simple functions φ and ψ. Indeed, if

$$\varphi = \sum_{i=1}^{m} a_i 1_{A_i} \quad \text{and} \quad \psi = \sum_{j=1}^{n} b_j 1_{B_j},$$

then

$$\max\{\varphi, \psi\} = \sum_{\substack{1\leq i\leq m \\ 1\leq j\leq n}} \max\{a_i, b_j\} 1_{A_i \cap B_j} \quad \text{and} \quad \min\{\varphi, \psi\} = \sum_{\substack{1\leq i\leq m \\ 1\leq j\leq n}} \min\{a_i, b_j\} 1_{A_i \cap B_j}$$

Thus $(\max\{\varphi_n, \psi_n\})$ and $(\min\{\varphi_n, \psi_n\})$ are both sequences of simple functions. They are also increasing sequences since both (φ_n) and (ψ_n) are increasing. We will show that they converge to $\max\{f, g\}$ and $\min\{f, g\}$ respectively.

Let $x \in X$ be arbitrary and assume without loss of generality that $f(x) \geq g(x)$. If $f(x) = g(x)$ then both $\varphi_n(x) \to f(x)$ and $\psi_n(x) \to f(x)$, and so clearly $\max\{\varphi_n(x), \psi_n(x)\} \to f(x)$ and $\min\{\varphi_n(x), \psi_n(x)\} \to f(x)$, so assume $f(x) > g(x)$. Set $\varepsilon = f(x) - g(x)$ and choose $N \in \mathbb{N}$ such that $n \geq N$ implies

$$|f(x) - \varphi_n(x)| < \frac{\varepsilon}{2} \quad \text{and} \quad |g(x) - \psi_n(x)| < \frac{\varepsilon}{2}.$$

Then $n \geq N$ implies $\varphi_n(x) > \psi_n(x)$. Therefore $n \geq N$ implies

$$\begin{aligned} |\max\{f(x), g(x)\} - \max\{\varphi_n(x), \psi_n(x)\}| &= |f(x) - \varphi_n(x)| \\ &< \frac{\varepsilon}{2} \\ &< \varepsilon \end{aligned}$$

and $n \geq N$ implies

$$\begin{aligned} |\min\{f(x), g(x)\} - \min\{\varphi_n(x), \psi_n(x)\}| &= |g(x) - \psi_n(x)| \\ &< \frac{\varepsilon}{2} \\ &< \varepsilon. \end{aligned}$$

Therefore $\max\{\varphi_n(x), \psi_n(x)\} \to f(x)$ and $\min\{\varphi_n(x), \psi_n(x)\} \to g(x)$. Since x was arbitrary, it follows that the sequences $(\max\{\varphi_n, \psi_n\})$ and $(\min\{\varphi_n, \psi_n\})$ converges pointwise to $\max\{f, g\}$ and $\min\{f, g\}$ respectively. □

38 Homework 5

38.1 Criterion for Function to be Measurable Using Rational Numbers

Proposition 38.1. *Let $f\colon X \to \mathbb{R}$ be a function. Then f is measurable if and only if for every $q \in \mathbb{Q}$ the set $f^{-1}(-\infty, q)$ is measurable.*

Proof. If f is measurable, then certainly $f^{-1}(-\infty, c) \in \mathcal{M}$ for any $c \in \mathbb{R}$ (and hence for any $c \in \mathbb{Q}$). Conversely, suppose $f^{-1}(-\infty, q) \in \mathcal{M}$ for any $q \in \mathbb{Q}$. Let $c \in \mathbb{R}$. For each $n \in \mathbb{N}$ choose $q_n \in \mathbb{Q}$ such that

$$c < q_n < c + \frac{1}{n}.$$

Such a choice for each n can be made since $\mathbb{Q}$ is dense in $\mathbb{R}$. We claim that

$$f^{-1}(-\infty, c] = \bigcap_{n=1}^{\infty} f^{-1}(-\infty, q_n)$$

To see this, first note that the inclusion

$$f^{-1}(-\infty, c] \subseteq \bigcap_{n=1}^{\infty} f^{-1}(-\infty, q_n)$$

is clear since each $f^{-1}(-\infty, q_n)$ contains $f^{-1}(-\infty, c)$ (as $c < q_n$). For the reverse inclusion, suppose $x \in \bigcap_{n=1}^{\infty} f^{-1}(-\infty, q_n)$, so $f(x) < q_n$ for all n. Since $q_n \to c$, this implies $f(x) \leq c$. Thus $x \in f^{-1}(-\infty, c]$. It follows that f is measurable. □

Remark 55. Note that we needed to use the fact that $\mathbb{Q}$ is dense in $\mathbb{R}$ in order to prove this.

38.2 Alternative Definition for Measurability of Function

Before we answer this problem, we give a more general definition of what it means for a function to be measurable with respect to σ-algebras $\mathcal{M}$ and $\mathcal{N}$. Then we show that this more general definition is equivalent to the definition we've been using when $\mathcal{N}$ is the Borel σ-algebra on $\mathbb{R}$.

Definition 38.1. Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces and let $f\colon X \to Y$ be a function. We say f is **measurable with respect to** $\mathcal{M}$ **and** $\mathcal{N}$ if $f^{-1}(\mathcal{N}) \subseteq \mathcal{M}$ where

$$f^{-1}(\mathcal{N}) = \{f^{-1}(B) \mid B \in \mathcal{N}\}.$$

In other words, f is measurable with respect to $\mathcal{M}$ and $\mathcal{N}$ if for all $B \in \mathcal{N}$ we have

$$f^{-1}(B) = \{x \in X \mid f(x) \in B\} \in \mathcal{M}.$$

If $\mathcal{M} = \mathcal{N}$, then we will just say f is measurable with respect to $\mathcal{M}$. If the σ-algebras $\mathcal{M}$ and $\mathcal{N}$ are clear from context, then we will just say f is measurable.

Let us now show that when $Y = \mathbb{R}$ and $\mathcal{N} = \mathcal{B}(\mathbb{R})$, that this definition is equivalent to the definition we gave in class. We first prove the following two propositions:

Proposition 38.2. *Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces and let $f \colon X \to Y$ be a function. Suppose that $\mathcal{N}$ is generated as a σ-algebra by the collection $\mathcal{C}$ of subsets of Y. Then $f^{-1}(\mathcal{N}) \subseteq \mathcal{M}$ if and only if $f^{-1}(\mathcal{C}) \subseteq \mathcal{M}$.*

Proof. One direction is clear, so we just prove the other direction. Suppose $f^{-1}(\mathcal{C}) \subseteq \mathcal{M}$. Observe that

$$\{B \in \mathcal{P}(Y) \mid f^{-1}(B) \in \mathcal{M}\}$$

is a σ-algebra which contains $\mathcal{C}$. Indeed, it is a σ-algebra since f^{-1} maps the emptyset set to the emptyset and maps the whole space Y to the whole space X, and since f^{-1} commutes with unions and complements. Furthemore, this σ-algebra contains $\mathcal{C}$ since $f^{-1}(\mathcal{C}) \subseteq \mathcal{M}$. Since $\mathcal{N}$ is the *smallest* σ-algebra which contains $\mathcal{C}$, it follows that

$$\mathcal{N} \subseteq \{B \in \mathcal{P}(Y) \mid f^{-1}(B) \in \mathcal{M}\}.$$

In particular, if $B \in \mathcal{N}$, then $f^{-1}(B) \in \mathcal{M}$. Thus f is measurable. □

Proposition 38.3. *Let $\mathcal{C} = \{(-\infty, c) \mid c \in \mathbb{R}\}$. Then $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$.*

Proof. Let $\mathcal{I}_n$ be the collection of all subintervals of $[n, n+1)$ and let $\mathcal{B}_n = \sigma(\mathcal{I}_n)$. So

$$\mathcal{B}(\mathbb{R}) = \{E \subseteq \mathbb{R} \mid E \cap [n, n+1) \in \mathcal{B}_n \text{ for all } n \in \mathbb{Z}\}.$$

Let $c \in \mathbb{R}$. Then since $(-\infty, c) \cap [n, n+1]$ is a subinterval of $[n, n+1]$ for all $n \in \mathbb{Z}$, it follows that $(-\infty, c) \in \mathcal{B}$ for all $n \in \mathbb{Z}$. Thus $\mathcal{C} \subseteq \mathcal{B}$ which implies $\sigma(\mathcal{C}) \subseteq \mathcal{B}$ (as $\sigma(\mathcal{C})$ is the *smallest* σ-algebra which contains $\mathcal{C}$). Conversely, note that $\sigma(\mathcal{C})$ contains all subintervals of $[n, n+1)$ for all $n \in \mathbb{Z}$. Thus $\sigma(\mathcal{C}) \supseteq \mathcal{B}_n$ for all $n \in \mathbb{Z}$ (as $\mathcal{B}_n$ is the *smallest* σ-algebra which contains all subintervals of $[n, n+1)$. Since $\mathcal{B}(\mathbb{R}) = \sigma\left(\bigcup_{n \in \mathbb{Z}} \mathcal{B}_n\right)$, it follows that $\mathcal{B}(\mathbb{R}) \subseteq \sigma(\mathcal{C})$. □

Corollary 15. *Let $(X, \mathcal{M})$ be a measurable space and let $f \colon X \to \mathbb{R}$ be a function. Then f is measurable with respect to $\mathcal{M}$ and $\mathcal{B}(\mathbb{R})$ if and only if $f^{-1}(-\infty, c) \in \mathcal{M}$ for all $c \in \mathbb{R}$.*

Proof. Follows from Proposition (38.3) and Proposition (38.2). □

Corollary 16. *Let $(X, \mathcal{M})$ be a measurable space and let $\mathcal{B}(\mathbb{R})$ be the Borel σ-algebra on $\mathbb{R}$. Suppose that $\mathcal{C} \subseteq \mathcal{B}(\mathbb{R})$ is a collection of sets in $\mathcal{B}(\mathbb{R})$ such that $\mathcal{B}(\mathbb{R}) \subseteq \sigma(\mathcal{C})$. Then $f \colon X \to \mathbb{R}$ is measurable with respect to $\mathcal{M}$ and $\mathcal{B}(\mathbb{R})$ if and only if $f^{-1}(C) \in \mathcal{M}$ for all $C \in \mathcal{C}$.*

Proof. Follows from Proposition (38.2) and from Corollary (15). □

38.3 Monotone Increasing Function and Continuous Function are Borel Measurable

Problem 3.i

Proposition 38.4. *Let $f \colon \mathbb{R} \to \mathbb{R}$ be a continuous function. Then f is measurable with respect to $\mathcal{B}(\mathbb{R})$.*

Proof. For each $q, r \in \mathbb{Q}$ and $n \in \mathbb{N}$, let

$$\mathrm{B}_{1/n}(q) = \{x \in \mathbb{R} \mid |x - q| < 1/n\}$$

Then the collection

$$\mathscr{B} = \{\mathrm{B}_{1/n}(q) \mid n \in \mathbb{N} \text{ and } q \in \mathbb{Q}\}$$

forms a countable basis for the usual topology on $\mathbb{R}$. In particular, if U be an open subset of $\mathbb{R}$, then we can express U as a union of the form

$$U = \bigcup_{\lambda \in \Lambda} B_\lambda$$

where $B_\lambda \in \mathscr{B}$ and where the index set Λ is *countable*. In particular, it follows that $\tau(\mathscr{B}) \subseteq \mathcal{B}(\mathbb{R})$, where $\tau(\mathscr{B})$ is the usual Euclidean topology on $\mathbb{R}$. Thus since $(-\infty, c)$ is an open subset of $\mathbb{R}$ for any $c \in \mathbb{R}$, it follows that $f^{-1}(-\infty, c)$ can be expressed a countable union of open subsets of $\mathbb{R}$ (by definition of what it means to be continuous, the inverse image of an open set under f is open). Since every open subset of $\mathbb{R}$ is Borel measurable, it follows that $f^{-1}(-\infty, c) \in \mathcal{B}(\mathbb{R})$. Thus f is measurable with respect to $\mathcal{B}(\mathbb{R})$. □

Problem 3.ii

Proposition 38.5. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be a monotone increasing function. Then f is $\mathcal{B}(\mathbb{R})$-measurable.*

Proof. Let $c \in \mathbb{R}$. We want to show that $f^{-1}(-\infty, c) \in \mathcal{B}(\mathbb{R})$. If $f^{-1}(-\infty, c) = \emptyset$ or $f^{-1}(-\infty, c) = \mathbb{R}$, then we are done, so assume $f^{-1}(-\infty, c) \neq \emptyset$ and $f^{-1}(-\infty, c) \neq \mathbb{R}$. Choose $y \in \mathbb{R}$ such that $c \leq f(y)$. Observe that if $x \in f^{-1}(-\infty, c)$, then

$$f(x) < c \leq f(y),$$

which implies $x \leq y$ since f is monotone increasing. Thus y is an upper bound of the set $f^{-1}(-\infty, c)$. Since $f^{-1}(-\infty, c)$ is nonempty and bounded above, it follows that its supremum exists. Denote its supremum by y_0. So

$$y_0 = \sup\{x \in \mathbb{R} \mid f(x) < c\}.$$

We claim that

$$f^{-1}(-\infty, c) = \begin{cases} (-\infty, y_0) & \text{if } f(y_0) \geq c \\ (-\infty, y_0] & \text{if } f(y_0) < c. \end{cases}$$

Indeed, since y_0 is an upper bound $f^{-1}(-\infty, c)$, it must be greater than or equal to all elements in $f^{-1}(-\infty, c)$. In other words, if $x \in f^{-1}(-\infty, c)$, then $x \leq y_0$. Thus

$$f^{-1}(-\infty, c) \subseteq \begin{cases} (-\infty, y_0) & \text{if } f(y_0) \geq c \\ (-\infty, y_0] & \text{if } f(y_0) < c. \end{cases}$$

Conversely, suppose $x \in (-\infty, y_0)$, so $x < y_0$. Then x is not an upper bound of the set $f^{-1}(-\infty, c)$ (since y_0 is the *least* upper bound), which means that there exists an $x' \in \mathbb{R}$ such that $x \leq x'$ and $f(x') < c$. But since $f(x) \leq f(x')$, this implies $f(x) < c$, and hence $x \in f^{-1}(-\infty, c)$. Thus

$$f^{-1}(-\infty, c) \supseteq \begin{cases} (-\infty, y_0) & \text{if } f(y_0) \geq c \\ (-\infty, y_0] & \text{if } f(y_0) < c. \end{cases}$$

In any case, we see that $f^{-1}(-\infty, c) \in \mathcal{B}(\mathbb{R})$. □

38.4 Sum of Nonnegative Measurable Functions Commutes with Integral

Proposition 38.6. *Let $(X, \mathcal{M}, \mu)$ be a measure space. Suppose $(f_n\colon X \to [0, \infty])$ is a sequence of nonnegative measurable functions. Define $f\colon X \to [0, \infty]$ by*

$$f(x) = \sum_{n=1}^{\infty} f_n(x)$$

for all $x \in X$. Then f is a nonnegative measurable function and

$$\int_X f \mathrm{d}\mu = \sum_{n=1}^{\infty} \int_X f_n \mathrm{d}\mu.$$

Proof. For each $N \in \mathbb{N}$, let $s_N = \sum_{n=1}^{N} f_n$. Then s_N converges pointwise to f since

$$\begin{aligned} f(x) &= \sum_{n=1}^{\infty} f_n(x) \\ &= \lim_{N \to \infty} \sum_{n=1}^{N} f_n(x) \\ &= \lim_{N \to \infty} s_N(x) \end{aligned}$$

for all $x \in X$. Each s_N is a nonnegative measurable function since it is a finite sum of nonnegative measurable functions, and so (s_N) is a sequence of nonnegative functions which converges pointwise to f. This implies f is a nonnegative measurable function. Furthermore, s_N is an increasing sequence since if $M \leq N$, then

$$\begin{aligned} s_M(x) &= \sum_{n=1}^{M} f_n(x) \\ &\leq \sum_{n=1}^{N} f_n(x) \\ &= s_N(x) \end{aligned}$$

for all $x \in X$, where the inequality follows from the fact that each f_n is nonnegative. Therefore we may apply the Monotone Convergence Theorem to obtain

$$\begin{aligned} \int_X f \mathrm{d}\mu &= \lim_{N\to\infty} \int_X s_N \mathrm{d}\mu \\ &= \lim_{N\to\infty} \int_X \sum_{n=1}^{N} f_n \mathrm{d}\mu \\ &= \lim_{N\to\infty} \sum_{n=1}^{N} \int_X f_n \mathrm{d}\mu \\ &= \sum_{n=1}^{\infty} \int_X f_n \mathrm{d}\mu, \end{aligned}$$

where we obtained the third line from the fourth line from the fact that this is a finite sum. □

38.5 Defining Measure From Integral Weighted by Nonnegative Measurable Function

Proposition 38.7. *Let $(X, \mathcal{M}, \mu)$ be measure space and let $g\colon X \to [0, \infty)$ be a nonnegative measurable function. Define $\nu_g\colon \mathcal{M} \to [0, \infty]$ by*

$$\nu_g(E) = \int_X g 1_E \mathrm{d}\mu$$

for all $E \in \mathcal{M}$. Then ν is a measure on $(X, \mathcal{M})$.

Proof. First note that

$$\begin{aligned} \nu_g(\emptyset) &= \int_X g 1_\emptyset \mathrm{d}\mu \\ &= \int_X g \cdot 0 \cdot \mathrm{d}\mu \\ &= \int_X 0 \cdot \mathrm{d}\mu \\ &= 0. \end{aligned}$$

Next we show that ν_g is finitely additive. Let $(E_n)_{n=1}^N$ be a finite sequence of pairwise disjoint sets in $\mathcal{M}$. Then

$$\begin{aligned} \nu_g\left(\bigcup_{n=1}^{N} E_n\right) &= \int_X g 1_{\bigcup_{n=1}^N E_n} \mathrm{d}\mu \\ &= \int_X g \sum_{n=1}^{N} 1_{E_n} \mathrm{d}\mu \\ &= \int_X \sum_{n=1}^{N} g 1_{E_n} \mathrm{d}\mu \\ &= \sum_{n=1}^{N} \int_X g 1_{E_n} \mathrm{d}\mu \\ &= \sum_{n=1}^{N} \nu_g(E_n), \end{aligned}$$

where we used the fact that each $g1_{E_n}$ is a nonnegative measurable function in order to commute the finite sum with the integral. Thus it follows that ν_g is finitely additive.

It remains to show that ν_g is countably subadditive. In problem 3 in HW4, we showed that for any nonnegative simple function $\varphi\colon X \to [0,\infty)$, the function $\nu_\varphi\colon \mathcal{M} \to [0,\infty]$ defined by

$$\nu_\varphi(E) = \int_X \varphi 1_E \mathrm{d}\mu$$

is a measure. We will use this fact in what follows:

Choose an increasing sequence $(\varphi_n\colon X \to [0,\infty])$ of nonnegative simple functions which converges pointwise to g (we can do this since g is a nonnegative measurable function). Then $(\varphi_n 1_E)$ is an increasing sequence of nonnegative simple functions which converges pointwise to $g1_E$ for all $E \in \mathcal{M}$. It follows from the Monotone Convergence Theorem that

$$\begin{aligned} \nu_{\varphi_n}(E) &= \int_X \varphi_n 1_E \mathrm{d}\mu \\ &\to \int_X g1_E \mathrm{d}\mu \\ &= \nu_g(E) \end{aligned}$$

for any $E \in \mathcal{M}$. In particular, for any $E \in \mathcal{M}$ and $\varepsilon > 0$, we can find a $N_{E,\varepsilon} \in \mathbb{N}$ (which depends on E and ε) such that

$$\nu_g(E) < \nu_{\varphi_n}(E) + \varepsilon \tag{91}$$

for all $n \geq N_{E,\varepsilon}$. However we will don't need estimate $\nu_g(E)$ to this level of precision. We just need to know thta for any $\varepsilon > 0$, we can find an $n \in \mathbb{N}$ such that (91) holds.

Now let (E_k) be a sequence of measurable sets and let $\varepsilon > 0$. Choose $n \in \mathbb{N}$ such that

$$\nu_g\left(\bigcup_{k=1}^\infty E_k\right) < \nu_{\varphi_n}\left(\bigcup_{k=1}^\infty E_k\right) + \varepsilon$$

Then we have

$$\begin{aligned} \nu_g\left(\bigcup_{k=1}^\infty E_k\right) &\leq \nu_{\varphi_n}\left(\bigcup_{k=1}^\infty E_k\right) + \varepsilon \\ &\leq \sum_{k=1}^\infty \nu_{\varphi_n}(E_k) + \varepsilon \\ &\leq \sum_{k=1}^\infty \nu_g(E_k) + \varepsilon \end{aligned}$$

where we obtained the second line from the first line using countable subadditivity of ν_{φ_n}, and where we obtained the third line from the second line from the fact that $\varphi_n 1_{E_k} \leq g1_{E_k}$ for all k and from monotonicity of integration. Taking $\varepsilon \to 0$ gives us countable subadditivity of ν_g. □

38.6 Decreasing Version of MCT

Proposition 38.8. *Let $(X, \mathcal{M}, \mu)$ be a measure space. Suppose $(f_n\colon X \to [0,\infty])$ is a decreasing sequence of nonnegative measurable functions which converges pointwise to $f\colon X \to [0,\infty]$. If $\int_X f_1 \mathrm{d}\mu < \infty$, then*

$$\lim_{n\to\infty} \int_X f_n \mathrm{d}\mu = \int_X f \mathrm{d}\mu. \tag{92}$$

Proof. For each $n \in \mathbb{N}$, set $g_n = f_n - f_{n+1}$. Since (f_n) is a decreasing sequence, each g_n is nonnegative. Furthermore each g_n is measurable since it is a difference of two measurable functions. Set $g = \sum_{n=1}^\infty g_n$ and observe that

$$\begin{aligned}
g &= \sum_{n=1}^{\infty} g_n \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} g_n \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} (f_n - f_{n+1}) \\
&= \lim_{N\to\infty} (f_1 - f_{N+1}) \\
&= f_1 - f.
\end{aligned}$$

It follows from the monotone convergence theorem that

$$\begin{aligned}
\int_X f_1 \mathrm{d}\mu - \int_X f \mathrm{d}\mu &= \int_X (f_1 - f)\mathrm{d}\mu \\
&= \int_X g \mathrm{d}\mu \\
&= \sum_{n=1}^{\infty} \int_X g_n \mathrm{d}\mu && \text{MCT} \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \int_X g_n \mathrm{d}\mu \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \int_X (f_n - f_{n+1})\mathrm{d}\mu \\
&= \lim_{N\to\infty} \int_X \sum_{n=1}^{N} (f_n - f_{n+1})\mathrm{d}\mu \\
&= \lim_{N\to\infty} \int_X (f_1 - f_{N+1})\mathrm{d}\mu \\
&= \int_X f_1 \mathrm{d}\mu - \lim_{N\to\infty} \int_X f_{N+1} \mathrm{d}\mu.
\end{aligned}$$

Since $\int_X f_1 \mathrm{d}\mu < 0$, we can cancel it from both sides to get (92). □

38.7 Generalized Fatou's Lemma

Proposition 38.9. *Fatou's Lemma remains valid if the hypothesis that all $f_n \colon X \to [0,\infty]$ are nonnegative measurable functions is replaced by the hypothesis that $f_n \colon X \to \mathbb{R}$ are measurable and there exists a nonnegative integrable function $g \colon X \to [0,\infty]$ such that $-g \le f_n$ pointwise for all $n \in \mathbb{N}$.*

Proof. Observe that $(g + f_n)$ is a sequence of nonnegative measurable functions which converges pointwise to the nonnegative measurable function $g+f$. Then it follows from Fatou's Lemma that

$$\begin{aligned}
\int_X g\mathrm{d}\mu + \int_X f\mathrm{d}\mu &= \int_X g\mathrm{d}\mu + \int_X ((g+f)-g)\mathrm{d}\mu \\
&= \int_X g\mathrm{d}\mu + \int_X (g+f)\mathrm{d}\mu - \int_X g\mathrm{d}\mu \\
&= \int_X (g+f)\mathrm{d}\mu \\
&\leq \liminf_{n\to\infty} \int_X (g+f_n)\mathrm{d}\mu \\
&= \liminf_{n\to\infty} (\int_X g\mathrm{d}\mu + \int_X (g+f_n)\mathrm{d}\mu - \int_X g\mathrm{d}\mu) \\
&= \liminf_{n\to\infty} (\int_X g\mathrm{d}\mu + \int_X ((g+f_n)-g)\mathrm{d}\mu) \\
&= \liminf_{n\to\infty} (\int_X g\mathrm{d}\mu + \int_X f_n\mathrm{d}\mu) \\
&= \int_X g\mathrm{d}\mu + \liminf_{n\to\infty} \int_X f_n\mathrm{d}\mu.
\end{aligned}$$

Since $\int_X g\mathrm{d}\mu < \infty$, we can cancel it from both sides to obtain

$$\int_X f\mathrm{d}\mu \leq \liminf_{n\to\infty} \int_X f_n\mathrm{d}\mu.$$

□

38.8 Integral Computations (Using DCT and Decreasing MCT)

Exercise 2. Compute the following integrals

$$\lim_{n\to\infty} \int_0^\infty \frac{\sin(x/n)}{(1+x/n)^n}\mathrm{d}x \tag{93}$$

$$\lim_{n\to\infty} \int_0^1 \frac{1+nx^2}{(1+x^2)^n}\mathrm{d}x \tag{94}$$

Solution 1. We first compute (93). For each $n \in \mathbb{N}$, let $f_n = \sin(x/n)(1+x/n)^{-n}$. Observe that each f_n is measurable since each f_n is continuous (the denominator is only zero when $x=-n$). Let us check that each f_n is integrable: we have

$$\begin{aligned}
\int_0^\infty |f_n|\mathrm{d}x &= \int_0^\infty \left|\frac{\sin(x/n)}{(1+x/n)^n}\right| \\
&\leq \int_0^\infty \left|(1+x/n)^{-n}\right| \mathrm{d}x \\
&= \int_0^\infty (1+x/n)^{-n}\mathrm{d}x \\
&\leq \int_0^\infty e^{-x}\mathrm{d}x \\
&= 1.
\end{aligned}$$

for all $n \in \mathbb{N}$. Thus each f_n is integrable.

Next we observe that f_n converges pointwise to 0 since $(1+x/n)^n \to e^x$ and $\sin(x/n) \to 0$ as $n \to \infty$ for all $x \in \mathbb{R}$. Finally, note that $e^{-x} \geq |f_n|$ pointwise and e^{-x} is integrable ($\int_0^\infty |e^{-x}|\mathrm{d}x = 1$). It follows from the Lebesgue Dominated Convergence Theorem that

$$\lim_{n\to\infty} \int_0^\infty \frac{\sin(x/n)}{(1+x/n)^n}\mathrm{d}x = \int_0^\infty 0\mathrm{d}x = 0.$$

Next we compute (94). For each $n \in \mathbb{N}$, let $f_n = (1+nx^2)(1+x^2)^{-n}$. Observe that each f_n is measurable since each f_n is continuous (the denominator is never zero for any $x \in \mathbb{R}$). Furthermore, each f_n is nonnegative (since taking squares makes everything nonnegative). We claim that f_n is a decreasing sequence. Indeed

$$\begin{aligned}\frac{f_n}{f_{n+1}} &= \left(\frac{1+nx^2}{(1+x^2)^n}\right)\left(\frac{(1+x^2)^{n+1}}{1+(n+1)x^2}\right) \\ &= \frac{(1+nx^2)(1+x^2)}{1+(n+1)x^2} \\ &= \frac{nx^4+(n+1)x^2+1}{(n+1)x^2+1} \\ &\geq \frac{(n+1)x^2+1}{(n+1)x^2+1} \\ &= 1.\end{aligned}$$

Thus (f_n) is a decreasing sequence which is bounded below, so it must converge pointwise to some function. For $x=0$, it's easy to see that $f_n(0) \to 0$. For $x \neq 0$, we use L'Hopital's rule to get

$$\begin{aligned}\lim_{n\to\infty} f_n &= \lim_{n\to\infty} \frac{1+nx^2}{(1+x^2)^n} \\ &= \lim_{n\to\infty} \frac{x^2}{\ln(1+x^2)(1+x^2)^n} \\ &= 0.\end{aligned}$$

Thus (f_n) converges pointwise to 0. Since

$$\begin{aligned}\int_0^1 f_1 \mathrm{d}x &= \int_0^1 \frac{1+x^2}{1+x^2}\mathrm{d}x \\ &= \int_0^1 \mathrm{d}x \\ &= 1 \\ &< \infty,\end{aligned}$$

it follows from problem 6 that

$$\begin{aligned}\lim_{n\to\infty} \int_0^1 \frac{1+nx^2}{(1+x^2)^n}\mathrm{d}x &= \lim_{n\to\infty}\int_0^1 f_n \mathrm{d}x \\ &= \int_0^1 0\mathrm{d}x \\ &= 0.\end{aligned}$$

39 Homework 6

39.1 Necessary and Sufficient Condition For Integrable Function Being Zero Almost Everywhere

Proposition 39.1. *Let $f \in L^1(X, \mathcal{M}, \mu)$ and suppose that $\int_X f 1_E \mathrm{d}\mu = 0$ for every $E \in \mathcal{M}$. Then $f = 0$ almost everywhere.*

Proof. Let $A^+ = \{f^+ \neq 0\}$ and $A^- = \{f^- \neq 0\}$. Then A^+ and A^- are measurable sets since f^+ and f^- are measurable functions. Since f agrees with f^+ on A^+, we have

$$\begin{aligned}\int_X f^+ \mathrm{d}\mu &= \int_X f^+ 1_{A^+}\mathrm{d}\mu \\ &= \int_X f 1_{A^+}\mathrm{d}\mu \\ &= 0.\end{aligned}$$

Similarly, since $-f$ agrees with f^- on A^-, we have

$$\begin{aligned}\int_X f^- \mathrm{d}\mu &= \int_X f^- 1_{A^-} \mathrm{d}\mu \\ &= \int_X -f 1_{A^-} \mathrm{d}\mu \\ &= -\int_X f 1_{A^-} \mathrm{d}\mu \\ &= 0.\end{aligned}$$

It follows that

$$\begin{aligned}\int_X |f| \mathrm{d}\mu &= \int_X (f^+ + f^-) \mathrm{d}\mu \\ &= \int_X f^+ \mathrm{d}\mu + \int_X f^- \mathrm{d}\mu \\ &= 0.\end{aligned}$$

Thus $f = 0$ almost everywhere (by a proposition proved in class). $\square$

39.2 Integrable Function Takes Value Infinity on a Set of Measure Zero

Proposition 39.2. *Let $f\colon X \to [0, \infty]$ be a nonnegative measurable function such that $\int_X f \mathrm{d}\mu < \infty$. Then*

1. $\mu(\{f = \infty\}) = 0$;
2. *f does not need to be bounded almost everywhere.*

Proof. 1. Assume for a contradiction that $\mu(\{f = \infty\}) > 0$. Then for any $M \in \mathbb{N}$, we have

$$M 1_{\{f=\infty\}} \leq M f.$$

Therefore

$$\begin{aligned}\infty &> \int_X f \mathrm{d}\mu \\ &\geq \int_X M 1_{\{f=\infty\}} \mathrm{d}\mu \\ &= M\mu(\{f = \infty\}).\end{aligned}$$

Taking $M \to \infty$ gives us a contradiction.

2. To see that f does not need to be bounded, consider $X = [0, 1]$ and $f(x) = x^{-1/2}$. Then

$$\int_0^1 x^{-1/2} \mathrm{d}x = 2,$$

but f is not bounded almost everywhere. Indeed, for any $M \in \mathbb{N}$, the set $[0, 1/M^2]$ has nonzero measure and $f|_{[0,1/M^2]} \geq M$. $\square$

39.3 $L^p(X, \mathcal{M}, \mu)$ is the Completion of Space of Simple Functions

Problem 3.a

Lemma 39.1. *Let (X, d) be a metric space and let (x_n) be a Cauchy sequence in X. Suppose there exists a subsequence $(x_{\pi(n)})$ of the sequence (x_n) such that $x_{\pi(n)} \to x$ for some $x \in X$. Then $x_n \to x$.*

Proof. Let $\varepsilon > 0$. Since $(x_{\pi(n)})$ is convergent, there exists an $N \in \mathbb{N}$ such that $\pi(n) \geq N$ implies

$$\mathrm{d}(x_{\pi(n)}, x) < \frac{\varepsilon}{2}.$$

Since (x_n) is Cauchy, there exists $M \in \mathbb{N}$ such that $m, n \geq M$ implies

$$d(x_m, x_n) < \frac{\varepsilon}{2}.$$

Choose such M and N and assume without loss of generality that $N \geq M$. Then $n \geq N$ implies

$$\begin{aligned} d(x_n, x) &\leq d(x_{\pi(n)}, x_n) + d(x_{\pi(n)}, x) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon. \end{aligned}$$

It follows that $x_n \to x$. □

Lemma 39.2. *Let X be a normed linear space. Then $\mathcal{X}$ is a Banach space if and only if every absolutely convergent series in $\mathcal{X}$ is convergent.*

Proof. Suppose first that every absolutely convergent series in $\mathcal{X}$ is convergent. Let (x_n) be a Cauchy sequence in $\mathcal{X}$. To show that (x_n) is convergent, it suffices to show that a subsequence of (x_n) is convergent, by Lemma (39.1). Choose a subsequence $(x_{\pi(n)})$ of (x_n) such that

$$\|x_{\pi(n)} - x_{\pi(n-1)}\| < \frac{1}{2^n}$$

and for all $n \in \mathbb{N}$ (we can do this since (x_n) is Cauchy). Then the series $\sum_{n=1}^{\infty}(x_{\pi(n)} - x_{\pi(n-1)})$ is absolutely convergent since

$$\begin{aligned} \sum_{n=1}^{\infty} \|x_{\pi(n)} - x_{\pi(n-1)}\| &< \sum_{n=1}^{\infty} \frac{1}{2^n} \\ &= 1. \end{aligned}$$

Therefore it must be convergent, say $\sum_{n=1}^{\infty}(x_{\pi(n)} - x_{\pi(n-1)}) \to x$. On the other hand, for each $n \in \mathbb{N}$, we have

$$x_{\pi(n)} - x_{\pi(1)} = \sum_{m=1}^{n} (x_{\pi(m)} - x_{\pi(1)}).$$

In particular, $x_{\pi(n)} \to x - x_{\pi(1)}$ as $n \to \infty$. Thus $(x_{\pi(n)})$ is a convergent subsequence of (x_n).

Conversely, suppose $\mathcal{X}$ is a Banach space and suppose $\sum_{n=1}^{\infty} x_n$ is absolutely convergent. Let $\varepsilon > 0$ and choose $K \in \mathbb{N}$ such that $N \geq M \geq K$ implies

$$\sum_{n=M}^{N} \|x_n\| < \varepsilon.$$

Then $N \geq M \geq K$ implies

$$\begin{aligned} \left\| \sum_{n=1}^{N} x_n - \sum_{n=1}^{M} x_n \right\| &= \left\| \sum_{n=M}^{N} x_n \right\| \\ &\leq \sum_{n=M}^{N} \|x_n\| \\ &< \varepsilon. \end{aligned}$$

It follows that the sequence of partial sums $(\sum_{n=1}^{N} x_n)_N$ is Cauchy. Since $\mathcal{X}$ is a Banach space, it follows that $\sum_{n=1}^{\infty} x_n$ is convergent. □

Proposition 39.3. *Let $1 < p < \infty$. Then $L^p(X, \mathcal{M}, \mu)$ is a Banach space.*

Proof. By Lemma (39.2), it suffices to show that every absolutely convergent series in $L^p(X, \mathcal{M}, \mu)$ is convergent. Suppose (f_n) is a sequence in $L^p(X, \mathcal{M}, \mu)$ such that $\sum_{n=1}^{\infty} \|f_n\|_p < \infty$. For each $N \in \mathbb{N}$, set $s_N = (\sum_{n=1}^{N} f_n)$. We want to show that (s_N) is convergent in $L^p(X, \mathcal{M}, \mu)$. For each $N \in \mathbb{N}$, define

$$G_N = \sum_{n=1}^{N} |f_n| \quad \text{and} \quad G = \sum_{n=1}^{\infty} |f_n|.$$

Observe that (G_N^p) is increasing sequence of nonnegative measurable (in fact integrable) functions which converges pointwise to G^p. Therefore by MCT it follows that

$$\begin{aligned}\|G\|_p &= \|G^p\|_1^{1/p}\\ &= \lim_{N\to\infty} \|G_N^p\|_1^{1/p}\\ &= \lim_{N\to\infty}\|G_N\|_p.\end{aligned}$$

In particular, since

$$\begin{aligned}\|G_N\|_p &\leq \sum_{n=1}^{N}\|f_n\|_p\\ &\leq \sum_{n=1}^{\infty}\|f_n\|_p\end{aligned}$$

for all N, we have

$$\begin{aligned}\|G\|_p &\leq \sum_{n=1}^{\infty}\|f_n\|\\ &< \infty.\end{aligned}$$

This implies $G \in L^p(X, \mathcal{M}, \mu)$. Since $\|G^p\|_1 = \|G\|_p^p < \infty$, Proposition (39.2) implies $\{G^p = \infty\}$ has measure zero, which implies $\{G = \infty\}$ has measure zero. Define $F\colon X \to \mathbb{R}$ by

$$F(x) = \begin{cases} 0 & \text{if } G(x) = \infty. \\ \sum_{n=1}^{\infty} f_n(x) & \text{if } G(x) < \infty. \end{cases}$$

for all $x \in X$. Observe that $F(x)$ lands in $\mathbb{R}$ since if $G(x) < \infty$, then $\sum_{n=1}^{\infty} f_n(x)$ is absolutely convergent (and hence convergent since $\mathbb{R}$ is complete). Since $|F| \leq G$ and $G \in L^p(X, \mathcal{M}, \mu)$, we see that $F \in L^p(X, \mathcal{M}, \mu)$. Finally, observe that

$$\begin{aligned}\lim_{N\to\infty}\|s_N - F\|_p^p &= \lim_{N\to\infty}\int_X |s_N - F|^p \mathrm{d}\mu\\ &= \lim_{N\to\infty}\int_X \left|\sum_{n=N+1}^{\infty} f_n\right|^p \mathrm{d}\mu.\\ &= \int_X \lim_{N\to\infty}\left|\sum_{n=N+1}^{\infty} f_n\right|^p \mathrm{d}\mu\\ &= \int_X 0\mathrm{d}\mu\\ &= 0.\end{aligned}$$

where we applied DCT to get from the second step to the third step with G^p being the dominating function. □

Problem 3.b

Proposition 39.4. *Let $1 < p < \infty$. Then the set of simple functions in $L^p(X, \mathcal{M}, \mu)$ is a dense subspace of $L^p(X, \mathcal{M}, \mu)$.*

Proof. Let $f \in L^p(X, \mathcal{M}, \mu)$. Decompose f into its positive and negative parts

$$f = f^+ - f^-.$$

There exists an increasing sequence (φ_n) of nonnegative simple functions which converges to f^+ pointwise. Similarly, there exists an increasing sequence (ψ_n) of nonnegative simple functions which converges to f^- pointwise. Then $(\varphi_n + \psi_n)$ is an increasing sequence of nonnegative simple functions which converges pointwise to $|f|$. Also note that $(\varphi_n - \psi_n)$ is a sequence of simple functions which converges pointwise to f. We claim that $\|s_n - f\|_p \to 0$ as $n \to \infty$. Indeed, it suffices to show that $\||s_n - f|^p\|_1 \to 0$ since $\|s_n - f\|_p = \||s_n - f|^p\|_1^{1/p}$ for all

n. To this, we'll use DCT. Clearly $(|s_n - f|^p)$ is a sequence of measurable functions which converges pointwise to 0. Also observe that

$$\begin{aligned}
|s_n - f|^p &\leq (|s_n| + |f|)^p \\
&= (|\varphi_n + \psi_n| + |f|)^p \\
&= (\varphi_n + \psi_n + |f|)^p \\
&\leq (|f| + |f|)^p \\
&\leq 2^p |f|^p.
\end{aligned}$$

Thus $2^p|f|^p$ is a dominating function, which means we can apply DCT. Therefore

$$\begin{aligned}
\lim_{n\to\infty} \int_X |s_n - f|^p \mathrm{d}\mu &= \int_X \lim_{n\to\infty} |s_n - f|^p \mathrm{d}\mu \\
&= \int_X 0 \mathrm{d}\mu \\
&= 0.
\end{aligned}$$

□

39.4 If $\mu(X) < \infty$, then Uniform Convergence Implies Integral Convergence

Problem 4.a

Proposition 39.5. *Assume that $\mu(X) < \infty$. Suppose $(f_n \colon X \to \mathbb{R})$ is a sequence of integrable functions such that $f_n \to f$ uniformly. Then f is integrable and*

$$\lim_{n\to\infty} \int_X f_n \mathrm{d}\mu = \int_X f \mathrm{d}\mu. \tag{95}$$

Proof. Let $\varepsilon > 0$ and choose $N \in \mathbb{N}$ such that $n \geq N$ implies

$$|f(x) - f_n(x)| < \frac{\varepsilon}{\mu(X)}$$

for all $x \in X$. Then

$$\begin{aligned}
\int_X |f| \mathrm{d}\mu &= \int_X |f_N + f - f_N| \mathrm{d}\mu \\
&\leq \int_X |f_N| \mathrm{d}\mu + \int_X |f - f_n| \mathrm{d}\mu \\
&< \int_X |f_N| \mathrm{d}\mu + \frac{\varepsilon}{\mu(X)} \mu(X) \\
&< \int_X |f_N| \mathrm{d}\mu + \varepsilon \\
&< \infty.
\end{aligned}$$

It follows that f is integrable. Now observe that $n \geq N$ implies

$$\begin{aligned}
\left| \int_X f \mathrm{d}\mu - \int_X f_n \mathrm{d}\mu \right| &= \left| \int_X (f - f_n) \mathrm{d}\mu \right| \\
&\leq \int_X |f - f_n| \mathrm{d}\mu \\
&< \frac{\varepsilon}{\mu(X)} \mu(X). \\
&= \varepsilon.
\end{aligned}$$

This implies (95). □

39.5 If $\mu(X) < \infty$ and $1 \leq p < q < \infty$, then $L^q(X, \mathcal{M}, \mu) \subseteq L^p(X, \mathcal{M}, \mu)$

Problem 4.b

Proposition 39.6. *Assume that $\mu(X) < \infty$. Let $1 \leq p < q < \infty$. Then $L^q(X, \mathcal{M}, \mu) \subseteq L^p(X, \mathcal{M}, \mu)$.*

Proposition 39.7.

Proof. Let $f \in L^q(X, \mathcal{M}, \mu)$. We want to show that $f \in L^p(X, \mathcal{M}, \mu)$. Let

$$A = \{x \in X \mid |f|(x) > 1\}.$$

Then $|f|^p 1_A < |f|^q 1_A$, thus

$$\begin{aligned} \int_X |f|^p \mathrm{d}\mu &= \int_X (|f|^p 1_A + |f|^p 1_{A^c}) \mathrm{d}\mu \\ &= \int_X |f|^p 1_A \mathrm{d}\mu + \int_X |f|^p 1_{A^c} \mathrm{d}\mu \\ &\leq \int_X |f|^q 1_A \mathrm{d}\mu + \int_X 1_{A^c} \mathrm{d}\mu \\ &\leq \|f\|_q + \mu(A^c) \\ &< \infty. \end{aligned}$$

It follows that $f \in L^p(X, \mathcal{M}, \mu)$. □

39.6 Generalized Dominated Convergence Theorem

Proposition 39.8. *Let $(f_n \colon X \to \mathbb{R})$ and $(g_n \colon X \to [0, \infty))$ be two sequences of integrable functions which converge almost everywhere to integrable functions $f \colon X \to \mathbb{R}$ and $g \colon X \to \mathbb{R}$ respectively. Suppose $|f_n| \leq g_n$ for all n and $\|g_n\|_1 \to \|g\|_1$. Then*

$$\lim_{n \to \infty} \int_X f_n \mathrm{d}\mu = \int_X f \mathrm{d}\mu.$$

Proof. Observe that $(g_n - f_n)$ is a sequence of nonnegative measurable functions. Thus by Fatou's Lemma, we have

$$\begin{aligned} \int_X g \mathrm{d}\mu - \int_X f \mathrm{d}\mu &= \int_X (g - f) \mathrm{d}\mu \\ &\leq \liminf_{n \to \infty} \int_X (g_n - f_n) \mathrm{d}\mu \\ &= \int_X g \mathrm{d}\mu - \limsup_{n \to \infty} \int f_n \mathrm{d}\mu, \end{aligned}$$

where we used the fact that $\|g_n\|_1 \to \|g\|_1$ to get from the second line to the third line. Subtracting $\int_X g \mathrm{d}\mu$ from both sides and canceling the sign gives us

$$\limsup_{n \to \infty} \int_X f_n \mathrm{d}\mu \leq \int_X f \mathrm{d}\mu.$$

Now we apply the same argument with functions $g_n + f_n$ in place of $g_n - f_n$, and we obtain

$$\liminf_{n \to \infty} \int f_n \mathrm{d}\mu \geq \int_X f \mathrm{d}\mu.$$

□

39.7 Almost Everywhere Convergence Plus Integral Convergence Implies L^1 Convergence

Proposition 39.9. *Let $(f_n \colon X \to \mathbb{R})$ be a sequence of integrable functions that converge almost everywhere to an integrable function $f \colon X \to \mathbb{R}$. Then $\|f_n - f\|_1 \to 0$ if and only if $\|f_n\|_1 \to \|f\|_1$.*

Proof. Suppose $\|f_n - f\|_1 \to 0$. Then

$$\begin{aligned}\lim_{n\to\infty} |\|f_n\|_1 - \|f\|_1| &\leq \lim_{n\to\infty} \|f_n - f\|_1 \\ &= 0.\end{aligned}$$

Thus $\|f_n\|_1 \to \|f\|$.

Conversely, suppose $\|f_n\|_1 \to \|f\|_1$. For each $n \in \mathbb{N}$, set $g_n = |f_n| + |f|$, and set $g = 2|f|$. Then $|f_n - f| \leq g_n$, also g_n converges pointwise almost everywhere to g, also

$$\begin{aligned}\|g_n\|_1 &= \int_X (|f_n| + |f|)\mathrm{d}\mu \\ &= \int_X |f_n|\mathrm{d}\mu + \int_X |f|\mathrm{d}\mu \\ &= \|f_n\|_1 + \|f\|_1 \\ &\to 2\|f\|_1 \\ &= \|g\|,\end{aligned}$$

and $f_n - f$ converges pointwise almost everywhere to 0. It follows from problem 5 that

$$\begin{aligned}\|f_n - f\|_1 &\to \|0\|_1 \\ &= 0.\end{aligned}$$

□

If $f\colon X \to \mathbb{R}$ is Integral, then $n\mu(\{|f| > n\}) \to 0$ as $n \to \infty$

Proposition 39.10. *Let $f\colon X \to \mathbb{R}$ be an integral function. Then*

$$\lim_{n\to\infty} n\mu(\{|f| > n\}) = 0.$$

Proof. First we consider the case for integrable simple functions, say

$$\varphi = \sum_{i=1}^{n} a_i 1_{A_i}, \tag{96}$$

where (96) is expressed in canonical form. Being integral here means $\mu(A_i) \neq \infty$ for all $1 \leq i \leq n$. In particular, $|\varphi|$ is bounded above by some N. Thus $n \geq N$ implies

$$\begin{aligned}n\mu(\{\varphi > n\}) &\geq n \cdot 0 \\ &= 0.\end{aligned}$$

Therefore

$$\lim_{n\to\infty} n\mu(\{\varphi > n\}) = 0.$$

Now we prove it for any integral function $f\colon X \to \mathbb{R}$. First note that since $\mu(\{|f| > n\}) \geq \mu(\{f > n\})$, we may assume that f is nonnegative. Using the fact that the set of all integrable simple functions is dense in $L^1(X, \mathcal{M}, \mu)$, choose a nonnegative integrable simple function φ such that $\varphi \leq f$ and $\|f - \varphi\|_1 < \varepsilon$. Let M be an upper bound for φ. Then we have

$$\begin{aligned}\lim_{n\to\infty} n\mu(\{f > n\}) &= \lim_{n\to\infty} n\mu(\{\varphi > n\} \cup \{f - \varphi \geq n - \varphi\}) \\ &\leq \lim_{n\to\infty} n\mu(\{\varphi > n\} \cup \{f - \varphi \geq n - M\}) \\ &\leq \lim_{n\to\infty} n\mu(\{\varphi > n\}) + \lim_{n\to\infty} n\mu(\{f - \varphi \geq n - M\}) \\ &= \lim_{n\to\infty} n\mu(\{f - \varphi \geq n - M\}) \\ &\leq \lim_{n\to\infty} \frac{n}{n - M}\|f - \varphi\|_1 \\ &< \lim_{n\to\infty} \frac{n\varepsilon}{n - M} \\ &= \varepsilon.\end{aligned}$$

Taking $\varepsilon \to 0$ gives us our desired result. □

39.8 Young's Inequality

Exercise 3. Let $x, y \geq 0$ and $0 < \gamma < 1$. Prove that

$$x^{\gamma} y^{1-\gamma} \leq \gamma x + (1-\gamma) y. \tag{97}$$

Deduce the Young's Inequality.

Solution 2. We may assume that $x, y > 0$ since otherwise it is trivial. Set $t = x/y$ and rewrite (97) as

$$t^{\gamma} - \gamma t \leq 1 - \gamma. \tag{98}$$

Thus, to show (97) for all $x, y > 0$, we just need to show (98) for all $t > 0$. To see why (98) holds, define $f \colon \mathbb{R}_{>0} \to \mathbb{R}$ by

$$f(t) = t^{\gamma} - \gamma t$$

for all $t \in \mathbb{R}_{>0}$. Observe that f is a smooth function on $\mathbb{R}_{>0}$, with it's first derivative and second derivative given by

$$f'(t) = \gamma t^{\gamma-1} - \gamma \quad \text{and} \quad f''(t) = \gamma(\gamma-1) t^{\gamma-2}$$

for all $t \in \mathbb{R}_{>0}$. Observe that

$$\begin{aligned} f'(t) = 0 &\iff \gamma t^{\gamma-1} = \gamma \\ &\iff t^{\gamma-1} = 1 \\ &\iff t = 1, \end{aligned}$$

where the last if and only if follows from the fact that t is a positive real number. Also, we clearly have $f''(t) < 0$ for all $t \in \mathbb{R}_{>0}$. Thus, since f is concave down on all of $\mathbb{R}_{>0}$, and $f'(t) = 0$ if and only if $t = 1$, it follows that f has a global maximum at $t = 1$. In particular, we have

$$\begin{aligned} t^{\gamma} - \gamma t &= f(t) \\ &\leq f(1) \\ &\leq 1^{\gamma} - \gamma \cdot 1 \\ &= 1 - \gamma \end{aligned}$$

for all $t \in \mathbb{R}_{>0}$.

With (97) established, we now prove Young's Inequality: Let $a, b \geq 0$ and let $1 \leq p, q < \infty$ such that $1/p + 1/q = 1$. We want to show that

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Set $\gamma = 1/p$ (so $1 - \gamma = 1/q$), $a = x^{\gamma}$, and $b = y^{1-\gamma}$. Then Young's Inequality becomes (97), which was proved above.

40 Exam 2

40.1 Prove function is measurable. Prove set is measurable

Exercise 4. Let $(X, \mathcal{M}, \mu)$ be a measure space and let $f \colon X \to \mathbb{R}$ be a measurable function. Prove that $f/(1+f^2)$ is also a measurable function.

Solution 3. By a proposition proved in class (see Appendix for details), the product and sum of two measurable functions is measurable. Therefore both f and $1 + f^2$ are measurable. It remains to show that $f/(1+f^2)$ is measurable. To see this, note that for any strictly positive measurable function $h \colon X \to (0, \infty)$, the function $1/h$ is measurable. Indeed, for any $c > 0$, we have

$$\begin{aligned} x \in \left\{ \frac{1}{h} < c \right\} &\iff \frac{1}{h(x)} < c \\ &\iff 1 < c h(x) \\ &\iff x \in \left\{ h > \frac{1}{c} \right\}. \end{aligned}$$

Thus $\{1/h < c\} = \{h > 1/c\} \in \mathcal{M}$. If $c \leq 0$, then we have $\{1/h < c\} = \emptyset \in \mathcal{M}$ since h is strictly positive. In either case, we see that $1/h$ is measurable. In particular, since $1 + f^2$ is a strictly positive measurable function, we see that $1/(1 + f^2)$ is a measurable function. Therefore the product $f/(1 + f^2)$ is a measurable function.

Exercise 5. Let $(X, \mathcal{M}, \mu)$ be a measure space and let $f\colon X \to \mathbb{R}$ be a measurable function. Prove that $\{-1 \leq f \leq 1\}$ is a measurable function.

Solution 4. Since f is measurable, we have

$$\begin{aligned}\{f \leq 1\} &= \bigcap_{n=1}^{\infty} \left\{ f < 1 + \frac{1}{n} \right\} \\ &\in \mathcal{M}.\end{aligned}$$

Similarly we have

$$\begin{aligned}\{f \geq -1\} &= \{f < -1\}^c \\ &\in \mathcal{M}.\end{aligned}$$

Therefore

$$\begin{aligned}\{-1 \leq f \leq 1\} &= \{f \geq -1\} \cap \{f \leq 1\} \\ &\in \mathcal{M}.\end{aligned}$$

40.2 $f_n \xrightarrow{\text{pwae}} f$ implies $f_n \xrightarrow{\text{m}} f$

Exercise 6. Let $(X, \mathcal{M}, \mu)$ be a finite measure space. Suppose $(f_n\colon X \to \mathbb{R})$ is a sequence of measurable functions and suppose $f\colon X \to \mathbb{R}$ is another measurable function. Prove that

$$\bigcup_{k=1}^{\infty} \limsup_{n\to\infty} \{|f_n - f| \geq 1/k\} = \{\lim_{n\to\infty} f_n \neq f\} \tag{99}$$

Solution 5. Observe that

$$\begin{aligned} x \in \bigcup_{k=1}^{\infty} \limsup_{n\to\infty} \{|f_n - f| \geq 1/k\} &\iff x \in \limsup_{n\to\infty} \{|f_n - f| \geq 1/k\} \text{ for some } k \\ &\iff x \in \{|f_{\pi_k(n)} - f| \geq 1/k\} \text{ for some } k \text{ for some subsequence } (\pi_k(n))_{n\in\mathbb{N}} \text{ of } (n)_{n\in\mathbb{N}} \\ &\iff |f_{\pi_k(n)}(x) - f(x)| \geq 1/k \text{ for some } k \text{ for some subsequence } (\pi_k(n))_{n\in\mathbb{N}} \text{ of } (n)_{n\in\mathbb{N}} \\ &\iff x \in \{\lim_{n\to\infty} f_n \neq f\} \end{aligned}$$

where the last if and only if follows from the fact that the distance $|f_n(x) - f(x)|$ is frequently greater than $1/k$, which means $f_n(x) \not\to f(x)$. This gives us (99).

Exercise 7. Let $(X, \mathcal{M}, \mu)$ be a finite measure space. Suppose $(f_n\colon X \to \mathbb{R})$ is a sequence of measurable functions and suppose $f\colon X \to \mathbb{R}$ is another measurable function. Prove that if $\lim_{n\to\infty} f_n(x) = f(x)$ for a.e. $x \in X$, then for all $k \in \mathbb{N}$, we have

$$\lim_{N\to\infty} \mu\{|f_N - f| \geq 1/k\} = 0 \tag{100}$$

Solution 6. Let $k \in \mathbb{N}$. Then observe that

$$\begin{aligned}
0 &= \mu \left\{ \lim_{n\to\infty} f_n \neq f \right\} \\
&= \mu \left\{ \bigcup_{m=1}^{\infty} \limsup_{n\to\infty} \{|f_n - f| \geq 1/m\} \right\} \\
&\geq \mu \left\{ \limsup \{|f_n - f| \geq 1/k\} \right\} \\
&= \mu \left(\bigcap_{N=1}^{\infty} \bigcup_{n\geq N} \{|f_n - f| \geq 1/k\} \right) \\
&= \lim_{N\to\infty} \mu \left(\bigcup_{n\geq N} \{|f_n - f| \geq 1/k\} \right) \\
&\geq \lim_{N\to\infty} \mu\{|f_N - f| \geq 1/k\}
\end{aligned}$$

where we used the fact that $\mu(X) < \infty$ to get from the fourth line to the fifth line. This gives us (100).

Exercise 8. Let $(X, \mathcal{M}, \mu)$ be a finite measure space. Suppose $(f_n \colon X \to \mathbb{R})$ is a sequence of measurable functions and suppose $f \colon X \to \mathbb{R}$ is another measurable function. Prove that if $f_n \to f$ a.e, then $f_n \to f$ in measure.

Solution 7. Suppose $f_n \to f$ a.e and let $\varepsilon > 0$. Choose $k \in \mathbb{N}$ such that $1/k < \varepsilon$. Then by part (b), we have

$$\begin{aligned}
\lim_{n\to\infty} \mu\{|f_n - f| \geq \varepsilon\} &\leq \lim_{n\to\infty} \mu\{|f_n - f| \geq 1/k\} \\
&= 0.
\end{aligned}$$

This implies $f_n \to f$ in measure.

Exercise 9. Let $(X, \mathcal{M}, \mu)$ be a finite measure space. Suppose $(f_n \colon X \to \mathbb{R})$ is a sequence of measurable functions and suppose $f \colon X \to \mathbb{R}$ is another measurable function. Prove that if $f_n \xrightarrow{L^2} f$, then $f_n \to f$ in measure.

Solution 8. Let $g \in L^2(X, \mathcal{M}, \mu)$. Since $\mu(X) < \infty$, we also have $1_X \in L^2(X, \mathcal{M}, \mu)$. By Hölder's inequality, we have

$$\begin{aligned}
\|g\|_1 &\leq \|g\|_2 \cdot \|1_X\|_2 \\
&= \sqrt{\mu(X)} \|g\|_2.
\end{aligned}$$

In particular, $f_n \xrightarrow{L^2} f$ implies $f_n \xrightarrow{L^1} f$ which implies $f_n \to f$ in measure (proved in class).

40.3 Integral Computation (Using Descending MCT)

Exercise 10. Compute the following limit

$$\lim_{n\to\infty} \int_{(0,1)} \frac{1+nx}{(1+x)^n} \mathrm{d}x$$

Solution 9. For each $n \in \mathbb{N}$, let $f_n = (1+nx)(1+x)^{-n}$. Observe that each f_n is measurable since each f_n is continuous (the denominator is never zero for any $x \in \mathbb{R}$). Furthermore, each f_n is nonnegative (since taking squares makes everything nonnegative). We claim that f_n is a decreasing sequence. Indeed

$$\begin{aligned}
\frac{f_n}{f_{n+1}} &= \left(\frac{1+nx}{(1+x)^n} \right) \left(\frac{(1+x)^{n+1}}{1+(n+1)x} \right) \\
&= \frac{(1+nx)(1+x)}{1+(n+1)x} \\
&= \frac{nx^2 + (n+1)x + 1}{(n+1)x + 1} \\
&> 1
\end{aligned}$$

Thus (f_n) is a decreasing sequence which is bounded below, so it must converge pointwise to some function. To see what it converges to, we use L'Hopital's rule:

$$\begin{aligned}
\lim_{n\to\infty} f_n &= \lim_{n\to\infty} \frac{1+nx}{(1+x)^n} \\
&= \lim_{n\to\infty} \frac{x}{\ln(1+x)(1+x)^n} \\
&= 0.
\end{aligned}$$

Thus (f_n) converges pointwise to 0. Since

$$\begin{aligned}
\int_0^1 f_1 \mathrm{d}x &= \int_0^1 \frac{1+x}{1+x}\mathrm{d}x \\
&= \int_0^1 \mathrm{d}x \\
&= 1 \\
&< \infty,
\end{aligned}$$

it follows from (decreasing version of MCT) that

$$\begin{aligned}
\lim_{n\to\infty} \frac{1+nx}{(1+x)^n}\mathrm{d}x &= \lim_{n\to\infty} \int_0^1 f_n \mathrm{d}x \\
&= \int_0^1 0 \mathrm{d}x \\
&= 0.
\end{aligned}$$

40.4 Strictly Positive Measurable Function Gives Rise to Strictly Positive Measure

Exercise 11. Let $(X, \mathcal{M}, \mu)$ be a measure space. Suppose that $f\colon X \to \mathbb{R}$ is a measurable function such that $f(x) > 0$ for all $x \in X$. Prove that $\int_X 1_E f \mathrm{d}\mu > 0$ for every measurable $E \in \mathcal{M}$ such that $\mu(E) > 0$.

Solution 10. Let $E \in \mathcal{M}$ such that $\mu(E) > 0$. For each $n \in \mathbb{N}$, define

$$F_n := \{f \geq 1/n\}.$$

Since $f(x) > 0$ for all $x \in X$, we have

$$\begin{aligned}
0 &< \mu(E) \\
&= \mu\left(\bigcup_{n=1}^{\infty} F_n \cap E\right) \\
&\leq \sum_{n=1}^{\infty} \mu(F_n \cap E).
\end{aligned}$$

The strict inequality implies $\mu(F_n \cap E) > 0$ for some $n \in \mathbb{N}$. Choose such an $n \in \mathbb{N}$, then we have

$$\begin{aligned}
\int_X f 1_E \mathrm{d}\mu &\geq \int_X f 1_{E\cap F_n} \mathrm{d}\mu \\
&\geq \int_X \frac{1}{n} \cdot 1_{E\cap F_n} \mathrm{d}\mu \\
&= \mu(E \cap F_n)/n \\
&> 0.
\end{aligned}$$

40.5 Problem Involving MCT

Exercise 12. Let $f\colon [0,1] \to [0,\infty)$ be a nonnegative measurable function. Prove that if

$$\lim_{n\to\infty} \int_{[0,1]} f 1_{[0, \frac{n}{n+1}]} \mathrm{d}x \leq 1$$

for all $n \in \mathbb{N}$, then f is integrable and $\int_{[0,1]} f \mathrm{d}x \leq 1$.

Solution 11. Observe that since $(n/(n+1))$ is an increasing sequence which converges to 1, the sequence $(f1_{[0,\frac{n}{n+1}]})$ is an increasing sequence of nonnegative measurable functions which converges pointwise to f. It follows from MCT that

$$\begin{aligned}\int_{[0,1]} f\mathrm{d}x &= \lim_{n\to\infty}\int_{[0,1]} f1_{[0,\frac{n}{n+1}]}\mathrm{d}x \\ &\leq 1.\end{aligned}$$

In particular, f is integrable and $\int_{[0,1]} f\mathrm{d}x \leq 1$.

40.6 Nonnegative Measurable Function Induces Finite Measure

Exercise 13. Let $(X,\mathcal{M},\mu)$ be a measure space and let $f\colon X\to\mathbb{R}$ be a nonnegative measurable function such that $\int_X f\mathrm{d}\mu<\infty$. Prove that $\nu\colon\mathcal{M}\to\mathbb{R}$ defined by

$$\nu(E)=\int_X f1_E\mathrm{d}\mu$$

is a finite measure on $(X,\mathcal{M})$.

Solution 12. This was proved in the homework, but we include it for completeness.

First we prove it for nonnegative simple functions:

Proposition 40.1. *Let $\phi\colon X\to[0,\infty)$ be a nonnegative simple function. Define a function $\nu\colon\mathcal{M}\to[0,\infty]$ by*

$$\nu(E)=\int_X \phi1_E\mathrm{d}\mu$$

for all $E\in\mathcal{M}$. Then ν is a measure on $(X,\mathcal{M})$.

Proof. First note that

$$\begin{aligned}\nu(\emptyset) &= \int_X \phi1_\emptyset\mathrm{d}\mu \\ &= \int_X \phi\cdot 0\cdot\mathrm{d}\mu \\ &= \int_X 0\cdot\mathrm{d}\mu \\ &= 0.\end{aligned}$$

Now we show that ν is finitely additive. Let $(E_n)_{n=1}^N$ be a finite sequence of pairwise disjoint sets in $\mathcal{M}$. Then

$$\begin{aligned}\nu\left(\bigcup_{n=1}^N E_n\right) &= \int_X \phi1_{\bigcup_{n=1}^N E_n}\mathrm{d}\mu \\ &= \int_X \phi\sum_{n=1}^N 1_{E_n}\mathrm{d}\mu \\ &= \sum_{n=1}^N\int_X \phi1_{E_n}\mathrm{d}\mu \\ &= \sum_{n=1}^N \nu(E_n),\end{aligned}$$

where we used the fact that each $\phi1_{E_n}$ is a nonnegative simple function in order to commute the finite sum with the integral. Thus it follows that ν is finitely additive. It remains to show that ν is countably subadditive. Let (E_n) be a sequence of sets in $\mathcal{M}$. We want to show that

$$\int_X \phi1_{\bigcup_{n=1}^\infty E_n}\mathrm{d}\mu\leq\sum_{n=1}^\infty\int_X\phi1_{E_n}\mathrm{d}\mu. \tag{101}$$

To do this, we will show that the sum on the righthand side in (101) is greater than or equal to all integrals of the form $\int \varphi d\mu$ where $\varphi\colon X \to [0,\infty]$ is a simple function such that $\varphi \leq \phi 1_{\bigcup_{n=1}^\infty E_n}$. Then the inequality (101) will follow from the fact that the integral on the lefthand side in (101) is the supremum of this set. So let $\varphi\colon X \to [0,\infty]$ be a simple function such that $\varphi \leq \phi 1_{\bigcup_{n=1}^\infty E_n}$. Write φ and ϕ in terms of their canonical forms, say

$$\varphi = \sum_{i=1}^k a_i 1_{A_i} \quad \text{and} \quad \phi = \sum_{j=1}^m b_j 1_{B_j}.$$

So $a_i \neq a_{i'}$ and $A_i \cap A_{i'} = \emptyset$ whenever $i \neq i'$ and $b_j \neq b_{j'}$ and $B_j \cap B_{j'} = \emptyset$ whenever $j \neq j'$. Observe that the canonical representation of $\phi 1_{\bigcup_{n=1}^\infty E_n}$ is given by

$$\begin{aligned}
\phi 1_{\bigcup_{n=1}^\infty E_n} &= \left(\sum_{j=1}^m b_j 1_{B_j}\right) 1_{\bigcup_{n=1}^\infty E_n} \\
&= \sum_{j=1}^m b_j 1_{B_j} 1_{\bigcup_{n=1}^\infty E_n} \\
&= \sum_{j=1}^m b_j 1_{\bigcup_{n=1}^\infty B_j \cap E_n},
\end{aligned}$$

where this representation is the canonical representation since $b_j \neq b_{j'}$ and

$$\left(\bigcup_{n=1}^\infty B_j \cap E_n\right) \cap \left(\bigcup_{n=1}^\infty B_{j'} \cap E_n\right) = \emptyset$$

whenever $j \neq j'$ (since $B_j \cap B_{j'} = \emptyset$). Therefore we have

$$\begin{aligned}
\int_X \varphi d\mu &\leq \int_X \phi 1_{\bigcup_{n=1}^\infty E_n} d\mu \\
&= \sum_{j=1}^m b_j \mu\left(\bigcup_{n=1}^\infty B_j \cap E_n\right) \\
&\leq \sum_{j=1}^m b_j \sum_{n=1}^\infty \mu(B_j \cap E_n) \\
&= \sum_{n=1}^\infty \sum_{j=1}^m b_j \mu(B_j \cap E_n) \\
&= \sum_{n=1}^\infty \sum_{j=1}^m \int_X b_j 1_{B_j \cap E_n} d\mu \\
&= \sum_{n=1}^\infty \int_X \sum_{j=1}^m b_j 1_{B_j \cap E_n} d\mu \\
&= \sum_{n=1}^\infty \int_X \sum_{j=1}^m b_j \left(1_{B_j} 1_{E_n}\right) d\mu \\
&= \sum_{n=1}^\infty \int_X \left(\sum_{j=1}^m b_j 1_{B_j}\right) 1_{E_n} d\mu \\
&= \sum_{n=1}^\infty \int_X \phi 1_{E_n} d\mu,
\end{aligned}$$

where we used monotonicity of integration in the first line and where we used countable subadditivity of μ to get from the second line to the third line.

□

Now we prove it for more general nonnegative measurable functions

Proposition 40.2. *Let $(X, \mathcal{M}, \mu)$ be measure space and let $g\colon X \to [0,\infty)$ be a nonnegative measurable function. Define $\nu_g\colon \mathcal{M} \to [0,\infty]$ by*

$$\nu_g(E) = \int_X g1_E \mathrm{d}\mu$$

for all $E \in \mathcal{M}$. Then ν is a measure on $(X, \mathcal{M})$. Furthermore, if $\int_X g\mathrm{d}\mu < \infty$, then $(X, \mathcal{M}, \nu)$ is a finite measure space.

Proof. First note that

$$\begin{aligned}
\nu_g(\emptyset) &= \int_X g1_\emptyset \mathrm{d}\mu \\
&= \int_X g \cdot 0 \cdot \mathrm{d}\mu \\
&= \int_X 0 \cdot \mathrm{d}\mu \\
&= 0.
\end{aligned}$$

Next we show that ν_g is finitely additive. Let $(E_n)_{n=1}^N$ be a finite sequence of pairwise disjoint sets in $\mathcal{M}$. Then

$$\begin{aligned}
\nu_g\left(\bigcup_{n=1}^N E_n\right) &= \int_X g1_{\bigcup_{n=1}^N E_n} \mathrm{d}\mu \\
&= \int_X g \sum_{n=1}^N 1_{E_n} \mathrm{d}\mu \\
&= \int_X \sum_{n=1}^N g1_{E_n} \mathrm{d}\mu \\
&= \sum_{n=1}^N \int_X g1_{E_n} \mathrm{d}\mu \\
&= \sum_{n=1}^N \nu_g(E_n),
\end{aligned}$$

where we used the fact that each $g1_{E_n}$ is a nonnegative measurable function in order to commute the finite sum with the integral. Thus it follows that ν_g is finitely additive.

It remains to show that ν_g is countably subadditive. In problem 3 in HW4, we showed that for any nonnegative simple function $\varphi\colon X \to [0,\infty)$, the function $\nu_\varphi\colon \mathcal{M} \to [0,\infty]$ defined by

$$\nu_\varphi(E) = \int_X \varphi 1_E \mathrm{d}\mu$$

is a measure. We will use this fact in what follows:

Choose an increasing sequence $(\varphi_n\colon X \to [0,\infty])$ of nonnegative simple functions which converges pointwise to g (we can do this since g is a nonnegative measurable function). Then $(\varphi_n 1_E)$ is an increasing sequence of nonnegative simple functions which converges pointwise to $g1_E$ for all $E \in \mathcal{M}$. It follows from the Monotone Convergence Theorem that

$$\begin{aligned}
\nu_{\varphi_n}(E) &= \int_X \varphi_n 1_E \mathrm{d}\mu \\
&\to \int_X g1_E \mathrm{d}\mu \\
&= \nu_g(E)
\end{aligned}$$

for any $E \in \mathcal{M}$. In particular, for any $E \in \mathcal{M}$ and $\varepsilon > 0$, we can find a $N_{E,\varepsilon} \in \mathbb{N}$ (which depends on E and ε) such that

$$\nu_g(E) < \nu_{\varphi_n}(E) + \varepsilon \tag{102}$$

for all $n \geq N_{E,\varepsilon}$. However we will don't need estimate $\nu_g(E)$ to this level of precision. We just need to know thta for any $\varepsilon > 0$, we can find an $n \in \mathbb{N}$ such that (102) holds.

Now let (E_k) be a sequence of measurable sets and let $\varepsilon > 0$. Choose $n \in \mathbb{N}$ such that

$$\nu_g\left(\bigcup_{k=1}^{\infty} E_k\right) < \nu_{\varphi_n}\left(\bigcup_{k=1}^{\infty} E_k\right) + \varepsilon$$

Then we have

$$\begin{aligned}
\nu_g\left(\bigcup_{k=1}^{\infty} E_k\right) &\leq \nu_{\varphi_n}\left(\bigcup_{k=1}^{\infty} E_k\right) + \varepsilon \\
&\leq \sum_{k=1}^{\infty} \nu_{\varphi_n}(E_k) + \varepsilon \\
&\leq \sum_{k=1}^{\infty} \nu_g(E_k) + \varepsilon
\end{aligned}$$

where we obtained the second line from the first line using countable subadditivity of ν_{φ_n}, and where we obtained the third line from the second line from the fact that $\varphi_n 1_{E_k} \leq g 1_{E_k}$ for all k and from monotonicity of integration. Taking $\varepsilon \to 0$ gives us countable subadditivity of ν_g.

Finally, for the last part, we note that

$$\begin{aligned}
\nu(X) &= \int_X g 1_X \mathrm{d}\mu \\
&= \int_X g \mathrm{d}\mu \\
&< \infty.
\end{aligned}$$

□

Exercise 14. Let $(X, \mathcal{M}, \mu)$ be a measure space and let $f\colon X \to \mathbb{R}$ be a nonnegative measurable function such that $\int_X f \mathrm{d}\mu < \infty$. Prove that for any sequence (E_n) of measurable sets such that $\sum_{n=1}^{\infty} \nu(E_n) < \infty$, we have

$$\lim_{n\to\infty} 1_{E_n} f = 0$$

for μ a.e. x.

Solution 13. First recall three propositions we proved in the homework:

Proposition 40.3. *Let $(X, \mathcal{M}, \mu)$ be a measure space and let (E_n) be a sequence in $\mathcal{M}$. Then*

$$\mu\left(\liminf E_n\right) \leq \liminf \mu(E_n)$$

Proof. Note that the sequence

$$\left(\bigcap_{n\geq N} E_n\right)_{N\in\mathbb{N}}$$

is an ascending sequence in N. Therefore we have

$$\begin{aligned}
\mu\left(\liminf E_n\right) &= \mu\left(\bigcup_{N=1}^{\infty}\left(\bigcap_{n\geq N} E_n\right)\right) \\
&= \liminf \mu\left(\bigcap_{n\geq N} E_n\right) \\
&\leq \lim_{N\to\infty} \inf\left\{\mu(E_n) \mid n \geq N\right\} \\
&= \liminf \mu(E_n),
\end{aligned}$$

where we obtained the third line from the second line since

$$\mu\left(\bigcap_{n\geq N} E_n\right) \leq \mu(E_n)$$

for all $n \geq N$ by monotonicity of μ. □

Proposition 40.4. *Let $(X, \mathcal{M}, \mu)$ be a measure space and let (E_n) be a sequence in $\mathcal{M}$ such that $\mu\left(\bigcup_{n=1}^{\infty} E_n\right) < \infty$. Then*

$$\mu\left(\limsup E_n\right) \geq \limsup \mu(E_n)$$

Proof. Note that the sequence

$$\left(\bigcup_{n \geq N} E_n\right)_{N \in \mathbb{N}}$$

is a descending sequence in N. This together with the fact that $\mu\left(\bigcup_{n=1}^{\infty} E_n\right) < \infty$ implies

$$\begin{aligned}
\mu\left(\limsup E_n\right) &= \mu\left(\bigcap_{N=1}^{\infty}\left(\bigcup_{n \geq N} E_n\right)\right) \\
&= \lim_{N \to \infty} \mu\left(\bigcup_{n \geq N} E_n\right) \\
&\geq \lim_{N \to \infty} \sup\left\{\mu(E_n) \mid n \geq N\right\} \\
&= \limsup \mu(E_n),
\end{aligned}$$

where we obtained the third line from the second line since

$$\mu\left(\bigcup_{n \geq N} E_n\right) \geq \mu(E_n)$$

for all $n \geq N$ by monotonicity of μ. □

Proposition 40.5. *Let $(X, \mathcal{M}, \mu)$ be a measure space and let (E_n) be a sequence in $\mathcal{M}$ such that $\sum_{n=1}^{\infty} \mu(E_n) < \infty$. Then*

$$\mu\left(\limsup E_n\right) = 0.$$

Proof. Note that the sequence

$$\left(\bigcup_{n \geq N} E_n\right)_{N \in \mathbb{N}}$$

is a descending sequence in N. This together with the fact that $\mu\left(\bigcup_{n=1}^{\infty} E_n\right) < \infty$ implies

$$\begin{aligned}
\mu\left(\limsup E_n\right) &= \mu\left(\bigcap_{N=1}^{\infty}\left(\bigcup_{n \geq N} E_n\right)\right) \\
&= \lim_{N \to \infty} \mu\left(\bigcup_{n \geq N} E_n\right) \\
&\leq \lim_{N \to \infty} \sum_{n=N}^{\infty} \mu(E_n) \\
&= 0,
\end{aligned}$$

where the last equality follows since $\sum_{n=1}^{\infty} \mu(E_n) < \infty$. □

Now we prove part (b). Let (E_n) be a sequence of measurable sets such that

$$\sum_{n=1}^{\infty} \nu(E_n) < \infty.$$

Then observe that

$$\begin{aligned}
\int_X \lim_{n \to \infty} f 1_{E_n} \mathrm{d}\mu &= \int_X \liminf f 1_{E_n} \mathrm{d}\mu \\
&\leq \liminf \int_X f 1_{E_n} \mathrm{d}\mu \\
&= \liminf \nu(E_n) \\
&\leq \limsup \nu(E_n) \\
&\leq \nu(\limsup E_n) \\
&= 0.
\end{aligned}$$

where we applied Fatou's Lemma to get the second line from the first line. It follows that $\lim_{n\to\infty} f1_{E_n} = 0$ almost everywhere (by a proposition proved in class).

Appendix

Problem 1

Proposition 40.6. *Let $f,g\colon X \to \mathbb{R}$ be measurable functions and let $a \in \mathbb{R}$. Then af, $|f|$, f^2, $f+g$, fg, $\max\{f,g\}$, and $\min\{f,g\}$ are all measurable.*

Proof. We first show af is measurable. If $a = 0$, then af is the zero function, which is measurable. So assume $a \neq 0$. Then we have

$$(af)^{-1}(-\infty, c) = \begin{cases} f^{-1}(-\infty, c/a) \in \mathcal{M} & \text{if } a > 0 \\ f^{-1}(c/a, \infty) \in \mathcal{M} & \text{if } a < 0 \end{cases}$$

αf is measurable αf is measurable.
Observe that

$$\begin{aligned} x \in (f+g)^{-1}(-\infty, c) &\iff f(x) + g(x) < c \\ &\iff \text{there exists an } r \in \mathbb{Q} \text{ such that } f(x) < r \text{ and } r < c - g(x) \\ &\iff \text{there exists an } r \in \mathbb{Q} \text{ such that } x \in f^{-1}(-\infty, r) \cap g^{-1}(-\infty, c-r). \\ &\iff x \in \bigcup_{r \in \mathbb{Q}} f^{-1}(-\infty, r) \cap g^{-1}(-\infty, c-r). \end{aligned}$$

Therefore

$$(f+g)^{-1}(-\infty, c) = \bigcup_{r \in \mathbb{Q}} f^{-1}(-\infty, r) \cap g^{-1}(-\infty, c-r) \in \mathcal{M}.$$

We first prove f^2 is measurable:

$$(f^2)^{-1}(c, \infty) = \begin{cases} f^{-1}(\sqrt{c}, \infty) \cup f^{-1}(-\infty, -\sqrt{c}) \in \mathcal{M} & c \geq 0 \\ E \in \mathcal{M} & c < 0 \end{cases}$$

Next, note that

$$fg = \frac{1}{4}\left((f+g)^2 - (f-g)^2\right) \in \mathcal{M}.$$

Finally note that

$$\max\{f,g\} = \frac{1}{2}(|f+g| + |f-g|)$$

and

$$\min\{f,g\} = \frac{1}{2}(|f+g| - |f-g|).$$

□

Proposition 40.7. *Let $(f_n\colon X \to \mathbb{R})$ be a sequence of measurable functions. Then $\sup f_n$, $\inf f_n$, $\limsup f_n$, and $\liminf f_n$ are all measurable. In particular, of*

$$\lim_{n\to\infty} f_n(x)$$

exists for all $x \in X$. The corresponding function is also measurable.

Proof. Let $c \in \mathbb{R}$. Then we have

$$(\sup f_n)^{-1}(c, \infty) = \bigcup_n f_n^{-1}(c, \infty) \in \mathcal{M}.$$

Similarly, we have

$$(\inf f_n)^{-1}(-\infty, c) = \bigcup_n f_n^{-1}(-\infty, c) \in \mathcal{M}.$$

Also we have

$$\limsup f_n = \inf_k \sup_{n \geq k} \in \mathcal{M}.$$

Similarly, we have

$$\liminf f_n = \sup_k \inf_{n \geq k} f_n \in \mathcal{M}$$

□

Problem 2

Proposition 40.8. *If $f_n \xrightarrow{L^1} f$, then $f_n \xrightarrow{\mathrm{m}} f$.*

Proof. Suppose $f_n \xrightarrow{L^1} f$ and let $\varepsilon, \delta > 0$. Choose $N_{\varepsilon,\delta} \in \mathbb{N}$ such that $n \geq N_{\varepsilon,\delta}$ implies

$$\|f_n - f\|_1 < \varepsilon\delta.$$

Then it follows from Chebyshev's inequality that

$$\begin{aligned}
\mu\left(\{\{x \in X \mid |f_n(x) - f(x)| \geq \varepsilon\}\}\right) &\leq \frac{1}{\varepsilon}\|f_n - f\|_1 \\
&< \frac{1}{\varepsilon}\varepsilon\delta \\
&= \delta.
\end{aligned}$$

Thus $f_n \xrightarrow{\mathrm{m}} f$. □

Problem 3

Proposition 40.9. *Let $(X, \mathcal{M}, \mu)$ be a measure space. Suppose $(f_n \colon X \to [0,\infty])$ is a decreasing sequence of nonnegative measurable functions which converges pointwise to $f \colon X \to [0,\infty]$. If $\int_X f_1 \mathrm{d}\mu < \infty$, then*

$$\lim_{n\to\infty} \int_X f_n \mathrm{d}\mu = \int_X f \mathrm{d}\mu. \tag{103}$$

Proof. For each $n \in \mathbb{N}$, set $g_n = f_{n+1} - f_n$. Since (f_n) is a decreasing sequence, each g_n is nonnegative. Furthermore each g_n is measurable since it is a difference of two measurable functions. Set $g = \sum_{n=1}^{\infty} g_n$ and observe that

$$\begin{aligned}
g &= \sum_{n=1}^{\infty} g_n \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} g_n \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} (f_{n+1} - f_n) \\
&= \lim_{N\to\infty} (f_N - f_1) \\
&= f - f_1.
\end{aligned}$$

It follows from problem 4 that

$$\begin{aligned}
\int_X f\mathrm{d}\mu - \int_X f_1\mathrm{d}\mu &= \int_X (f - f_1)\mathrm{d}\mu \\
&= \int_X g\mathrm{d}\mu \\
&= \sum_{n=1}^{\infty} \int_X g_n\mathrm{d}\mu \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \int_X g_n\mathrm{d}\mu \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \int_X (f_{n+1} - f_n)\mathrm{d}\mu \\
&= \lim_{N\to\infty} \int_X \sum_{n=1}^{N} (f_{n+1} - f_n)\mathrm{d}\mu \\
&= \lim_{N\to\infty} \int_X (f_N - f_1)\mathrm{d}\mu \\
&= \lim_{N\to\infty} \int_X f_N\mathrm{d}\mu - \int_X f_1\mathrm{d}\mu.
\end{aligned}$$

Since $\int_X f_1\mathrm{d}\mu < 0$, we can cancel it from both sides to get (92). □

Problem 6

Proposition 40.10. *If $\int_X |f|\mathrm{d}\mu = 0$, then $\mu(\{f \neq 0\}) = 0$.*

Proof. Note that $\{f \neq 0\} = \{|f| \neq 0\}$. Also $\{|f| \neq 0\} = \bigcup_{n=1}^{\infty}\{|f| \geq 1/n\}$. Thus

$$\begin{aligned}
\mu(\{|f| \neq 0\}) &= \mu\left(\bigcup_{n=1}^{\infty}\{|f| \geq 1/n\}\right) \\
&= \lim_{n\to\infty} \mu(\{|f| \geq 1/n\}) \\
\text{(C-M)} &\leq \lim_{n\to\infty} n\int_X |f|\mathrm{d}\mu \\
&= 0.
\end{aligned}$$

□

Appendix

41 Pseudometric Spaces

Recall that a metric space is a pair (X, d) where X is a set and where $\mathrm{d}\colon X \times X \to \mathbb{R}$ is a metric on X which satisfies

1. (Identity of Indiscernibles) $\mathrm{d}(x,y) = 0$ if and only if $x = y$;
2. (Symmetry) $\mathrm{d}(x,y) = \mathrm{d}(y,x)$ for all $x, y \in X$,
3. (Triangle Inequality) $\mathrm{d}(x,z) \leq \mathrm{d}(x,y) + \mathrm{d}(y,z)$ for all $x, y, z \in X$.

If we weaken the "Identity of Indiscernibles" axiom, then we get a pseudometric space:

Definition 41.1. A **pseudometric** on a set X is a function $\mathrm{d}\colon X \times X \to \mathbb{R}$ which satisfies the following three properties:

1. (Reflexivity) $\mathrm{d}(x,x) = 0$ for all $x \in X$;
2. (Symmetry) $\mathrm{d}(x,y) = \mathrm{d}(y,x)$ for all $x,y \in X$,
3. (Triangle Inequality) $\mathrm{d}(x,z) \leq \mathrm{d}(x,y) + \mathrm{d}(y,z)$ for all $x,y,z \in X$.

If d is a pseudometric on a set X, then we call the pair (X, d) a **pseudometric space**. If the pseudometric is understood from context, then we often denote a pseudometric space by X instead of (X, d).

Remark 56. Given the three axioms above, we also have $\mathrm{d}(x,y) \geq 0$ for all $x,y \in X$. Indeed,

$$\begin{aligned} 0 &= \mathrm{d}(x,x) \\ &\leq \mathrm{d}(x,y) + \mathrm{d}(y,x) \\ &= \mathrm{d}(x,y) + \mathrm{d}(x,y) \\ &= 2\mathrm{d}(x,y). \end{aligned}$$

This implies $\mathrm{d}(x,y) \geq 0$.

41.1 Topology Induced by Pseudometric Space

Proposition 41.1. *Let (X, d) be a pseudometric space. For each $x \in X$ and $r > 0$, define*

$$\mathrm{B}_r^{\mathrm{d}}(x) := \{y \in X \mid \mathrm{d}(x,y) < r\},$$

and let

$$\mathcal{B}^{\mathrm{d}} = \{\mathrm{B}_r(x) \mid x \in X \text{ and } r > 0\}.$$

Finally, let $\tau(\mathcal{B}^{\mathrm{d}})$ be the smallest topology on X which contains $\mathcal{B}^{\mathrm{d}}$. Then $\mathcal{B}^{\mathrm{d}}$ is a basis for $\tau(\mathcal{B}^{\mathrm{d}})$.

Remark 57. We often remove the d in the superscript in $\mathrm{B}_r^{\mathrm{d}}(x)$ and $\mathcal{B}^{\mathrm{d}}$ whenever context is clear.

Proof. First note that $\mathcal{B}$ covers X. Indeed, for any $r > 0$, we have

$$X \subseteq \bigcup_{x \in X} \mathrm{B}_r(x).$$

Next, let $\mathrm{B}_r(x)$ and $\mathrm{B}_{r'}(x')$ be two members of $\mathcal{B}$ which have nontrivial intersection and let $x'' \in \mathrm{B}_r(x) \cap \mathrm{B}_{r'}(x')$. Set

$$r'' = \min\{r' - \mathrm{d}(x', x''), r - \mathrm{d}(x, x'')\}.$$

We claim that $\mathrm{B}_{r''}(x'') \subseteq \mathrm{B}_r(x) \cap \mathrm{B}_{r'}(x')$. Indeed, assume without loss of generality that $r'' = r - \mathrm{d}(x, x'')$. Let $y \in \mathrm{B}_{r''}(x'')$. Then

$$\begin{aligned} \mathrm{d}(y,x) &\leq \mathrm{d}(y,x'') + \mathrm{d}(x'',x) \\ &< r - \mathrm{d}(x,x'') + \mathrm{d}(x'',x) \\ &= r - \mathrm{d}(x'',x) + \mathrm{d}(x'',x) \\ &= r \end{aligned}$$

implies $y \in \mathrm{B}_r(x)$. Similarly,

$$\begin{aligned} \mathrm{d}(y,x') &\leq \mathrm{d}(y,x'') + \mathrm{d}(x'',x') \\ &< r' - \mathrm{d}(x',x'') + \mathrm{d}(x'',x') \\ &= r' - \mathrm{d}(x'',x') + \mathrm{d}(x'',x') \\ &= r' \end{aligned}$$

implies $y \in \mathrm{B}_{r'}(x')$. Thus $\mathrm{B}_{r''}(x'') \subseteq \mathrm{B}_r(x) \cap \mathrm{B}_{r'}(x')$, and so $\mathcal{B}$ is a basis for $\tau(\mathcal{B})$. □

Definition 41.2. The topology $\tau(\mathcal{B})$ in Proposition (41.1) is called the **topology induced by the pseudometric** d. We also denote this topology by τ_{d}.

41.1.1 Subspace topology agrees with topology induced by pseudometric

Let (X, d) be a pseudometric space and let $A \subseteq X$. Then the pseudometric on X restricts to a pseudometric on A. We denote this restriction by $\mathrm{d}|_A$. Thus there are two natural topologies on A. One is the subspace is the subspace topology given by

$$\tau \cap A := \{U \cap A \mid U \in \tau\}.$$

The other is the topology induced by the pseudometric $\mathrm{d}|_A$ given by

$$\tau_{\mathrm{d}|_A} := \tau(\mathcal{B}^{\mathrm{d}|_A}).$$

The next proposition tells us that these are actually the same.

Proposition 41.2. *Let (X, d) be a pseudometric space and let $A \subseteq X$. Then*

$$\tau_{\mathrm{d}} \cap A = \tau_{\mathrm{d}|_A}.$$

Proof. Let $a \in A$ and $r > 0$. Then

$$\begin{aligned} \mathrm{B}_r^{\mathrm{d}|_A}(a) &= \{b \in A \mid \mathrm{d}|_A(a, b) < r\} \\ &= \{b \in A \mid \mathrm{d}(a, b) < r\} \\ &= A \cap \{x \in X \mid \mathrm{d}(a, x) < r\} \\ &= A \cap \mathrm{B}_r^{\mathrm{d}}(a). \end{aligned}$$

It follows that $\tau_{\mathrm{d}|_A}$ and $\tau_{\mathrm{d}} \cap A$ have the same basis, and hence $\tau_{\mathrm{d}} \cap A = \tau_{\mathrm{d}|_A}$. □

41.1.2 Convergence in (X, d)

Concepts like convergence and completion still make sense in pseudometric spaces. Let us state these definitions in the context of pseudometric spaces now.

Definition 41.3. Let (X, d) be a pseudometric space and let (x_n) be a sequence in X.

1. We say the sequence (x_n) converges to $x \in X$ if for all $\varepsilon > 0$ there exists an $N_\varepsilon \in \mathbb{N}$ (we write ε in the subscript because N_ε depends on ε, however we usually omit ε and just write N) such that

$$n \geq N_\varepsilon \text{ implies } \mathrm{d}(x_n, x) < \varepsilon.$$

   In this case, we say (x_n) is a **convergent sequence** and that it **converges** to x. We denote this by $x_n \to x$ as $n \to \infty$, or $\lim_{n \to \infty} x_n = x$, or even just $x_n \to x$.

2. We say the sequence (x_n) is **Cauchy** if for all $\varepsilon > 0$ there exists an $N_\varepsilon \in \mathbb{N}$ such that

$$n, m \geq N_\varepsilon \text{ implies } \mathrm{d}(x_n, x_m) < \varepsilon.$$

41.1.3 Completeness in (X, d)

In a metric space, every Cauchy sequence is convergence but the converse may not hold. The same thing is true for pseudometric spaces.

Proposition 41.3. *Let (x_n) be a sequence in X, let $x \in X$, and suppose $x_n \to x$. Then (x_n) is Cauchy.*

Proof. Let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies

$$\mathrm{d}(x_n, x) < \varepsilon/2.$$

Then $n, m \geq N$ implies

$$\begin{aligned} \mathrm{d}(x_n, x_m) &\leq \mathrm{d}(x_n, x) + \mathrm{d}(x, x_m) \\ &< \varepsilon/2 + \varepsilon/2 \\ &= \varepsilon. \end{aligned}$$

This implies (x_n) is Cauchy. □

Thus, the concept of completeness makes sense in a pseudometric space.

Definition 41.4. Let (X, d) be a pseudometric space. We say (X, d) is **complete** if every Cauchy sequence in (X, d) is a convergent.

41.2 Metric Obtained by Pseudometric

Unless otherwise specified, we let (X, d) be a pseudometric space throughout the remainder of this section. There is a natural way to obtain a metric space from (X, d) which we now describe as follows: define a relation $\sim$ on X by

$$x \sim y \text{ if and only if } \mathrm{d}(x, y) = 0.$$

Then $\sim$ is an equivalence relation. Indeed, we have reflexivity of $\sim$ since $\mathrm{d}(x, x) = 0$ for all $x \in X$, we have symmetry of $\sim$ since $\mathrm{d}(x, y) = \mathrm{d}(y, x)$ for all $x, y \in X$, and we have transitivity of $\sim$ since d satisfies the triangle inequality: if $x \sim y$ and $y \sim z$, then

$$\begin{aligned} \mathrm{d}(x, z) &\leq \mathrm{d}(x, y) + \mathrm{d}(y, z) \\ &= 0 + 0 \\ &= 0. \end{aligned}$$

Thus $\mathrm{d}(x, z) = 0$ which implies $x \sim z$.

Therefore we may consider the quotient space of X with respect to the equivalence relation above. We shall denote this quotient space by $[X] := X/\sim$. A coset in $[X]$ which is represented by $x \in X$ will be written as $[x]$. There is a natural **projection map** $\pi\colon X \to [X]$ that sends $x \in X$ to its equivalence class $[x]$. Since π is surjective, any subset of $[X]$ has the form

$$[A] = \{[a] \in [X] \mid a \in A\}$$

for some $A \subseteq X$. We are ready now to define the metric on $[X]$.

Theorem 41.1. *Define* $[\mathrm{d}]\colon [X] \times [X] \to \mathbb{R}$ *by*

$$[\mathrm{d}]([x], [y]) = \mathrm{d}(x, y) \tag{104}$$

for all $[x], [y] \in [X]$. *Then* $[\mathrm{d}]$ *is a metric on* $[X]$. *It is called the metric* ***induced*** *by the pseudometric* $[\mathrm{d}]$.

Proof. We first show that (104) is well-defined. Indeed, choose different coset representatives of $[x]$ and $[y]$, say x' and y' respectively (so $\mathrm{d}(x, x') = 0$ and $\mathrm{d}(y, y') = 0$). Then

$$\begin{aligned} [\mathrm{d}]([x'], [y']) &= \mathrm{d}(x', y') \\ &\leq \mathrm{d}(x', x) + \mathrm{d}(x, y) + \mathrm{d}(y, y') \\ &= \mathrm{d}(x, y) \\ &= [\mathrm{d}]([x], [y]). \end{aligned}$$

Thus $[\mathrm{d}]$ is well-defined.

Next we show that $[\mathrm{d}]$ is in fact a metric on $[X]$. First we check $[\mathrm{d}]$ is symmetric. Let $[x], [y] \in [X]$. Then

$$\begin{aligned} [\mathrm{d}]([x], [y]) &= \mathrm{d}(x, y) \\ &= \mathrm{d}(y, x) \\ &= [\mathrm{d}]([y], [x]). \end{aligned}$$

Thus $[\mathrm{d}]$ is symmetric. Next we check $[\mathrm{d}]$ satisfies triangle inequality. Let $[x], [y], [z] \in [X]$. Then

$$\begin{aligned} [\mathrm{d}]([x], [z]) &= \mathrm{d}(x, z) \\ &\leq \mathrm{d}(x, y) + \mathrm{d}(y, z) \\ &= [\mathrm{d}]([x], [y]) + [\mathrm{d}]([y], [z]). \end{aligned}$$

Thus $[\mathrm{d}]$ satisfies triangle inequality. Finally we check $[\mathrm{d}]$ satisfies identify of indiscernables. Let $[x], [y] \in [X]$ and suppose $[\mathrm{d}]([x], [y]) = 0$. Then

$$\begin{aligned} 0 &= [\mathrm{d}]([x], [y]) \\ &= \mathrm{d}(x, y) \end{aligned}$$

implies $x \sim y$ by definition. Therefore $[x] = [y]$. Thus $[\mathrm{d}]$ satisfies identify of indiscernables. $\square$

41.2.1 Completeness in (X, d) is equivalent to completeness in $([X], [\mathrm{d}])$

As in the case of the pseudometric d, the metric [d] induces a topology on $[X]$. We denote this topology by $\tau_{[\mathrm{d}]}$.

Proposition 41.4. *(X, d) is complete if and only if $([X], [\mathrm{d}])$ is complete.*

Proof. Suppose that (X, d) is complete. Let $([x_n])$ be a Cauchy sequence in $([X], [\mathrm{d}])$. We claim (x_n) is a Cauchy sequence in (X, d). Indeed, let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $m, n \geq N$ implies

$$[\mathrm{d}]([x_n], [x_m]) < \varepsilon.$$

Then $m, n \geq N$ implies

$$\begin{aligned} \mathrm{d}(x_n, x_m) &= [\mathrm{d}]([x_n], [x_m]) \\ &< \varepsilon. \end{aligned}$$

This implies (x_n) is a Cauchy sequence in (X, d). Since (X, d) is complete, the sequence converges to a (not necessarily unique) $x \in X$. Then we claim that $[x_n] \to [x]$. Indeed, let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies

$$\mathrm{d}(x_n, x) < \varepsilon.$$

Then $n \geq N$ implies

$$\begin{aligned} [\mathrm{d}]([x_n], [x]) &= \mathrm{d}(x_n, x) \\ &< \varepsilon. \end{aligned}$$

This implies $[x_n] \to [x]$. Thus $([X], [\mathrm{d}])$ is complete.

Conversely, suppose $([X], [\mathrm{d}])$ is complete. Let (x_n) be a Cauchy sequence in (X, d). We claim $([x_n])$ is a Cauchy sequence in $([X], [\mathrm{d}])$. Indeed, let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $m, n \geq N$ implies

$$\mathrm{d}(x_n, x_m) < \varepsilon.$$

Then $m, n \geq N$ implies

$$\begin{aligned} [\mathrm{d}]([x_n], [x_m]) &= \mathrm{d}(x_n, x_m) \\ &< \varepsilon. \end{aligned}$$

This implies (x_n) is a Cauchy sequence in $([X], [\mathrm{d}])$. Since $([X], [\mathrm{d}])$ is complete, the sequence converges to a unique $[x] \in [X]$. We claim that $x_n \to x$ (in fact it converges to any $y \in X$ such that $y \sim x$). Indeed, let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies

$$[\mathrm{d}]([x_n], [x]) < \varepsilon.$$

Then $n \geq N$ implies

$$\begin{aligned} \mathrm{d}(x_n, x) &= [\mathrm{d}]([x_n], [x]) \\ &< \varepsilon. \end{aligned}$$

This implies $x_n \to x$. Thus (X, d) is complete. □

41.3 Quotient Topology

Recall that we view X as a topological space with toplogy τ_{d}; the topology induced by the pseudometric d. It turns out that there are two natural topologies on $[X]$. One such topology is $\tau_{[\mathrm{d}]}$; the topology induced by the metric [d]. The other topology is called the **quotient topology with respect to** $\sim$, and is denoted by $[\tau_{\mathrm{d}}]$, where $[\tau_{\mathrm{d}}]$ is defined by

$$[\tau_{\mathrm{d}}] = \{[A] \subseteq [X] \mid \pi^{-1}([A]) \in \tau_{\mathrm{d}}\}.$$

In other words, we declare a subset $[A]$ of $[X]$ to be $[\tau_{\mathrm{d}}]$-open in $[X]$ if and only if

$$\begin{aligned} \pi^{-1}([A]) &= \{x \in X \mid x \sim a \text{ for some } a \in A\} \\ &= \{x \in X \mid \mathrm{d}(x, a) = 0 \text{ for some } a \in A\} \end{aligned}$$

is open in X. Since $\pi^{-1}(\emptyset) = \emptyset$ and $\pi^{-1}([X]) = X$, we see that both $\emptyset$ and $[X]$ are open in $[X]$. Furthermore, since

$$\pi^{-1}\left(\bigcup_{i \in I}[A_i]\right) = \bigcup_{i \in I} \pi^{-1}([A_i]) \text{ and } \pi^{-1}\left(\bigcap_{i \in I}[A_i]\right) = \bigcap_{i \in I} \pi^{-1}([A_i]),$$

we see that the collection of open sets in $[X]$ is closed under arbitrary unions and finite intersections. Therefore $[\tau_{\mathrm{d}}]$ is indeed a topology on $[X]$. Note that $[\tau_{\mathrm{d}}]$ was defined in such a way that it makes the projection map $\pi\colon X \to [X]$ continous.

41.3.1 Universal Mapping Property For Quotient Space

Quotient spaces satisfy the following universal mapping property.

Proposition 41.5. *Let $f\colon X \to Y$ be any continuous function which is constant on each equivalence class. Then there exists a unique continuous function $[f]\colon [X] \to Y$ such that $f = [f] \circ \pi$.*

Proof. We define $[f]\colon [X] \to Y$ by

$$[f]([x]) = f(x) \tag{105}$$

for all $x \in X$. We first show that (105) is well-defined. Suppose x and x' are two different representatives of the same coset (so $x \sim x'$). Then $f(x) = f(x')$ as f was assumed to be constant on equivalence classes, and so

$$\begin{aligned} [f]([x']) &= f(x') \\ &= f(x) \\ &= [f]([x]). \end{aligned}$$

Thus (105) is well-defined.

Next we want to show that $[f]$ is continuous. Let V be an open set in Y. Then

$$f^{-1}(V) = \pi^{-1}\left([f]^{-1}(V)\right)$$

is open in X. By the definition of quotient topology, this implies $[f]^{-1}(V)$ is open in $[X]$. This implies $[f]$ is continuous.

Finally, we want to show that $f = [f] \circ \pi$ holds. Let $x \in X$. Then we have

$$\begin{aligned} ([f] \circ \pi)(x) &= [f](\pi(x)) \\ &= [f]([x]) \\ &= f(x). \end{aligned}$$

It follows that $[f] \circ \pi = f$. This establishes existence of f.

For uniqueness, assume for a contradiction that $\overline{f}\colon [X] \to Y$ is a continuous function such that $f = \overline{f} \circ \pi$ and such that $\overline{f} \neq [f]$. Choose $[x] \in [X]$ such that $\overline{f}[x] \neq [f][x]$. Then

$$\begin{aligned} f(x) &= (\overline{f} \circ \pi)(x) \\ &= \overline{f}(\pi(x)) \\ &= \overline{f}([x]) \\ &\neq [f]([x]) \\ &= f(x), \end{aligned}$$

which gives us a contradiction. □

It follows from Proposition (41.5) that we have the following bijection of sets

$$\{f\colon X \to Y \mid f \text{ is continuous and constant on equivalence classes}\} \cong \{\text{continuous functions from } [X] \text{ to } Y\}.$$

In particular, if we want to study continuous functions out of $[X]$, then we just need to study the continuous functions out of X which are constant on equivalence classes.

Proposition 41.6. *Suppose (Y, d_Y) is a metric space and $f\colon (X, \mathrm{d}) \to (Y, \mathrm{d}_Y)$ is continuous. The f is constant on equivalence classes.*

Proof. Let $x, x' \in X$ such that $x \sim x'$. Thus $\mathrm{d}(x, x') = 0$. Let $\varepsilon > 0$. Choose $\delta > 0$ such that

$$\mathrm{d}(x, y) < \delta \text{ implies } \mathrm{d}_Y(f(x), f(y)) < \varepsilon.$$

We want to show that $f(x) = f(x')$. □

41.3.2 Open Equivalence Relation

An equivalence relation $\sim$ on a topological space X is said to be **open** if the projection map $\pi\colon X \to [X]$ is open. In other words, the equivalence relation $\sim$ on X is open if and only if for every open set U in X, the set

$$\pi^{-1}(\pi(U)) = \bigcup_{x\in U}[x]$$

of all points equivalent to some point of U is open. The imporance of open equivalence relations is that if $\mathscr{B}$ is a basis for X, then $[\mathscr{B}]$ is a basis for $[X]$.

Lemma 41.2. *Let $x \in X$ and $r > 0$. Then*

$$\mathrm{B}_r(x) = \pi^{-1}([\mathrm{B}_r(x)]).$$

In particular, π is an open mapping.

Proof. We have

$$\begin{aligned}\mathrm{B}_r(x) &\subseteq \pi^{-1}\left(\pi\left(\mathrm{B}_r(x)\right)\right)\\ &= \pi^{-1}([\mathrm{B}_r(x)]).\end{aligned}$$

For the reverse inclusion, let $y \in \pi^{-1}([\mathrm{B}_r(x)])$. Then $\mathrm{d}(y,z) = 0$ for some $z \in \mathrm{B}_r(x)$. Choose such a $z \in \mathrm{B}_r(x)$. Then

$$\begin{aligned}\mathrm{d}(y,x) &\leq \mathrm{d}(y,z) + \mathrm{d}(z,x)\\ &= \mathrm{d}(z,x)\\ &< r\end{aligned}$$

implies implies $y \in \mathrm{B}_r(x)$. Therefore

$$\pi^{-1}([\mathrm{B}_r(x)]) \subseteq \mathrm{B}_r(x).$$

Thus each subset in $[X]$ of the form $[\mathrm{B}_r(x)]$ is open in $[X]$.

To see that π is an open mapping, let U be an open set in X. Since the set of all open balls is a basis for τ_{d}, we can cover U by open balls, say

$$U = \bigcup_{i\in I}\mathrm{B}_{r_i}(x_i).$$

Then

$$\begin{aligned}\pi(U) &= \pi\left(\bigcup_{i\in I}\mathrm{B}_{r_i}(x_i)\right)\\ &= \bigcup_{i\in I}\pi\left(\mathrm{B}_{r_i}(x_i)\right)\\ &= \bigcup_{i\in I}[\mathrm{B}_{r_i}(x_i)]\\ &\in [\tau_{\mathrm{d}}].\end{aligned}$$

Thus π is an open mapping. □

41.3.3 Quotient Topology Agrees With Metric Topology

Theorem 41.3. *With the notation as above, we have*

$$[\tau_{\mathrm{d}}] = \tau_{[\mathrm{d}]}.$$

Proof. We first note that for each $x \in X$ and $r > 0$, we have

$$\begin{aligned}[\mathrm{B}_r(x)] &= \{[y] \in [X] \mid y \in \mathrm{B}_r(x)\}\\ &= \{[y] \in [X] \mid \mathrm{d}(y,x) < r\}\\ &= \{[y] \in [X] \mid [\mathrm{d}]([y],[x]) < r\}\\ &= \mathrm{B}_r([x]).\end{aligned}$$

In particular, $\tau_{[\mathrm{d}]}$ and $[\tau_{\mathrm{d}}]$ share a common basis. Therefore $\tau_{[\mathrm{d}]} = [\tau_{\mathrm{d}}]$. □

42 Completing a Normed Linear Space

Recall from linear analysis that if $(\mathcal{X}, \|\cdot\|)$ is a normed linear space, then $\mathcal{X}$[18] may or may not be complete with respect to the topology induced by $\|\cdot\|$. In other words, there may exist a Cauchy sequence in $\mathcal{X}$ which does not converge in $\mathcal{X}$. For instance, $(\mathbb{Q}, |\cdot|)$ is not complete. Indeed, the sequence

$$3, 3.1, 3.14, 3.145, \dots$$

is Cauchy but does not converge in $\mathbb{Q}$; it converges to π which is irrational. All is not lost however because we can always **complete** a normed linear space:

Theorem 42.1. *Every normed linear space can be completed. More precisely, if $(X, \|\cdot\|)$ is a normed linear space, then there exists a normed linear space $(\widetilde{\mathcal{X}}, \widetilde{\|\cdot\|})$ such that*

1. *$(\widetilde{\mathcal{X}}, \widetilde{\|\cdot\|})$ is a complete normed linear space (that is, it is a Banach space).*
2. *There exists a isometric embedding $\iota\colon (\mathcal{X}, \|\cdot\|) \to (\widetilde{\mathcal{X}}, \widetilde{\|\cdot\|})$ with dense image. This means that*
   (a) *ι is a linear map which preserves norms: $\|\iota(x)\| = \widetilde{\|x\|}$ for all $x \in \mathcal{X}$. In particular, this implies ι is injective;*
   (b) *the closure of $\iota(\mathcal{X})$ in $\widetilde{\mathcal{X}}$ is equal to all of $\widetilde{\mathcal{X}}$.*

*We call $(\widetilde{\mathcal{X}}, \widetilde{\|\cdot\|})$ the **completion** of $(\mathcal{X}, \|\cdot\|)$.*

42.1 Constructing Completions

Let $(\mathcal{X}, \|\cdot\|)$ be a normed linear space. We denote by $\mathscr{C}_{\mathcal{X}}$ to be the set of all Cauchy sequences in $\mathcal{X}$. We can give $\mathscr{C}_{\mathcal{X}}$ the structure of a $\mathbb{C}$-vector space as follows: let $(x_n), (y_n) \in \mathscr{C}_{\mathcal{X}}$ and let $\lambda, \mu \in \mathbb{C}$. Then we define

$$\lambda(x_n) + \mu(y_n) := (\lambda x_n + \mu y_n). \tag{106}$$

One easily checks that scalar multiplication and addition defined as in (106) gives $\mathscr{C}_{\mathcal{X}}$ the structure of a $\mathbb{C}$-vector space.

42.1.1 Seminorm

A natural contender for a norm on $\mathscr{C}_{\mathcal{X}}$ is the map $\|\cdot\|\colon \mathscr{C}_{\mathcal{X}} \to \mathbb{C}$ defined by

$$\|(x_n)\| := \lim_{n\to\infty} \|x_n\| \tag{107}$$

for all $(x_n) \in \mathscr{C}_{\mathcal{X}}$. In fact, (107) will not be a norm, but rather a seminorm[19]. Before we explain this however, let us first show that the righthand side of (107) converges in $\mathbb{C}$.

Proposition 42.1. *Let (x_n) be a Cauchy sequence of vectors in $\mathcal{X}$. Then $(\|x_n\|)$ is a Cauchy sequence of complex numbers in $\mathbb{C}$. In particular, (107) converges in $\mathbb{C}$.*

Proof. Let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $m, n \geq N$ implies

$$\|x_n - x_m\| < \varepsilon.$$

Then $m, n \geq N$ implies

$$\begin{aligned} |\|x_n\| - \|x_m\|| &\leq \|x_n - x_m\| \\ &< \varepsilon. \end{aligned}$$

It follows that $(\|x_n\|)$ is a Cauchy sequence of complex numbers in $\mathbb{C}$. The latter statement in the proposition follows from the fact that $\mathbb{C}$ is complete. □

[18] To simplify notation, we often write $\mathcal{X}$ rather than $(\mathcal{X}, \|\cdot\|)$. Context will make it clear which norm we are equipping $\mathcal{X}$ with.
[19] See Appendix for notes on pseudonorms

42.1.2 Quotienting Out To get an Inner-Product

As mentioned above, (107) is not a norm. It is what's called a seminorm:

Definition 42.1. Let V be a vector space over $\mathbb{C}$. A map $\|\cdot\|: V \to \mathbb{C}$ is called a **seminorm** if it satisfies the following properties:

1. Absolute homogeneity: $\|\lambda x\| = |\lambda|\|x\|$ for all $x \in V$ and $\lambda \in \mathbb{C}$;
2. Subadditivity: $\|x+y\| \leq \|x\| + \|y\|$ for all $x, y \in V$;
3. Semipositive definiteness: $\|x\| \geq 0$ for all $x \in V$.

To see why (107) is a seminorm, note that absolute homogeneity and subadditivity are clear. What makes (107) a seminorm and not a norm is that we have only have semipositive definiteness:

$$\begin{aligned}\|(x_n)\| &= \lim_{n\to\infty} \|x_n\| \\ &\geq 0.\end{aligned}$$

In particular, we may have $\|x_n\| \to 0$ with $(x_n) \neq 0$. To remedy this situation, we define

$$\mathscr{C}_{\mathcal{X}}^0 := \{(x_n) \in \mathscr{C}_{\mathcal{X}} \mid \|x_n\| \to 0\}.$$

Then $\mathscr{C}_{\mathcal{X}}^0$ is a subspace of $\mathscr{C}_{\mathcal{X}}$. Indeed, if $\lambda, \mu \in \mathbb{C}$ and $(x_n), (y_n) \in \mathscr{C}_{\mathcal{X}}^0$, then

$$\begin{aligned}\|\lambda x_n + \mu y_n\| &\leq \|\lambda x_n\| + \|\mu y_n\| \\ &= |\lambda|\|x_n\| + |\mu|\|y_n\| \\ &\to 0\end{aligned}$$

and hence $(\lambda x_n + \mu y_n) \in \mathscr{C}_{\mathcal{X}}^0$. Therefore we obtain a quotient space $\mathscr{C}_{\mathcal{X}}/\mathscr{C}_{\mathcal{X}}^0$. Now we claim that the pseudo-norm (107) induces a genuine norm, which we denote again by $\|\cdot\|$, on $\mathscr{C}_{\mathcal{X}}/\mathscr{C}_{\mathcal{X}}^0$, defined by

$$\|\overline{(x_n)}\| := \lim_{n\to\infty} \|x_n\|. \tag{108}$$

for all $\overline{(x_n)}$[20] in $\mathscr{C}_{\mathcal{X}}/\mathscr{C}_{\mathcal{X}}^0$. We need to be sure that (108) is well-defined. Let (x_n') be different representative of the cosets $\overline{(x_n)}$ (so $\|x_n' - x_n\| \to 0$). Then

$$\begin{aligned}\|(x_n')\| &= \lim_{n\to\infty} \|x_n'\| \\ &= \lim_{n\to\infty} \|x_n' + x_n - x_n\| \\ &\leq \lim_{n\to\infty} \|x_n\| + \lim_{n\to\infty} \|x_n' - x_n\| \\ &= \lim_{n\to\infty} \|x_n\| \\ &= \|(x_n)\|.\end{aligned}$$

Thus (108) is well-defined (meaning it is independent of the choice of representatives of cosets).

Now absolute homogeneity of (108) and subadditivity of (108) are clear. This time however, we also have positive definiteness: if $\overline{(x_n)} \in \mathscr{C}_{\mathcal{X}}/\mathscr{C}_{\mathcal{X}}^0$ such that $\|\overline{(x_n)}\| = 0$, then

$$\begin{aligned}\lim_{n\to\infty} \|x_n\| &= \|\overline{(x_n)}\| \\ &= 0,\end{aligned}$$

which implies $\overline{(x_n)} = 0$ in $\mathscr{C}_{\mathcal{X}}/\mathscr{C}_{\mathcal{X}}^0$.

[20] When we write $\overline{(x_n)}$ for a coset in $\mathscr{C}_{\mathcal{X}}/\mathscr{C}_{\mathcal{X}}^0$, then it is implicitly understood that (x_n) is an element $\mathscr{C}_{\mathcal{X}}$ which represents the coset $\overline{(x_n)}$ in $\mathscr{C}_{\mathcal{X}}/\mathscr{C}_{\mathcal{X}}^0$.

42.1.3 The map $\iota\colon \mathcal{X} \to \mathscr{C}_{\mathcal{X}}/\mathscr{C}^0_{\mathcal{X}}$

Let $\iota\colon \mathcal{X} \to \mathscr{C}_{\mathcal{X}}/\mathscr{C}^0_{\mathcal{X}}$ be defined by

$$\iota(x) = \overline{(x)}$$

for all $x \in \mathcal{X}$, where (x) is a constant sequence in $\mathscr{C}_{\mathcal{X}}$.

Definition 42.2. An **isometry** between normed linear space $(\mathcal{X}_1, \|\cdot\|_1)$ and $(\mathcal{X}_2, \|\cdot\|_2)$ is an operator $T\colon \mathcal{X}_1 \to \mathcal{X}_2$ such that

$$\|Tx\|_2 = \|x\|_1$$

for all $x \in \mathcal{X}_1$.

Remark 58. Note that an isometry is automatically injective. Indeed, let $x \in \ker T$. Then

$$\begin{aligned}\|x\| &= \|Tx\| \\ &= \|0\| \\ &= 0\end{aligned}$$

implies $x = 0$. Thus we sometimes call an isometry an **isometric embedding**.

Proposition 42.2. *The map $\iota\colon \mathcal{X} \to \mathscr{C}_{\mathcal{X}}/\mathscr{C}^0_{\mathcal{X}}$ is an isometry.*

Proof. Linearity of ι is clear. Let $x \in \mathcal{X}$. Then

$$\begin{aligned}\|\iota(x)\| &= \|\overline{(x)}\| \\ &= \lim_{n\to\infty} \|x\| \\ &= \|x\|.\end{aligned}$$

Thus ι is an isometry. □

Proposition 42.3. *The image of $\mathcal{X}$ under ι is dense in $\mathscr{C}_{\mathcal{X}}/\mathscr{C}^0_{\mathcal{X}}$. In other words, the closure of $\iota(\mathcal{X})$ in $\mathscr{C}_{\mathcal{X}}/\mathscr{C}^0_{\mathcal{X}}$ is all of $\mathscr{C}_{\mathcal{X}}/\mathscr{C}^0_{\mathcal{X}}$.*

Proof. Let $\overline{(x_n)}$ be a coset in $\mathscr{C}_{\mathcal{X}}/\mathscr{C}^0_{\mathcal{X}}$. To show that the closure of $\iota(\mathcal{X})$ is is all of $\mathscr{C}_{\mathcal{X}}/\mathscr{C}^0_{\mathcal{X}}$, we construct a sequence of cosets in $\iota(\mathcal{X})$ which converges to $\overline{(x_n)}$. Let $\varepsilon > 0$ and choose $N \in \mathbb{N}$ such that $n, m \geq N$ implies

$$\|x_n - x_m\| < \varepsilon/2.$$

Then $n, m \geq N$ implies

$$\begin{aligned}\|\iota(x_m) - \overline{(x_n)}\| &= \lim_{n\to\infty} \|x_m - x_n\| \\ &< \lim_{n\to\infty} \varepsilon \\ &= \varepsilon.\end{aligned}$$

Thus, $(\iota(x_m))$ is a sequence of cosets in $\iota(\mathcal{X})$ which converges to $\overline{(x_n)}$. □

42.1.4 $\mathscr{C}_{\mathcal{X}}/\mathscr{C}^0_{\mathcal{X}}$ is Complete

Proposition 42.4. *$\mathscr{C}_{\mathcal{X}}/\mathscr{C}^0_{\mathcal{X}}$ is a Banach space.*

Proof. Let $(\overline{x}^n)$ be a Cauchy sequence of cosets in $\mathscr{C}_{\mathcal{X}}/\mathscr{C}^0_{\mathcal{X}}$ where

$$\overline{x}^n = \overline{(x^n_k)}_{k\in\mathbb{N}}$$

for each $n \in \mathbb{N}$. Throughout the remainder of this proof, let $\varepsilon > 0$.

Since each $x^n = (x^n_k)_{k\in\mathbb{N}}$ is a Cauchy sequence of elements in $\mathcal{X}$, there exists a $\pi(n) \in \mathbb{N}$ such that $k, l \geq \pi(n)$ implies

$$\|x^n_k - x^n_l\| < \frac{\varepsilon}{3}.$$

For each $n \in \mathbb{N}$, choose such $\pi(n) \in \mathbb{N}$ in such a way so $\pi(n) \geq \pi(m)$ whenever $n \geq m$.

Step 1: We show that the sequence $(x^n_{\pi(n)})$ of elements in $\mathcal{X}$ is a Cauchy sequence. Since $(\overline{x}^n)$ is a Cauchy sequence of cosets in $\mathscr{C}_{\mathcal{X}}/\mathscr{C}^0_{\mathcal{X}}$, there exists an $N \in \mathbb{N}$ such that $n \geq m \geq N$ implies

$$\|(\overline{x}^n) - (\overline{x}^m)\| = \lim_{k\to\infty} \|x^n_k - x^m_k\| < \frac{\varepsilon}{4}. \tag{109}$$

Choose such an $N \in \mathbb{N}$. It follows from (109) that for each $n \geq m \geq N$, there exists $\pi(n,m) \geq \pi(n)$ such that

$$\|x^n_k - x^m_k\| < \frac{\varepsilon}{3}$$

for all $k \geq \pi(n,m)$. Choose such $\pi(n,m)$ for each $n \geq m \geq N$. Then if $n \geq m \geq N$, we have

$$\begin{aligned}
\|x^n_{\pi(n)} - x^m_{\pi(m)}\| &= \|x^n_{\pi(n)} - x^n_{\pi(n,m)} + x^n_{\pi(n,m)} - x^m_{\pi(n,m)} + x^m_{\pi(n,m)} - x^m_{\pi(m)}\| \\
&\leq \|x^n_{\pi(n)} - x^n_{\pi(n,m)}\| + \|x^n_{\pi(n,m)} - x^m_{\pi(n,m)}\| + \|x^m_{\pi(n,m)} - x^m_{\pi(m)}\| \\
&< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 \\
&= \varepsilon.
\end{aligned}$$

Therefore $(x^n_{\pi(n)})$ is a Cauchy sequence of elements in $\mathcal{X}$ and hence represents a coset $\overline{(x^n_{\pi(n)})}$ in $\mathscr{C}_{\mathcal{X}}/\mathscr{C}^0_{\mathcal{X}}$.

Step 2: Let $x = (x^k_{\pi(k)})$[21]. We want to show that the sequence $(\overline{x}^n)$ of cosets in $\mathscr{C}_{\mathcal{X}}/\mathscr{C}^0_{\mathcal{X}}$ converges to the coset $\overline{x}$ in $\mathscr{C}_{\mathcal{X}}/\mathscr{C}^0_{\mathcal{X}}$. In particular, we need to find an $N \in \mathbb{N}$ such that $n \geq N$ implies

$$\|\overline{x}^n - \overline{x}\| = \lim_{k\to\infty} \|x^n_k - x^k_{\pi(k)}\| < \varepsilon$$

or in other words, $n \geq N$ implies

$$\|x^n_k - x^k_{\pi(k)}\| < \varepsilon$$

for all k sufficiently large. Since x is a Cauchy sequence of elements in $\mathcal{X}$, there exists an $M \in \mathbb{N}$ such that $n, k \geq M$ implies

$$\|x^n_{\pi(n)} - x^k_{\pi(k)}\| < 2\varepsilon/3.$$

Choose such an $M \in \mathbb{N}$. Then $n \geq M$ implies

$$\begin{aligned}
\|x^n_k - x^k_{\pi(k)}\| &\leq \|x^n_k - x^n_{\pi(n)}\| + \|x^n_{\pi(n)} - x^k_{\pi(k)}\| \\
&< \frac{\varepsilon}{3} + \frac{2\varepsilon}{3} \\
&= \varepsilon.
\end{aligned}$$

for all $k \geq \max\{M, \pi(n)\}$. □

Part IV
Functional Analysis

43 Introduction

Given a measure μ, the nth **moment** is by definition $\int_I t^n \mathrm{d}\mu(t)$ where Ij is a subinterval of $\mathbb{R}$. The moment problem says that if we are given a sequence (a_n) of real numbers, can we find a measure μ such that

$$a_n = \int_I t^n \mathrm{d}\mu(t).$$

[21] Note the change in index from n to k.

for all $n \in \mathbb{N}$. If $I = [0,1]$, then this is called the Hausdorff moment problem. If $I = [0,\infty)$, then this is called the Stieltjes moment problem. If $I = (-\infty,\infty)$, then this is called the Hamburger moment problem.

Let us start with some intuition on how we can solve this problem. For a function f and a measure μ, let us denote

$$\langle f, \mu \rangle = \int_I f \mathrm{d}\mu \tag{110}$$

In some sense, (110) behaves like an inner-product. Of course, f and μ are different types of mathematical objects; one is a function and the other is a measure. So for all functions f and measures μ.

43.1 Convex Sets

Proof. Let V be an $\mathbb{R}$-vector space and let C be a subset of V. We say C is **convex** if for all $t \in (0,1)$ and $x, y \in S$, we have $tx + (1-t)y \in C$. □

Proposition 43.1. *Let V be an $\mathbb{R}$-vector space and let C be a convex subset of V. Then for all $n \in \mathbb{N}$, $x_1, \dots, x_n \in C$, and $t_1, \dots, t_n \in (0,1)$ such that $\sum_{i=1}^n t_i = 1$, we have $\sum_{i=1}^n t_i x_i \in C$.*

Proof. Let $x = \sum_{i=1}^n t_i x_i$ and assume that n is minimal in the sense that if $x = \sum_{i'=1}^{n'} t'_{i'} x_{i'}$ is another representation of x, where each $x_{i'} \in S$ and $t'_{i'} \in (0,1)$ such that $\sum_{i'=1}^{n'} t'_{i'} = 1$, then we must have $n \leq n'$. Assume for a contradiction that $x \notin C$, so $n > 2$. Then observe that

$$\begin{aligned}
\sum_{i=1}^{n-1} \frac{t_i}{1-t_n}((1-t_n)x_i + t_n x_n) &= \sum_{i=1}^{n-1} t_i x_i + \left(\sum_{i=1}^{n-1} \frac{t_i}{1-t_n} \right) t_n x_n \\
&= \sum_{i=1}^{n-1} t_i x_i + \left(\frac{1-t_n}{1-t_n} \right) t_n x_n \\
&= \sum_{i=1}^{n-1} t_i x_i + t_n x_n \\
&= \sum_{i=1}^{n} t_i x_i \\
&= x,
\end{aligned}$$

gives another representation with $n-1$ terms, a contradiction. □

43.1.1 Convex Closure and Closed Convex Closure

Definition 43.1. Let V be an $\mathbb{R}$-vector space and let S be a subset of V. The **convex closure** of S is defined by

$$\operatorname{conv}(S) = \{ tx + (1-t)y \mid t \in (0,1) \text{ and } x, y \in S \}.$$

Moreover, suppose $\|\cdot\|$ is a norm on V, so that $(V, \|\cdot\|)$ is a normed linear space. The **closed convex closure** of S is defined to be the smallest closed convex set which contains S and is denoted by $\overline{\operatorname{conv}}(S)$.

Proposition 43.2. *With the notation as above,* $\operatorname{conv}(S)$ *is the smallest convex set which contains S. Furthermore, we have* $\overline{\operatorname{conv}(S)} = \overline{\operatorname{conv}}(S)$.

Proof. Let us first show that $\operatorname{conv}(S)$ is in fact a convex set. Let $s, t, t' \in (0,1)$ and let $x, x', y, y' \in S$. Then observe that

$$s(tx + (1-t)y) + (1-s)(t'x' + (1-t')y') = stx + s(1-t)y + (1-s)t'x' + (1-s)(1-t')y' \in \operatorname{conv}(S),$$

where we used Proposition (43.1) together with the fact that

$$st + s(1-t) + (1-s)t' + (1-s)(1-t') = 1.$$

It follows that $\operatorname{conv}(S)$ is convex. It is also the smallest convex set which contains S since if C is a convex set which contains S, then we must have $tx + (1-t)y \in C$ for all $t \in (0,1)$ and $x, y \in S$, which implies $\operatorname{conv}(S) \subseteq C$.

Now we will show $\overline{\text{conv}(S)} = \overline{\text{conv}}(S)$. To see this, first note that since $\overline{\text{conv}}(S)$ is convex, we have $\text{conv}(S) \subseteq \overline{\text{conv}}(S)$, and hence

$$\begin{aligned} \overline{\text{conv}(S)} &\subseteq \overline{\overline{\text{conv}}(S)} \\ &= \overline{\text{conv}}(S). \end{aligned}$$

For the reverse inclusion, it suffices to show that $\overline{\text{conv}(S)}$ is convex, since then $\overline{\text{conv}(S)}$ would be a closed convex set, and so $\overline{\text{conv}}(S) \subseteq \overline{\text{conv}(S)}$ by definition of $\overline{\text{conv}}(S)$. In fact, we will show that the closure of a convex set is convex. To this end, suppose C is a convex set and let $t \in (0,1)$ and $x, y \in \overline{C}$. Choose sequences (x_n) and (y_n) in C such that $x_n \to x$ and $y_n \to y$. Then $(tx_n + (1-t)y_n)$ is a sequence in C (as C is convex) which converges to $tx + (1-t)y$. It follows that $tx + (1-t)y \in \overline{C}$, and hence $\overline{C}$ is convex. □

43.1.2 Convex Closure Preserves Minkowski Sum

Definition 43.2. Let V be an $\mathbb{R}$-vector space and let S_1, S_2 be subsets of V. We define the **Minkowski sum** of S_1 and S_2 to be the set

$$S_1 + S_2 = \{x_1 + x_2 \mid x_1 \in S_1 \text{ and } x_2 \in S_2\}.$$

Proposition 43.3. *Let V be an $\mathbb{R}$-vector space and let C_1, C_2 be convex subsets of V. Then $C_1 + C_2$ is convex.*

Proof. Let $t \in (0,1)$, let $c_1, c_1' \in C_1$, and let $c_2, c_2' \in C_2$. Then we have

$$t(c_1 + c_2) + (1-t)(c_1' + c_2') = (tc_1 + (1-t)c_1') + (tc_2 + (1-t)c_2') \in C_1 + C_2,$$

since both C_1 and C_2 are convex. It follows that $C_1 + C_2$ is convex. □

Proposition 43.4. *Let V be an $\mathbb{R}$-vector space and let S_1, S_2 be subsets of V. Then we have*

$$\text{conv}(S_1 + S_2) = \text{conv}(S_1) + \text{conv}(S_2).$$

Proof. Note that $\text{conv}(S_1) + \text{conv}(S_2)$ is a convex set which contains $S_1 + S_2$. Thus

$$\text{conv}(S_1 + S_2) \subseteq \text{conv}(S_1) + \text{conv}(S_2).$$

For the reverse inclusion, let $z_1 \in \text{conv}(S_1)$ and $z_2 \in \text{conv}(S_2)$ and express them as $z_1 = t_1x_1 + (1-t_1)y_1$ and $z_2 = t_2x_2 + (1-t_1)y_2$ where $x_1, y_1 \in S_1$, $x_2, y_2 \in S_2$, and $t_1, t_2 \in (0,1)$. Then note that

$$\begin{aligned} z_1 + z_2 &= t_1x_1 + (1-t_1)y_1 + t_2x_2 + (1-t_2)y_2 \\ &= t_1x_1 + t_2x_2 + y_1 - t_1y_1 + y_2 - t_2y_2 \\ &= (t_1 - t_2)(x_1 + y_2) + t_2(x_1 + x_2) + (1-t_1)(y_1 + y_2), \end{aligned}$$

where $(t_1 - t_2) + t_2 + (1 - t_1) = 1$ and where $x_1 + y_2, x_1 + x_2, y_1 + y_2 \in S_1 + S_2$. It follows that $z_1 + z_2 \in \text{conv}(S_1 + S_2)$. Thus we have the reverse inclusion

$$\text{conv}(S_1 + S_2) \supseteq \text{conv}(S_1) + \text{conv}(S_2).$$

□

43.2 Convex Cones

Definition 43.3. Let V be an $\mathbb{R}$-vector space. A set $P \subseteq V$ is said to be a **convex cone** if

1. if $x, y \in P$ then $x + y \in P$
2. if $x \in P$ and $\alpha \geq 0$, then $\alpha x \in P$.

Given a convex cone $P \subseteq V$, we can define a partial order on V as follows: if $x, y \in V$, then we say $x \leq_P y$ if $y - x \in P$. To see that this is a preorder, note that reflexivity of $\leq_P$ follows from the fact that $0 \in P$. Transitivity of $\leq_P$ follows from the fact that P is closed under addition: if $x \leq_P y$ and $y \leq_P z$, then $z - x = (z-y) + (y-x) \in P$ shows $x \leq_P z$. Thus $\leq_P$ is in fact a preorder. If we assume in addition that $-P \cap P = 0$, then we also have antisymmetry of $\leq_P$. In this case, $\leq_P$ is a partial order. Note that, we will have $0 \leq_P x$ for all $x \in P$, thus it makes sense to call the elements of P the **positive** elements with respect to the preorder $\leq_P$.

43.3 Marcel Riesz Extension Theorem

Theorem 43.1. *(Marcel Riesz Extension Theorem) Let V be an $\mathbb{R}$-vector space, let $W \subseteq V$ be a subspace of V, and let $P \subseteq V$ be a convex cone. Suppose $V = W + P$ and $\psi\colon W \to \mathbb{R}$ is a linear functional such that $\psi|_{P\cap W} \geq 0$. Then there exists a linear functional $\widetilde{\psi}\colon V \to \mathbb{R}$ such that $\widetilde{\psi}|_W = \psi$ and $\widetilde{\psi}|_P \geq 0$.*

Proof. Let $v \in V \backslash W$. We will first show that we can extend ψ to a linear functional $\widetilde{\psi}\colon W + \mathbb{R}v \to \mathbb{R}$ such that $\widetilde{\psi}$ preserves the positivity condition. Define two sets $A = \{x \in W \mid -x \leq_P v\}$ and $B = \{y \in W \mid v \leq_P y\}$. Note that A and B are nonempty since $V = W + P$. We claim that

$$\sup\{-\psi(x) \mid x \in A\} \leq \inf\{\psi(y) \mid y \in B\}. \tag{111}$$

Indeed, let $x \in A$ and let $y \in B$. Then note that $-x \leq_P v \leq_P y$ implies $x + y \in C$. It follows that

$$\begin{aligned} 0 &\leq \psi(x+y) \\ &= \psi(x) + \psi(y). \end{aligned}$$

In other words, $-\psi(x) \leq \psi(y)$, which implies (111).

We set $\widetilde{\psi}(v)$ to be any number between $\sup\{-\psi(x) \mid x \in A\}$ and $\inf\{\psi(y) \mid y \in B\}$ and we define we define $\widetilde{\psi}\colon W + \mathbb{R}v \to \mathbb{R}$ by

$$\widetilde{\psi}(w + \lambda v) = \psi(w) + \lambda\widetilde{\psi}(v) \tag{112}$$

for all $w + \lambda v \in W + \mathbb{R}v$. Note that (112) is well-defined since v is linearly independent from W. It is easy to check that (112) gives us a linear functional $\widetilde{\psi}\colon W + \mathbb{R}v \to \mathbb{R}$ such that $\widetilde{\psi}|_W = \psi$. Furthermore we have

$$-\psi(x) \leq \widetilde{\psi}(v) \leq \psi(y)$$

for all $x \in A$ and $y \in B$. The only thing left is to check that $\widetilde{\psi}$ satisfies the positivity condition. Let $w + \lambda v \in P \cap (W + \mathbb{R}v)$. We consider the following cases:

Case 1: Assume that $\lambda > 0$. Then note that $(1/\lambda)w + v = (1/\lambda)(w + \lambda v) \in P$ since P is a convex cone. This implies $(1/\lambda)w \in A$. Thus

$$\begin{aligned} 0 &\leq \lambda(\psi((1/\lambda)w) + \widetilde{\psi}(v)) \\ &= \psi(w) + \lambda\widetilde{\psi}(v) \\ &= \widetilde{\psi}(w + \lambda v). \end{aligned}$$

Case 2: Assume that $\lambda < 0$. Then note that $(-1/\lambda)w - v = (-1/\lambda)(w + \lambda v) \in P$ since P is a convex cone. This implies $(-1/\lambda)w \in B$. Thus

$$\begin{aligned} 0 &\leq -\lambda(\psi((-1/\lambda)w) - \widetilde{\psi}(v)) \\ &= \psi(w) + \lambda\widetilde{\psi}(v) \\ &= \widetilde{\psi}(w + \lambda v). \end{aligned}$$

Case 3: Assume that $\lambda = 0$. Then $w \in P \cap W$, and hence $0 \leq \psi(w) = \widetilde{\psi}(w)$.

In all three cases, we see that the positivity condition is satisfied. Thus we can extend ψ to a linear functional on $W + \mathbb{R}v$ while preserving the positivity condition.

Now to extend ψ to all of V, we must appeal to Zorn's Lemma. More specifically, we define a partially ordered set $(\mathcal{F}, \leq)$ as follows: the underlying set $\mathcal{F}$ is given by

$$\mathcal{F} = \{\text{linear functionals } \psi'\colon W' \to \mathbb{R} \mid W' \supseteq W,\ \psi'|_W = \psi, \text{ and } \psi'|_{W'\cap C = P} \geq 0\}.$$

A member of $\mathcal{F}$ is denoted by an ordered pair: (ψ', W'). If (ψ_1, W_1) and (ψ_2, W_2) are two members of $\mathcal{F}$ then we say $(\psi_1, W_1) \leq (\psi_2, W_2)$ if $W_1 \subseteq W_2$ and $\psi_2|_{W_1} = \psi_1$. Observe that every totally ordered subset in $(\mathcal{F}, \leq)$ has an upper bound. Indeed, suppose $\{(\psi_i, W_i)\}_{i\in I}$ is a totally ordered subset in $(\mathcal{F}, \leq)$. Then if we set $W' = \bigcup_{i\in I} W_i$ and if we define $\psi'\colon W \to \mathbb{R}$ as follows: if $x \in W$, then $x \in W_i$ for some i and we set $\psi'(x) = \psi_i(x)$. Then it is easy to check that (ψ', W') is a member of $\mathcal{F}$ and that it is an upper bound of $\{(\psi_i, W_i)\}_{i\in I}$. Since $\mathcal{F}$ is nonempty (it contains (ψ, W)) and every totally ordered subset of $\mathcal{F}$ has an upper bound, we can apply Zorn's Lemma to obtain a *maximal* element in $(\mathcal{F}, \leq)$. This maximal element *must* be defined on all of V, otherwise we can extend it to a larger subspace as shown above and obtain a contradiction. □

43.4 Hausdorff Moment Problem

Now we consider $\mathcal{M} = C[0,1]$, $\mathcal{N} = P[0,1]$, and $\mathcal{P} = \{\text{nonnegative continuous functions on } [0,1]\}$. Thus $f \in \mathcal{P}$ if and only if $f(x) \geq 0$ for all $x \in [0,1]$. Clearly $\mathcal{P}$ is a convex cone. For $p \in \mathcal{N}$ we write it as

$$p(x) = b_n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0,$$

and we define

$$\psi(p) = b_n a_n + b_{n-1}a_{n-1} + \cdots + b_1 a_1 + b_0 a_0.$$

Note that $\psi(x^i) = a_i$. This is clearly a linear functional on $\mathcal{N}$. The first crucial step is to show $\psi(p) \geq 0$ for all $p \in \mathcal{P} \cap \mathcal{N}$. We'll need to use the following theorem of Bernstein:

Theorem 43.2. *(S. Bernstein) A polynomial p is non-negative on $[0,1]$ if and only if it can be represented as*

$$p(x) = A_0 x^n + A_1 x^{n-1}(1-x) + A_2 x^{n-2}(1-x)^2 + \cdots + A_{n-1}x(1-x)^{n-1} + A_n(1-x)^n$$

with $A_0, A_1, \ldots, A_n \geq 0$.

If $\psi(x^i(1-x)^j) \geq 0$ for all $i, j \geq 0$ then by the previous theorem of Bernstein, we will have $\psi(p) \geq 0$ for all $p \in \mathcal{P} \cap \mathcal{N}$. It turns out that this is a sufficient condition too. We write

$$x^i(1-x)^j = x^i \sum_{k=0}^{j} \binom{j}{k}(-1)^k x^k = \sum_{k=0}^{j}\binom{j}{k}(-1)^k x^{i+k}.$$

Thus

$$\begin{aligned}\psi(x^i(1-x)^j) &= \sum_{k=0}^{j}\binom{j}{k}(-1)^k \psi(x^{i+k}) \\ &= \sum_{k=0}^{j}\binom{j}{k}(-1)^k a_{i+k}.\end{aligned}$$

So we need to impose the condition

$$\sum_{k=0}^{j}\binom{j}{k}(-1)^k a_{i+k} \geq 0$$

for all $i, j \geq 0$. Under this condition, we have that all conditions of the Marcel Riesz extension theorem are satisfied, namely we need to check that $\mathcal{M} = \mathcal{P} + \mathcal{N}$. However this is clear: if $f \in \mathcal{M}$, then f is bounded, say $f \leq M$. Then

$$f = (f - M) + M,$$

where $f - M \in \mathcal{P}$ and $M \in \mathcal{N}$. So applying the Marcel Riesz extension theorem, there exists $\widetilde{\psi} \colon \mathcal{M} \to \mathbb{R}$ such that $\widetilde{\psi}(p) = \psi(p)$ for any polynomial p and $\widetilde{\psi}(f) \geq 0$ whenever $f \in \mathcal{P}$. The final important ingredient is the Riesz Representation Theorem:

43.4.1 Riesz Representation Theorem

Lemma 43.3. *(Dini's Theorem) Let X be a compact topological space and let $(f_n \colon X \to \mathbb{R})$ be an increasing sequence of continuous functions which converges pointwise to a continuous function $f \colon X \to \mathbb{R}$. Then (f_n) converges uniformly to f.*

Proof. Let $\varepsilon > 0$. For each $n \in \mathbb{N}$, let $g_n = f - f_n$ and let $E_n = \{g_n < \varepsilon\}$. Each g_n is continuous and thus each E_n is open. Since (f_n) is increasing, each (g_n) is decreasing, and thus the sequence of sets (E_n) is ascending. Since (f_n) converges pointwise to f, it follows that the collection $\{E_n\}$ forms an open cover of X. By compactness of X, we can choose a finite subcover of $\{E_n\}$, and since (E_n) is ascending, this means that there is an $N \in \mathbb{N}$ such that $E_N = X$. Choosing such an N, we see that $n \geq N$ implies

$$\begin{aligned}\varepsilon &> g_n(x) \\ &= f(x) - f_n(x) \\ &= |f(x) - f_n(x)|\end{aligned}$$

for all $x \in X$. It follows that (f_n) converges uniformly to f. □

Theorem 43.4. *(Riesz Representation Theorem) Let $\ell\colon C[0,1] \to \mathbb{R}$ be a linear functional such that $\ell(f) \geq 0$ for all $f \geq 0$. Then there exists a unique finite (positive) measure μ on $[0,1]$ such that*

$$\ell(f) = \int_0^1 f \mathrm{d}\mu$$

for all $f \in C[0,1]$.

Proof. Uniqueness is clear. Let's prove existence. Let $B[0,1]$ be the space of all bounded functions $f\colon [0,1] \to \mathbb{R}$ and let $N[0,1]$ be the space of all nonnegative bounded functions $f\colon [0,1] \to \mathbb{R}$. Clearly $B[0,1]$ contains $C[0,1]$ as subspace and it is easy to see that $B[0,1] = C[0,1] + N[0,1]$. Indeed, for any bounded function $f \in B[0,1]$ there exists a continuous function $g \in C[0,1]$ such that $g \leq f$. Then

$$f = (f - g) + g$$

where $f - g \in N[0,1]$ and $g \in C[0,1]$. Furthermore, $N[0,1]$ is a convex cone and by assumption we have $\ell(f) \geq 0$ for all $f \in C[0,1] \cap N[0,1]$. So by the Marcel Riesz extension theorem, there exists a linear functional $\widetilde{\ell}\colon B[0,1] \to \mathbb{R}$ such that $\widetilde{\ell}|_{C[0,1]} = \ell$ and $\widetilde{\ell}|_{N[0,1]} \geq 0$. Now we define a measure μ on $\mathcal{B}[0,1]$ by

$$\mu(E) = \widetilde{\ell}(1_E)$$

for each $E \in \mathcal{B}[0,1]$. We next show that μ is a measure. Let (E_n) be a sequence of pairwise disjoint sets in $\mathcal{B}[0,1]$. Then

$$\begin{aligned}\mu\left(\bigcup_{n=1}^{\infty} E_n\right) &= \widetilde{\ell}\left(1_{\bigcup_{n=1}^{\infty} E_n}\right) \\ &=\end{aligned}$$

Observe

$$f_n - f, f - f_n \leq |f_n - f| \leq \|f_n - f\|_{\sup}$$

By the positivity of $\widetilde{\ell}$ we have

$$\widetilde{\ell}(f_n - f), \widetilde{\ell}(f - f_n) \leq \widetilde{\ell}(\|f_n - f\|_{\sup}).$$

Equivalently,

$$|\widetilde{\ell}(f_n - f)| \leq \widetilde{\ell}(\|f_n - f\|_{\sup}) = \|f_n - f\|_{\sup}\widetilde{\ell}(1).$$

Therefore if $f_n \to f$ uniformly. Thus $\widetilde{\ell}$ is continuous with respect to the sup norm.

Now if (f_n) is an increasing sequence which converges pointwise to f, then $f_n \to f$ uniformly (Dini's Theorem). Thus if (f_n) is increasing and converges pointwise to f, then $\widetilde{\ell}(f_n) \to \widetilde{\ell}(f)$. Observe that $(1_{\bigcup_{n=1}^{N} E_n})$ is increasing and converges pointwise to $1_{\bigcup_{n=1}^{\infty} E_n}$. It follows that

$$\begin{aligned}\mu\left(\bigcup_{n=1}^{\infty} E_n\right) &= \widetilde{\ell}\left(1_{\bigcup_{n=1}^{\infty} E_n}\right) \\ &= \lim_{N\to\infty} \widetilde{\ell}\left(1_{\bigcup_{n=1}^{N} E_n}\right) \\ &= \lim_{N\to\infty} \widetilde{\ell}\left(\sum_{n=1}^{N} 1_{E_n}\right) \\ &= \lim_{N\to\infty} \sum_{n=1}^{N} \widetilde{\ell}(E_n) \\ &= \lim_{N\to\infty} \sum_{n=1}^{N} \mu(E_n) \\ &= \sum_{n=1}^{\infty} \mu(E_n).\end{aligned}$$

Thus μ is a Borel measure on $[0,1]$. It is finite since $\mu([0,1]) = \widetilde{\ell}(1_{[0,1]}) < \infty$. Let $f \in C[0,1]$. Choose an increasing sequence (φ_n) of simple functions which converges pointwise to f. Then by MCT we have

$$\int_0^1 \varphi_n \mathrm{d}\mu \to \int_0^1 f \mathrm{d}\mu.$$

If $\varphi = \sum_{k=1}^n a_k 1_{A_k}$, then

$$\begin{aligned}\int_0^1 \varphi \mathrm{d}\mu &= \sum_{k=1}^n a_k \mu(A_k) \\ &= \sum_{k=1}^n a_k \widetilde{\ell}(1_{A_k}) \\ &= \widetilde{\ell}\left(\sum_{k=1}^n a_k 1_{A_k}\right) \\ &= \widetilde{\ell}(\varphi).\end{aligned}$$

So $\widetilde{\ell}(\varphi_n) \to \widetilde{\ell}(f) = \ell(f)$. We have

$$\int_0^1 \varphi_n \mathrm{d}\mu \to \ell(f)$$

Thus $\widetilde{\ell}(f) = \int f \mathrm{d}\mu$ for any f continuous. □

Another formulation of the Riesz Representation Theorem is given by:

Theorem 43.5. *(Riesz Representation Theorem) For any bounded (with respect to the supremum norm) linear functional $\ell \colon C[0,1] \to \mathbb{R}$ such that $\ell(f) \geq 0$ for all $f \geq 0$, there exists a unique finite (signed) measure μ on $[0,1]$ such that*

$$\ell(f) = \int_0^1 f \mathrm{d}\mu.$$

And a more general version of the Riesz Representation Theorem is given by:

Theorem 43.6. *(Kakutani general version of the Riesz Representation Theorem) Let X be a compact Hausdorff topological space and let $C(X)$ be the Banach space of all continuous functions $f \colon X \to \mathbb{R}$ equipped with the supremum norm:*

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

For any bounded linear functional $\ell \colon C(X) \to \mathbb{R}$ there exists a unique Borel regular measure μ on X such that

$$\ell(f) = \int_X f \mathrm{d}\mu.$$

Let $f \in C[0,1]$. Then f is uniformly continuous. For each $n \in \mathbb{N}$ define a partition

$$0 < x_0^{(n)} < x_1^{(n)} < \cdots < x_n^{(n)} = 1$$

of $[0,1]$ such that none of these points are discontinuities of f and such that

$$|x_{i+1}^{(n)} - x_i^{(n)}| < \frac{2}{n}$$

for all $i = 0, 1, \dots, n$. Now define $\varphi_n \colon [0,1] \to \mathbb{R}$ by

$$\varphi_n(x) = \sum_{i=0}^{n-1} f(x_i^{(n)}) 1_{(x_i^{(n)}, x_{i+1}^{(n)}]}$$

for all $x \in [0,1]$. Since f is uniformly continuous, we see that (φ_n) converges uniformly to f. Therefore $\widetilde{\ell}(\varphi_n) \to \widetilde{\ell}(f)$ and $\int_0^1 \varphi_n \mathrm{d}\mu \to \int_0^1 f \mathrm{d}\mu$. So it suffices to show

$$\int_0^1 \varphi_n \mathrm{d}\mu = \widetilde{\ell}(\varphi_n).$$

Thus

$$
\begin{aligned}
\widetilde{\ell}(\varphi_n) &= \widetilde{\ell}(\sum_{i=0}^{n-1} f(x_i^{(n)}) 1_{(x_i^{(n)}, x_{i+1}^{(n)}]}) \\
&= \sum_{i=0}^{n-1} f(x_i^{(n)}) \widetilde{\ell}(1_{(x_i^{(n)}, x_{i+1}^{(n)}]}) \\
&= \int_0^1 \varphi_n \mathrm{d}\mu
\end{aligned}
$$

for all $n \in \mathbb{N}$.

Theorem 43.7. *(Hausdorff) A sequence (a_n) is a moment sequence of some finite Borel measure μ on $[0,1]$, that is,*

$$
a_n = \int_0^1 x^n \mathrm{d}\mu
$$

if and only if $(-1)^k(\Delta^k a)_n \geq 0$ for all $k, n \geq 0$ where $(\Delta a)_n = a_{n+1} - a_n$.

We have

$$
\begin{aligned}
\Delta^2 a &= \Delta(\Delta a) \\
&= (a_{n+2} - 2a_{n+1} + a_n)_n
\end{aligned}
$$

More generally

$$
\Delta^k a = (\sum_{i=n}^{n+k} (-1)^i \binom{n}{i} a_{n+i}).
$$

Sequences satisfying this condition

$$
((-1)^k \Delta^k a)_n \geq 0
$$

are called monotone sequences. Observe that

$$
(-1)^k(\Delta^k a)_n = \int_0^1 x^n (1-x)^k \mathrm{d}\mu \geq 0.
$$

43.5 Hahn-Banach Theorem

Definition 43.4. Let V be an $\mathbb{R}$-vector space. A **partial-seminorm** is a function $p\colon V \to \mathbb{R}$ which satisfies

1. (nonnegativity) $p \geq 0$, that is, $p(x) \geq 0$ for all $x \in V$.
2. (nonnegative homogeneity) $p(\lambda x) = \lambda p(x)$ for all $\lambda \geq 0$ and $x \in V$.
3. (subadditivity) $p(x+y) \leq p(x) + p(y)$ for all $x, y \in V$.

Remark 59. The terminology "partial-seminorm" is made up by me. Recall that a **seminorm** is a function $p\colon V \to \mathbb{R}$ which satisfies

1. (absolute homogeneity) $p(\lambda x) = |\lambda| p(x)$ for all $\lambda \in \mathbb{R}$ and $x \in V$.
2. (subadditivity) $p(x+y) \leq p(x) + p(y)$ for all $x, y \in V$.

It is easy to check that a seminorm is necessarily nonnegative. Thus every seminorm is a partial-seminorm. On the other hand, there are partial-seminorms which are not seminorms. Indeed, consider the function $p\colon \mathbb{R} \to \mathbb{R}$ defined by

$$
p(x) = \begin{cases} x & \text{if } x \geq 0 \\ -x/2 & \text{if } x < 0 \end{cases}
$$

for all $x \in \mathbb{R}$. It is easy to check that p is a partial-seminorm which is not a seminorm.

Theorem 43.8. *Let V be an $\mathbb{R}$-vector space equipped with a partial-seminorm $p\colon V \to \mathbb{R}$ and let U be a subspace of V. Then every linear functional $\varphi\colon U \to \mathbb{R}$ such that $|\varphi| \leq p|_U$ can be extended to a linear functional $\widetilde{\varphi}\colon V \to \mathbb{R}$ such that $\widetilde{\varphi}|_U = \varphi$ and $|\widetilde{\varphi}| \leq p$.*

Remark 60. Note that by $|\varphi| \leq p|_U$, we mean $|\varphi(u)| \leq p(u)$ for all $u \in U$.

Proof. Let $\varphi \colon U \to \mathbb{R}$ be a linear functional such that $|\varphi| \leq p|_U$. We will construct an extension of φ using Marcel Riesz's Extension Theorem. Let

$$P = \{(\lambda, v) \in \mathbb{R} \times V \mid p(v) \leq \lambda\}.$$

Then observe that P is a convex cone contained in the space $\mathbb{R} \times V$. Indeed, if $\alpha > 0$ and $(\lambda, v) \in P$, then $(\alpha\lambda, \alpha v) \in P$ since

$$\begin{aligned} p(\alpha v) &= \alpha p(v) \\ &\leq \alpha\lambda \end{aligned}$$

Also if $(\lambda_1, v_1), (\lambda_2, v_2) \in P$, then $(\lambda_1 + \lambda_2, v_1 + v_2) \in P$ since

$$\begin{aligned} p(v_1 + v_2) &\leq p(v_1) + p(v_2) \\ &= \lambda_1 + \lambda_2. \end{aligned}$$

Furthermore, we have $\mathbb{R} \times V = (\mathbb{R} \times U) + P$, since if $(\lambda, v) \in \mathbb{R} \times V$, then

$$(\lambda, v) = (\lambda - p(v), 0) + (p(v), v)$$

with $(\lambda - p(v), 0) \in \mathbb{R} \times U$ and $(p(v), v) \in P$. Finally define $\psi \colon \mathbb{R} \times U \to \mathbb{R}$ by

$$\psi(\lambda, u) = \lambda - \varphi(u)$$

for all $(\lambda, u) \in \mathbb{R} \times U$. Observe that $\psi|_{(\mathbb{R} \times U) \cap P} \geq 0$. Indeed, if $(\lambda, v) \in (\mathbb{R} \times U) \cap P$, then

$$\begin{aligned} \psi(\lambda, v) &= \lambda - \varphi(v) \\ &\geq \lambda - p(v) \\ &\geq 0 \end{aligned}$$

Thus we have all of the ingredients to apply the Marcel Riesz Extension Theorem: choose $\widetilde{\psi} \colon \mathbb{R} \times V \to \mathbb{R}$ such that $\widetilde{\psi}|_{\mathbb{R} \times U} = \psi$ and $\widetilde{\psi}|_P \geq 0$. Define $\widetilde{\varphi} \colon V \to \mathbb{R}$ by

$$\widetilde{\varphi}(v) = -\widetilde{\psi}(0, v)$$

for all $v \in V$. Note that if $u \in U$, then

$$\begin{aligned} \widetilde{\varphi}(u) &= -\widetilde{\psi}(0, u) \\ &= -\psi(0, u) \\ &= \varphi(u). \end{aligned}$$

Thus $\widetilde{\varphi}|_U = \varphi$. We claim $|\widetilde{\varphi}| \leq p$. To see this, assume for a contradiction that $v_0 \in V$ such that

$$\widetilde{\varphi}(v_0) > p(v_0).$$

Then using that $(p(x_0), x_0) \in P$, we have

$$\begin{aligned} 0 &\leq \widetilde{\psi}(p(x_0), x_0) \\ &= \widetilde{\psi}(0, x_0) + \widetilde{\psi}(p(x_0), 0) \\ &= -\widetilde{\varphi}(x_0) + \psi(p(x_0), 0) \\ &= -\widetilde{\varphi}(x_0) + p(x_0) \\ &< -p(x_0) + p(x_0) \\ &= 0, \end{aligned}$$

which is a contradiction. This establishes our claim and we are done. $\square$

In the setting of normed linear spaces, the Hahn-Banach Theorem says that any linear functional ℓ defined on a subspace $\mathcal{Y} \subseteq \mathcal{X}$ which is bounded on $\mathcal{Y}$ can be extended to a bounded linear functional $\widetilde{\ell}$ on $\mathcal{X}$ such that $\widetilde{\ell}|_{\mathcal{Y}} = \ell$ and $\|\widetilde{\ell}\|_{\mathcal{X}} = \|\ell\|_{\mathcal{Y}}$. This is an immediate consequence of our more general version that we have just proved.

Proposition 43.5. *Let $\mathcal{X}$ be a normed linear space and let x_0 be a nonzero vector in $\mathcal{X}$. Then there exists a bounded linear functional $\ell\colon \mathcal{X} \to \mathbb{R}$ with $\|\ell\| = 1$ such that $\ell(x_0) = \|x_0\|$.*

So if you have two points $a \neq b$ in $\mathcal{X}$, then there exists a bounded linear functional $\ell \in \mathcal{X}^*$ such that $\ell(a) \neq \ell(b)$.

Theorem 43.9. *Let $\mathcal{X}$ be a reflexive Banach space and let $\mathcal{Y}$ be a closed subspace of $\mathcal{X}$. Then for every $x \in \mathcal{X}$ there exists $y_0 \in \mathcal{Y}$ such that $\mathrm{d}(x, \mathcal{Y}) = \|x - y_0\|$.*

Remark 61. We can replace $\mathcal{Y}$ with a convex set.

Proof. Define a function $\varphi\colon \mathcal{Y} \to \mathbb{R}$ by

$$\varphi(y) = \|y - x\|$$

for all $y \in \mathcal{Y}$. □

44 Geometric Form of the Hahn-Banach Theorem

44.1 Gauge Functional

Definition 44.1. Let V be an $\mathbb{R}$-vector space and let S be a subset of V. A point $x \in S$ is said to be an **internal point** of S if for any $y \in V$, there exists $\varepsilon_{x,y} > 0$ such that $|t| < \varepsilon_{x,y}$ implies $x + ty \in S$. The set of all points internal points of S is called the **core** of S and is denoted by $\operatorname{core} S$.

Remark 62. Let us make several remarks about this definition.

1. We write $\varepsilon_{x,y}$ to emphasize that $\varepsilon_{x,y}$ *depends* on x and y. Usually we will just write ε instead of $\varepsilon_{x,y}$.

2. Note that if $0 \in \operatorname{core} S$, then $0 \in S$. Indeed, assuming $0 \in \operatorname{core} S$, then there exists $\varepsilon_{0,0} > 0$ such that $|t| < \varepsilon_{0,0}$ implies $0 = 0 + t \cdot 0 \in S$. The converse of course isn't true (take $S = \{0\}$).

3. Suppose V is equipped with a metric. Recall that a point $x \in S$ is said to be an **interior point** of S if there exists an $\varepsilon > 0$ such that $\mathrm{B}_\varepsilon(x) \subseteq S$. The set of all interior points of S is denoted $\operatorname{int} S$. It is easy to see that every interior point of S is an internal point of S. Thus $\operatorname{int} S \subseteq \operatorname{core} S$. If S happens to be open, then $S = \operatorname{int} S \subseteq \operatorname{core} S = S$ which forces $\operatorname{int} S = \operatorname{core} S$.

Definition 44.2. Let V be an $\mathbb{R}$-vector space and let $C \subseteq V$ be a convex set with 0 as an internal point. We define the **gauge functional** of C to be the function $\mathrm{p}_C\colon V \to \mathbb{R}$ given by

$$\mathrm{p}_C(x) = \inf\{\alpha > 0 \mid (1/\alpha)x \in C\}$$

for all $x \in V$.

Note that 0 be an internal point of C guarantees that $\mathrm{p}_C(x) < \infty$. Indeed, since 0 is an internal point, there exists an $\varepsilon > 0$ such that $tx \in C$ for all $|t| < \varepsilon$. In particular, if $\alpha > 1/\varepsilon$, then $1/\alpha < \varepsilon$, and hence $(1/\alpha)x \in C$. Thus we see that $\mathrm{p}_C(x) \leq 1/\varepsilon$. Thus having 0 be an internal point of C guarantees that $\mathrm{p}_C(x) < \infty$.

Example 44.1. Let $\mathcal{X}$ be a normed linear space. Then $\mathrm{p}_{\mathrm{B}_1[0]}(x) = \|x\|$ for all $x \in \mathcal{X}$.

44.1.1 Gauge Functional is a Partial-Seminorm

Proposition 44.1. *Let V be an $\mathbb{R}$-vector space and let $C \subseteq V$ be a convex set with 0 as an internal point. Then the gauge functional p_C is a partial-seminorm.*

Proof. We first show p_C is subadditive. Let $\varepsilon > 0$ and let $x, y \in V$. Set $a = \mathrm{p}_C(x) + \varepsilon/2$ and set $b = \mathrm{p}_C(y) + \varepsilon/2$. Then $a, b > 0$ and $(1/a)x, (1/b)x \in C$. Since C is convex, we see that

$$\frac{1}{a+b}(x+y) = \frac{a}{a+b}\left(\frac{1}{a}x\right) + \frac{b}{a+b}\left(\frac{1}{b}y\right) \in C.$$

It follows that

$$\begin{aligned} \mathrm{p}_C(x) + \mathrm{p}_C(y) + \varepsilon &= a + b \\ &\geq \mathrm{p}_C(x+y). \end{aligned}$$

Taking $\varepsilon \to 0$ shows that p_C is subadditive.

Next we show that p_C satisfies nonnegative homogeneity. Let $\lambda \geq 0$ and let $x \in V$. First note that if $\lambda = 0$, then since

$$\mathrm{p}_C(0) = \inf\{\alpha > 0 \mid (1/\alpha) \cdot 0 \in C\} = 0,$$

we have $0 = 0 \cdot \mathrm{p}_C(x) = \mathrm{p}_C(0 \cdot x)$. Thus we may assume $\lambda > 0$. Then

$$\begin{aligned} \mathrm{p}_C(\lambda x) &= \inf\{\alpha > 0 \mid (1/\alpha)\lambda x \in C\} \\ &= \lambda \inf\{\alpha > 0 \mid (1/\alpha) x \in C\} \\ &= \lambda \mathrm{p}_C(x). \end{aligned}$$

Finally note that p_C is nonnegative by definition. Thus p_C is a partial-seminorm. □

44.1.2 Properties of Gauge Functional

Proposition 44.2. *Let V be an $\mathbb{R}$-vector space and let $C \subseteq V$ be a convex set with 0 as an internal point. We have*

1. $C \subseteq \{\mathrm{p}_C \leq 1\}$.
2. $\operatorname{core} C = \{\mathrm{p}_C < 1\}$.

Proof. 1. Let $x \in C$. Then $(1/1)x \in C$ and hence $\mathrm{p}_C(x) \leq 1$.

2. Let $x \in \operatorname{core} C$. Then there exists $\varepsilon > 0$ such that $x + \varepsilon x \in C$. So

$$\begin{aligned} x + \varepsilon x &= (1+\varepsilon)x \\ &= \frac{1}{1/(1+\varepsilon)} x \end{aligned}$$

shows $\mathrm{p}_C(x) \leq 1/(1+\varepsilon) < 1$. Conversely, let $x \in V$ such that $\mathrm{p}_C(x) < 1$. Then there exists $0 < \alpha < 1$ such that $(1/\alpha)x \in C$. Now let $y \in V$. Since $0 \in \operatorname{core}(C)$, there exists $\varepsilon > 0$ such that $|t| < \varepsilon$ implies $ty \in C$. Then $|t| < \varepsilon$ implies

$$x + (1-\alpha)ty = \alpha(1/\alpha)x + (1-\alpha)ty \in C$$

since C is convex. In particular, setting $\delta = (1-\alpha)\varepsilon$, we see that $|t| < \delta$ implies $x + ty \in C$. □

44.1.3 Gauge Functional Induced from Partial-Seminorm

Recall from Proposition (44.1) that is C is a convex subset of a real vector space V such that $0 \in \operatorname{core} C$, then the gauge functional $\mathrm{p}_C \colon V \to \mathbb{R}$ is a partial-seminorm. We will now show a converse to this.

Proposition 44.3. *Let V be an $\mathbb{R}$-vector space, let $p \colon V \to \mathbb{R}$ be a partial-seminorm, and set $C = \{p \leq 1\}$. Then C is a convex set, and moreover, we have $\mathrm{p}_C = p$.*

Proof. Let $x, y \in C$ and $\alpha \in [0,1]$. Then

$$\begin{aligned} p((1-\alpha)x + \alpha y) &\leq p((1-\alpha)x) + p(\alpha y) \\ &= (1-\alpha)p(x) + \alpha p(y) \\ &\leq (1-\alpha) + \alpha \\ &= 1 \end{aligned}$$

implies $(1-\alpha)x + \alpha y \in C$. Thus C is a convex set.

Now assume there exists $x_0 \in V$ such that $\mathrm{p}_C(x_0) < p(x_0)$. Then there exists $\alpha \in \mathbb{R}$ such that

$$\mathrm{p}_C(x_0) \le \alpha < p(x_0)$$

and such that $(1/\alpha)x_0 \in C$. Then $p((1/\alpha)x_0) \le 1$ which is equivalent to $(1/\alpha)p(x_0) \le 1$ which implies $p(x_0) \le \alpha$. This is a contradiction. So $\mathrm{p}_C(x) \ge p(x)$ for all $x \in V$. Now assume there exists $x_0 \in V$ such that $p(x_0) < \mathrm{p}_C(x_0)$. Then there exists $\alpha \in \mathbb{R}$ such that

$$p(x_0) \le \alpha < \mathrm{p}_C(x_0).$$

Then $(1/\alpha)p(x_0) \le 1$. In other words, $p((1/\alpha)x_0) \le 1$ which is equivalent to $(1/\alpha)x_0 \in C$. This contradicts the fact that $\mathrm{p}_C(x_0)$ is the infimum of all such $\alpha > 0$. Therefore $p(x) \ge \mathrm{p}_C(x)$ for all $x \in V$. It follows that $p = \mathrm{p}_C$. □

Theorem 44.1. *Let V be an $\mathbb{R}$-vector space and let C be a nonempty convex subset of V such that $C = \operatorname{core} C$. Then for any $y \notin C$, there exists a hyperplane $\{\ell = \alpha\}$ where $\ell \colon V \to \mathbb{R}$ is some linear functional and $\alpha \in \mathbb{R}$ such that $y \in \{\ell = \alpha\}$ and $C \subseteq \{\ell < \alpha\}$.*

Proof. By translating if necessary, we may assume that $0 \in \operatorname{int} C$. This means it is possible to define the gauge potential p_C of C. Define $\ell \colon \mathbb{R}y \to \mathbb{R}$ by $\ell(ay) = a$ for all $ay \in \mathbb{R}y$. Notice if $a < 0$, then

$$\begin{aligned} \ell(ay) &= a \\ &< 0 \\ &\le \mathrm{p}_C(ay), \end{aligned}$$

and if $a > 0$, then

$$\begin{aligned} \ell(ay) &= a \\ &\le a\mathrm{p}_C(y) \\ &= \mathrm{p}_C(ay), \end{aligned}$$

where we used the fact that $\mathrm{p}_C(y) \ge 1$ since $y \notin \operatorname{core} C = C$. So we see that $\ell \le \mathrm{p}_C|_{\mathbb{R}y}$. Therefore by the Hahn-Banach Theorem, we can extend ℓ to $\widetilde{\ell} \colon V \to \mathbb{R}$ such that $\widetilde{\ell}|_{\mathbb{R}y} = \ell$ and $\widetilde{\ell} \le \mathrm{p}_C$. In particular, if $x \in C$, then

$$\widetilde{\ell}(x) \le \mathrm{p}_C(x) < 1.$$

Thus $C \subseteq \{\widetilde{\ell} < \alpha\}$ where $\alpha = 1$. Also clearly $\widetilde{\ell}(y) = 1$, and so we are done. □

44.1.4 First Geometric Form of Hahn-Banach

Theorem 44.2. *(first geometric form of Hahn-Banach) Let V be an $\mathbb{R}$-vector space and let $A, B \subseteq V$ be nonempty convex sets such that $A \cap B = \emptyset$. Suppose A satisfies $A = \operatorname{core} A$. Then there exists a hyperplane that separates A and B. More precisely, there exists a linear functional $\ell \colon V \to \mathbb{R}$ and $\alpha \in \mathbb{R}$ such that $A \subseteq \{\ell \le \alpha\}$ and $B \subseteq \{\ell \ge \alpha\}$.*

Proof. Set $C = A - B = \{a - b \mid a \in A, b \in B\}$. Then C is a nonempty convex set. Furthermore we have $\operatorname{int} C = C$. Indeed, let $a - b \in C$ and let $y \in V$. Choose $\varepsilon > 0$ such that $|t| < \varepsilon$ implies $a + ty \in A$. Then $|t| < \varepsilon$ implies $a - b + ty = (a + ty) - b \in C$. Finally note that $0 \notin C$ since A and B are disjoint from one another. By the previous result, there exists a linear functional $\ell \colon V \to \mathbb{R}$ and an $\beta \in \mathbb{R}$ such that $0 \in \{\ell = \beta\}$ and $C \subseteq \{\ell < \beta\}$. Note that since $\ell(0) = \beta$, we must necessarily have $\beta = 0$.

Now let $a \in A$ and $b \in B$. Since $a - b \in C$, we have $0 > \ell(a - b) = \ell(a) - \ell(b)$, that is, $\ell(a) < \ell(b)$. Therefore

$$\sup\{\ell(a) \mid a \in A\} \le \inf\{\ell(b) \mid b \in B\}.$$

So choose α between $\sup\{\ell(a) \mid a \in A\}$ and $\inf\{\ell(b) \mid b \in B\}$. Then $A \subseteq \{\ell \le \alpha\}$ and $B \subseteq \{\ell \ge \alpha\}$. □

44.1.5 Second Geometric Form of Hahn-Banach

Lemma 44.3. *Let $\mathcal{X}$ be a normed linear space, let A be a closed subset of $\mathcal{X}$, and let B be a compact subset of $\mathcal{X}$. Then $A+B$ is closed.*

Proof. Let $x \in \overline{A+B}$ and choose a sequence $(a_n + b_n)$ in $A+B$ such that $a_n + b_n \to x$. Since B is compact, there exist a convergent subsequence of (b_n), say $(b_{\pi(n)})$. In fact, by relabeling indices if necessary, we may assume that (b_n) is convergent, say $b_n \to b$ where $b \in B$. Now since $a_n + b_n \to x$ and $b_n \to b$, it follows easily that $a_n \to x - b$. Since A is closed, we must have $x - b \in A$. Thus $x = (x-b) + b$ shows $x \in A+B$, which implies $A + B = \overline{A+B}$, hence $A+B$ is closed. □

Theorem 44.4. *(second geometric form of Hahn-Banach) Let $\mathcal{X}$ be a normed linear space and let $A, B \subseteq \mathcal{X}$ be two nonempty convex sets such that $A \cap B = \emptyset$. Suppose A is closed and B is compact. Then there exists a closed hyperplane that strictly separates A and B. More precisely, there exists a bounded linear functional $\ell\colon V \to \mathbb{R}$ and $\alpha \in \mathbb{R}$ such that $A \subseteq \{\ell < \alpha\}$ and $B \subseteq \{\ell > \alpha\}$.*

Proof. Set $C = A - B = \{a - b \mid a \in A, b \in B\}$. Then C is a nonempty convex set. Furthermore, C is closed by Lemma (44.3) since $-B$ is compact and $A - B = A + (-B)$. Also $0 \notin C$ since A and B are disjoint from one another. Thus C^c is open and contains 0, which means there exists $r > 0$ such that $\mathrm{B}_r(0) \subseteq C^c$. In other words, $\mathrm{B}_r(0) \cap C = \emptyset$. By the previous first geometric form of Hahn-Banach, we can separate $\mathrm{B}_r(0)$ and C by a hyperplane, say $\{\ell = \alpha\}$. Then $\ell(a-b) \leq \ell(rx)$ for all $a \in A$, $b \in B$ and $x \in \mathrm{B}_1(0)$. It can be shown that $\ell\colon \mathcal{X} \to \mathbb{R}$ is bounded. Therefore

$$\ell(a-b) \leq \inf\{\ell(rx) \mid x \in \mathrm{B}_1(0)\} = -r\|\ell\|.$$

Now take $\varepsilon = (1/2)r\|\ell\| > 0$. Then

$$\ell(a) + \varepsilon \leq \ell(b) - \varepsilon$$

for all $a \in A$ and $b \in B$. This implies

$$\sup\{\ell(a) \mid a \in A\} < \inf\{\ell(b) \mid b \in B\}.$$

So choose α strictly between $\sup\{\ell(a) \mid a \in A\}$ and $\inf\{\ell(b) \mid b \in B\}$. Then $A \subseteq \{\ell < \alpha\}$ and $B \subseteq \{\ell > \alpha\}$. □

44.2 Lower Semicontinuity

Definition 44.3. Let $\mathcal{X}$ be a normed linear space. A function $\varphi\colon \mathcal{X} \to (-\infty, \infty]$ is said to be **lower semicontinuous** if for every $c \in \mathbb{R}$ the set $\{\varphi \leq c\}$ is closed.

Here are some basic facts:

1. φ is lower semicontinuous if and only if $\{(x,\lambda) \mid \varphi(x) \leq \lambda\}$ is a closed set in $\mathcal{X} \times \mathbb{R}$ for every $\lambda \in \mathbb{R}$.
2. φ_1 and φ_2 are lower semicontinuous implies $\varphi_1 + \varphi_2$ is lower semicontinuous.
3. $\{\varphi_i\}_{i \in I}$ is a collection of lower semicontinuous functions, then $\sup_{i \in I} \varphi_i$ is also lower semicontinuous.
4. if $K \subseteq \mathcal{X}$ is compact, then $\inf_{x \in \mathcal{K}} \varphi(x)$ is acheived.

44.3 Convexity

Definition 44.4. Let $\mathcal{X}$ be a normed linear space. A function $\varphi\colon \mathcal{X} \to (-\infty, \infty]$ is said to be **convex** if

$$\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y)$$

for all $x, y \in \mathcal{X}$ and $t \in [0,1]$.

Here are some basic facts:

1. φ is convex if and only if $\mathrm{epi}(\varphi) = \{(x,\lambda) \mid \varphi(x) \leq \lambda\}$ is a convex set in $\mathcal{X} \times \mathbb{R}$.
2. If φ_1 and φ_2 are convex, then $\varphi_1 + \varphi_2$ is convex.
3. If $\{\varphi_i\}_{i \in I}$ are all convex, then $\sup_{i \in I} \varphi_i$ is convex.
4. If φ is convex, then $\{\varphi \leq c\}$ is a convex set for all $c \in \mathbb{R}$. The converse is not true in general.

We usually assume both convexity and lower semicontinuity in optimization problems.

44.3.1 Conjugate Function

Definition 44.5. Let $\mathcal{X}$ be a normed linear space and let $\varphi\colon \mathcal{X} \to (-\infty, \infty]$ be a function such that $\varphi \neq \infty$.

1. We define the **conjugate function** of φ to be the function $\varphi^*\colon \mathcal{X}^* \to (-\infty, \infty]$ defined by

$$\varphi^*(\ell) = \sup_{x \in \mathcal{X}} (\ell(x) - \varphi(x))$$

   for all $\ell \in \mathcal{X}^*$. The conjugate function φ^* is sometimes called a **Fenchel transform** of φ or a **Legendre transform** of φ.

2. We define the **double conjugate function** of φ to be the function $\varphi^{**}\colon \mathcal{X} \to (-\infty, \infty]$ defined by

$$\varphi^{**}(x) = \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \varphi^*(\ell))$$

   for all $x \in \mathcal{X}$.

Example 44.2. Suppose $\mathcal{X} = \mathbb{R}$ and $\varphi\colon \mathcal{X} \to (-\infty, \infty]$ is given by

$$\varphi(x) = \frac{1}{p}|x|^p$$

for all $x \in \mathcal{X}$ where $1 < p < \infty$. Recall from the Riesz representation theorem for Hilbert, each $\ell \in \mathcal{X}^*$ has the form $\ell = \ell_y$ for a unique $y \in \mathbb{R}$ where $\ell_y(x) = yx$ for all $x \in \mathcal{X}$. Using this fact, suppose $\ell = \ell_y$ is in $\mathcal{X}^*$. Then we have

$$\begin{aligned}
\varphi^*(y) &:= \varphi^*(\ell_y) \\
&= \sup_{x \in \mathbb{R}} (\ell_y(x) - \varphi(x)) \\
&= \sup_{x \in \mathbb{R}} \left(yx - \frac{1}{p}|x|^p \right) \\
&= \sup_{x \in \mathbb{R}} \left(|y||x| - \frac{1}{p}|x|^p \right) \\
&= \frac{1}{q}|y|^q + \frac{1}{p}|y^{p/q}|^p - \frac{1}{p}|y^{p/q}|^p \\
&= \frac{1}{q}|y|^q,
\end{aligned}$$

where $1 < q < \infty$ such that $1/p + 1/q = 1$. Here, we used Young's inequality, which says

$$\frac{a^p}{p} + \frac{b^q}{q} \geq ab$$

for all $a, b \geq 0$, with equality acheived if and only if $a^p = b^q$.

The example above suggests that we have the following generalization of Young's inequality:

$$\varphi^*(\ell) + \varphi(x) \geq \ell(x)$$

for all $\ell \in \mathcal{X}^*$ and $x \in \mathcal{X}$. Indeed, this is a simple consequence of the definition of φ^*: for all $\ell \in \mathcal{X}^*$, we have

$$\begin{aligned}
\varphi^*(\ell) &= \sup_{x \in \mathcal{X}} (\ell(x) - \varphi(x)) \\
&\geq \varphi(x) - \ell(x)
\end{aligned}$$

for all $x \in \mathcal{X}$.

44.3.2 Fenchel-Moreau

Lemma 44.5. *Let $\mathcal{X}$ be a normed linear space and let $\varphi\colon \mathcal{X} \to (-\infty, \infty]$ be a lower semicontinuous convex function such that $\varphi \neq \infty$. Then $\varphi^* \neq \infty$.*

Proof. Choose $x_0 \in \mathcal{X}$ such that $\varphi(x_0) < \infty$ and choose $\lambda_0 \in \mathbb{R}$ such that $\lambda_0 < \varphi(x_0)$. Consider the normed linear space $\mathcal{X} \times \mathbb{R}$ and the subsets $A = \{(x, \lambda) \mid \varphi(x) \leq \lambda\}$ and $B = \{(x_0, \lambda_0)\}$. Then A is a nonempty closed convex set and B is a nonempty compact convex set. Furthermore A and B are disjoint from one another. Thus by the second geometric form of Hahn-Banach, there exists a bounded linear functional $\ell\colon \mathcal{X} \times \mathbb{R} \to \mathbb{R}$ and an $\alpha \in \mathbb{R}$ such that

$$A \subseteq \{\ell > \alpha\} \quad \text{and} \quad B \subseteq \{\ell < \alpha\}. \tag{113}$$

Define $\psi\colon \mathcal{X} \to \mathbb{R}$ by $\psi(x) = \ell(x, 0)$ for all $x \in \mathcal{X}$. Then ψ is a bounded linear functional because ℓ is a bounded linear functional and $\psi = \ell|_{\mathbb{R}\times\{0\}}$. Set $k = \ell(0, 1)$ and note that

$$\begin{aligned} \ell(x, \lambda) &= \ell(x, 0) + \ell(0, \lambda) \\ &= \psi(x) + \lambda k \end{aligned}$$

for all $(x, \lambda) \in \mathcal{X} \times \mathbb{R}$.

Now by (113), we have

$$\begin{cases} \psi(x) + \lambda k > \alpha & \text{if } (x, \lambda) \in A \\ \psi(x_0) + \lambda_0 k < \alpha \end{cases}$$

In particular, since $(x_0, \varphi(x_0)) \in A$, we have

$$\begin{aligned} 0 &< \psi(x_0) + \varphi(x_0)k - \alpha \\ &< \psi(x_0) + \varphi(x_0)k - \psi(x_0) - \lambda_0 k \\ &= \varphi(x_0)k - \lambda_0 k \\ &= (\varphi(x_0) - \lambda_0)k. \end{aligned}$$

Thus $k > 0$ since $\varphi(x_0) > \lambda_0$. Now using the fact that $(x, \varphi(x)) \in A$ for all $x \in \mathcal{X}$, we can divide $\psi(x) + \lambda k > \alpha$ by $-1/k$ to obtain

$$-\frac{1}{k}\psi(x) - \varphi(x) < -\frac{\alpha}{k}.$$

In particular, we see that

$$\begin{aligned} \varphi^*(-\psi/k) &= \sup_{x \in \mathcal{X}} (-\psi(x)/k - \varphi(x)) \\ &\leq -\frac{\alpha}{k} \\ &< \infty. \end{aligned}$$

So $\varphi^* \neq \infty$. □

Theorem 44.6. *(Fenchel-Moreau) If $\varphi\colon X \to (-\infty, \infty]$ is lower semicontinuous, convex, and $\varphi \neq \infty$, then $\varphi^{**} = \varphi$.*

Proof. Note that for every $\ell \in \mathcal{X}^*$ and $x \in \mathcal{X}$, we have

$$\begin{aligned} \ell(x) - \varphi^*(\ell) &= \ell(x) - \sup_{y \in \mathcal{X}} (\ell(y) - \varphi(y)) \\ &\leq \ell(x) - (\ell(x) - \varphi(x)) \\ &= \varphi(x). \end{aligned}$$

Therefore

$$\begin{aligned} \varphi^{**}(x) &= \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \varphi^*(\ell)) \\ &\leq \varphi(x). \end{aligned}$$

It remains to show $\varphi^{**}(x) \geq \varphi(x)$.

Step 1: Suppose $\varphi \geq 0$ and assume for a contradiction that $\varphi^{**}(x_0) < \varphi(x_0)$. We apply the second geometric form of Hahn-Banach again in the space $\mathcal{X} \times \mathbb{R}$ with sets $A = \{(x, \lambda) \mid \varphi(x) \leq \lambda\}$ and $B = \{(x_0, \varphi^{**}(x_0))\}$. By the same argument as in the proof of Lemma (44.5), there exists a bounded linear functional $\ell \in \mathcal{X}^*$, an $\alpha \in \mathbb{R}$, and a $k \in \mathbb{R}$ such that

$$\ell(x) + \lambda k > \alpha \tag{114}$$

for all $(x, \lambda) \in A$ and such that

$$\ell(x_0) + k\varphi^{**}(x_0) < \alpha \tag{115}$$

Note that we could have $\varphi(x_0) = \infty$, so we can't plug in $(x_0, \varphi(x_0))$ into (114) to conclude that $k > 0$ as in the proof of Lemma (44.5). However we can still show that $k \geq 0$. Indeed, assume for a contradiction that $k < 0$. Choose $y_0 \in \mathcal{X}$ such that $\varphi(y_0) < \infty$. Since $(y_0, \varphi(y_0)) \in A$, we have

$$\ell(y_0) + k\lambda \geq \ell(y_0) + k\varphi(y_0) > \alpha$$

for all $\lambda \geq \varphi(y_0)$. In particular, taking $\lambda \to \infty$ gives us $-\infty \geq \alpha$, which is a contradiction. So we must have $k \geq 0$. In order to proceed with the proof, we need to make k a little bigger, so choose $\varepsilon > 0$ so that $k + \varepsilon > 0$. Then just as in the proof of Lemma (44.5), we have

$$\varphi^*\left(-\frac{1}{k+\varepsilon}\ell\right) = \sup_{x \in \mathcal{X}}\left(-\frac{1}{k+\varepsilon}\ell(x) - \varphi(x)\right) \leq -\frac{\alpha}{k+\varepsilon}$$

and hence

$$\begin{aligned}
\ell(x_0) + (k+\varepsilon)\varphi^{**}(x_0) &= \ell(x_0) + (k+\varepsilon)\sup_{\ell \in \mathcal{X}^*}(\ell(x_0) - \varphi^*(\ell)) \\
&\geq \ell(x_0) + (k+\varepsilon)\left(-\frac{1}{k+\varepsilon}\ell(x_0) - \varphi^*\left(-\frac{1}{k+\varepsilon}\ell\right)\right) \\
&\geq \ell(x_0) + (k+\varepsilon)\left(-\frac{1}{k+\varepsilon}\ell(x_0) + \frac{\alpha}{k+\varepsilon}\right) \\
&= \ell(x_0) - \ell(x_0) + \alpha \\
&= \alpha.
\end{aligned}$$

By taking $\varepsilon \to 0$, we obtain

$$\ell(x_0) + k\varphi^{**}(x_0) \geq \alpha,$$

which contradicts (115). This contradiction proves that $\varphi^{**} \geq \varphi$, and hence $\varphi^{**} = \varphi$.

Step 2: Now consider the general case where we may not have $\varphi \geq 0$. Choose $\ell_0 \in \mathcal{X}^*$ such that $\varphi^*(\ell_0) < \infty$ (such ℓ_0 exists by Lemma (44.5)). Define $\varphi_1 \colon \mathcal{X} \to (-\infty, \infty]$ by

$$\varphi_1(x) = \varphi(x) - \ell_0(x) + \varphi^*(\ell_0).$$

Then φ_1 is convex, lower semicontinuous, and $\varphi_1 \not\equiv \infty$. In addition, we have $\varphi_1 \geq 0$. So by step 1, we obtain $\varphi_1^{**} = \varphi_1$. Now observe that

$$\begin{aligned}
\varphi_1^*(\ell) &= \sup_{x \in \mathcal{X}}(\ell(x) - \varphi_1(x)) \\
&= \sup_{x \in \mathcal{X}}(\ell(x) - \varphi(x) + \ell_0(x) - \varphi^*(\ell_0)) \\
&= \sup_{x \in \mathcal{X}}((\ell + \ell_0)(x) - \varphi(x)) - \varphi^*(\ell_0) \\
&= \varphi^*(\ell + \ell_0) - \varphi^*(\ell_0).
\end{aligned}$$

Therefore

$$\begin{aligned}
\varphi_1^{**}(x) &= \sup_{\ell \in \mathcal{X}^*}(\ell(x) - \varphi_1^*(\ell)) \\
&= \sup_{\ell \in \mathcal{X}^*}(\ell(x) - \varphi^*(\ell + \ell_0) + \varphi^*(\ell_0)) \\
&= \sup_{\ell + \ell_0 \in \mathcal{X}^*}((\ell + \ell_0)(x) - \varphi^*(\ell + \ell_0) - \ell_0(x) + \varphi^*(\ell_0)) \\
&= \varphi^{**}(x) - \ell_0(x) + \varphi^*(\ell_0).
\end{aligned}$$

So

$$\begin{aligned}\varphi^{**}(x) - \ell_0(x) + \varphi^*(\ell_0) &= \varphi_1^{**}(x) \\ &= \varphi_1(x) \\ &= \varphi(x) - \ell_0(x) + \varphi^*(\ell_0).\end{aligned}$$

Hence $\varphi^{**} = \varphi$.

□

44.3.3 Example

Example 44.3. Let $\mathcal{X}$ be a normed linear space and consider let $\varphi = \|\cdot\|$ be the norm function. Then φ is lower semicontinuous and convex. Let's compute the conjugate function

$$\begin{aligned}\varphi^*(\ell) &= \sup_{x\in\mathcal{X}}(\ell(x) - \|x\|) \\ &= \sup_{x\in\mathcal{X}} \|x\| \left(\ell\left(\frac{x}{\|x\|}\right) - 1\right).\end{aligned}$$

Now if $\|\ell\| > 1$, then there exists $x_0 \in \mathcal{X}$ such that $\|x_0\| = 1$ and $\ell(x_0) > 1$. Then for any $\lambda \in \mathbb{R}$, we have

$$\begin{aligned}\varphi^*(\ell) &= \sup_{x\in\mathcal{X}} \|x\| \left(\ell\left(\frac{x}{\|x\|}\right) - 1\right) \\ &\geq \|\lambda x_0\| \left(\ell\left(\frac{\lambda x_0}{\|\lambda x_0\|}\right) - 1\right) \\ &= |\lambda| \left(\ell(x_0) - 1\right),\end{aligned}$$

so by taking $\lambda \to \infty$, we see that $\varphi^*(\ell) = \infty$. On the other hand, if $\|\ell\| \leq 1$, then it is easy to check that $\varphi^*(\ell) = 0$. Thus

$$\varphi^*(\ell) = \begin{cases} 0 & \text{if } \|\ell\| \leq 1 \\ \infty & \text{if } \|\ell\| > 1 \end{cases}$$

For a set $E \subseteq \mathcal{X}$ nonempty we define

$$\mathrm{I}_E(x) = \begin{cases} 0 & \text{if } x \in E \\ \infty & \text{if } x \notin E \end{cases} = \log\left(\frac{1}{1_E(x)}\right)$$

So $\varphi^* = 1_{\mathrm{B}_1[0]}$ where

$$\mathrm{B}_1[0] = \{\ell \in \mathcal{X}^* \mid \|\ell\| \leq 1\}.$$

Now we have

$$\begin{aligned}\|x\| &= \varphi(x) \\ &= \varphi^{**}(x) \\ &= \sup_{\ell\in\mathcal{X}^*} (\ell(x) - \varphi^*(x)) \\ &= \sup_{\ell\in\mathcal{X}^*} (\ell(x) - \mathrm{I}_{\mathrm{B}_1[0]}(x)) \\ &= \sup_{\substack{\ell\in\mathcal{X}^* \\ \|\ell\|\leq 1}} \ell(x).\end{aligned}$$

This identity can be proved in a more elementary way by applying Hahn-Banach.

44.4 Support Functional

Definition 44.6. Let $\mathcal{X}$ be a normed linear space and let S be a subset of $\mathcal{X}$. We define $\mathrm{q}_S \colon \mathcal{X}^* \to (-\infty, \infty]$ by

$$\mathrm{q}_S(\ell) = \sup_{x\in S} \ell(x).$$

We call q_S the **support functional** of C.

44.4.1 Basic Properties of Support Functional

Proposition 44.4. *Let $\mathcal{X}$ be a normed linear space and let S be a subset of $\mathcal{X}$. Then*

1. q_S *is a partial-seminorm.*

2. $\mathrm{q}_S = \mathrm{q}_{\mathrm{conv}(S)} = \mathrm{q}_{\overline{\mathrm{conv}}(S)}$,

3. *Let S_1 and S_2 be subsets of $\mathcal{X}$. Then $\mathrm{q}_{S_1+S_2} = \mathrm{q}_{S_1} + \mathrm{q}_{S_2}$.*

4. *Let $\mathcal{K}$ be a closed subspace of $\mathcal{X}$. Then*
$$\mathrm{q}_{\mathcal{K}}(\ell) = \begin{cases} 0 & \text{if } \ell \in \mathcal{K}^{\perp} \\ \infty & \text{else} \end{cases}$$
where $\mathcal{K}^{\perp} = \{\ell \in \mathcal{X}^* \mid \ell|_{\mathcal{K}} = 0\}$.

Proof. 1. Clearly q_S is nonnegative since $\ell(0) = 0$ for all linear functionals $\ell \in \mathcal{X}^*$. Next, suppose $\lambda \geq 0$ and $\ell \in \mathcal{X}^*$. Then
$$\begin{aligned} \mathrm{q}_S(\lambda\ell) &= \sup_{x\in S} \ell(\lambda x) \\ &= \sup_{x\in S} \lambda\ell(x) \\ &= \lambda \sup_{x\in S} \ell(x) \\ &= \lambda \mathrm{q}_S(\ell). \end{aligned}$$
Similarly, suppose $\ell_1, \ell_2 \in \mathcal{X}^{\times}$. Then
$$\begin{aligned} \mathrm{q}_S(\ell_1+\ell_2) &= \sup_{x\in S}\{(\ell_1+\ell_2)(x)\} \\ &= \sup_{x\in S}\{\ell_1(x)+\ell_2(x)\} \\ &\leq \sup_{x\in S}\{\ell_1(x)\} + \sup_{x\in S}\{\ell_2(x)\} \\ &= \mathrm{q}_S(\ell_1) + \mathrm{q}_S(\ell_2). \end{aligned}$$
Thus q_S is a partial-seminorm.

2. Since $S \subseteq \mathrm{conv}(S) \subseteq \overline{\mathrm{conv}}(S)$, we clearly have $\mathrm{q}_S \leq \mathrm{q}_{\mathrm{conv}(S)} \leq \mathrm{q}_{\overline{\mathrm{conv}}(S)}$. Conversely, let $\ell \in \mathcal{X}^*$ and let $tx + (1-t)y \in \mathrm{conv}(S)$ where $t \in (0,1)$ and $x, y \in S$. Then observe that
$$\begin{aligned} \ell(tx+(1-t)y) &= t\ell(x) + (1-t)\ell(y) \\ &\leq t\sup_{z\in S}\ell(z) + (1-t)\sup_{z\in S}\ell(z) \\ &= t\mathrm{q}_S(\ell) + (1-t)\mathrm{q}_S(\ell) \\ &= \mathrm{q}_S(\ell). \end{aligned}$$
It follows that $\mathrm{q}_{\mathrm{conv}(S)}(\ell) \leq \mathrm{q}_S(\ell)$, and since ℓ was arbitrary, we have $\mathrm{q}_{\mathrm{conv}(S)} \leq \mathrm{q}_S$. To show $\mathrm{q}_{\overline{\mathrm{conv}}(S)} \leq \mathrm{q}_{\mathrm{conv}(S)}$, we will prove something more general: if E is a subset of $\mathcal{X}$, then $\mathrm{q}_{\overline{E}} \leq \mathrm{q}_E$. Indeed, let $\ell \in \mathcal{X}^*$, let $x \in \overline{E}$, and choose a sequence (x_n) of elements in E such that $x_n \to x$. Then observe that
$$\begin{aligned} \ell(x) &= \lim_{n\to\infty} \ell(x_n) \\ &\leq \sup_{y\in E} \ell(y) \\ &= \mathrm{q}_E(\ell). \end{aligned}$$
It follows that $\mathrm{q}_{\overline{E}}(\ell) \leq \mathrm{q}_E(\ell)$, and since ℓ was arbitrary, we have $\mathrm{q}_{\overline{E}} \leq \mathrm{q}_E$.

3. Let $x_1 + x_2 \in S_1 + S_2$ and let $\ell \in \mathcal{X}^*$. Then observe that

$$\begin{aligned}\ell(x_1 + x_2) &= \ell(x_1) + \ell(x_2) \\ &\leq \sup_{y_1 \in S_1} \ell(y_1) + \sup_{y_2 \in S_2} \ell(y_2) \\ &= \mathrm{q}_{S_1}(\ell) + \mathrm{q}_{S_2}(\ell) \\ &= (\mathrm{q}_{S_1} + \mathrm{q}_{S_2})(\ell)\end{aligned}$$

It follows that $\mathrm{q}_{S_1+S_2}(\ell) \leq (\mathrm{q}_{S_1} + \mathrm{q}_{S_2})(\ell)$, and since ℓ was arbitrary, we have $\mathrm{q}_{S_1+S_2} \leq \mathrm{q}_{S_1} + \mathrm{q}_{S_2}$. Conversely, let $\ell \in \mathcal{X}^*$, let $\varepsilon > 0$, and choose $x_1 \in S_1$ and $x_2 \in S_2$ such that $\ell(x_1) + \varepsilon/2 > \mathrm{q}_{S_1}(\ell)$ and $\ell(x_2) + \varepsilon/2 > \mathrm{q}_{S_2}(\ell)$. Then observe that

$$\begin{aligned}(\mathrm{q}_{S_1} + \mathrm{q}_{S_2})(\ell) &= \mathrm{q}_{S_1}(\ell) + \mathrm{q}_{S_2}(\ell) \\ &< \ell(x_1) + \frac{\varepsilon}{2} + \ell(x_2) + \frac{\varepsilon}{2} \\ &= \ell(x_1) + \ell(x_2) + \varepsilon \\ &= \ell(x_1 + x_2) + \varepsilon \\ &\leq \mathrm{q}_{S_1+S_2}(\ell) + \varepsilon.\end{aligned}$$

By taking $\varepsilon \to 0$, we see that $(\mathrm{q}_{S_1} + \mathrm{q}_{S_2})(\ell) \leq \mathrm{q}_{S_1+S_2}(\ell)$, and since ℓ was arbitrary, we have $\mathrm{q}_{S_1} + \mathrm{q}_{S_2} \leq \mathrm{q}_{S_1+S_2}$.

4. Let $\ell \in \mathcal{X}^*$. First suppose that $\ell \in \mathcal{K}^\perp$. Then $\ell(x) = 0$ for all $x \in \mathcal{K}$. Thus

$$\begin{aligned}\mathrm{q}_{\mathcal{K}}(\ell) &= \sup_{x \in \mathcal{K}} \ell(x) \\ &= \sup_{x \in \mathcal{K}} 0 \\ &= 0.\end{aligned}$$

Now suppose that $\ell \notin \mathcal{K}^\perp$. Choose $x \in \mathcal{K}$ such that $\ell(x) \neq 0$ and let $\lambda \geq 0$. Then observe that

$$\begin{aligned}\lambda \ell(x) &= \ell(\lambda x) \\ &\leq \sup_{y \in \mathcal{K}} \ell(y) \\ &= \mathrm{q}_{\mathcal{K}}(\ell).\end{aligned}$$

Taking $\lambda \to \infty$ gives us $\mathrm{q}_{\mathcal{K}}(\ell) = \infty$. □

44.4.2 Examples of Support Functionals

Example 44.4. Suppose $C = \{x_0\}$. Then $\mathrm{q}_{\{x_0\}}(\ell) = \ell(x_0)$.

Example 44.5. Suppose $C = \mathrm{B}_1[0]$, then $\mathrm{q}_{\mathrm{B}_1[0]} = \|\ell\|$.

Example 44.6. Suppose $C = \mathrm{B}_R[0]$, then $\mathrm{q}_{\mathrm{B}_R[0]} = R\|\ell\|$. Recall that the gauge functional in this case is $\mathrm{p}_{\mathrm{B}_R[0]}(x) = \|x\|/R$. More generally, we have

$$\begin{aligned}\mathrm{q}_{\mathrm{B}_R[x_0]}(x) &= \mathrm{q}_{\{x_0\}+\mathrm{B}_R[0]}(x) \\ &= \mathrm{q}_{\{x_0\}}(x) + \mathrm{q}_{\mathrm{B}_R[0]}(x) \\ &= \ell(x_0) + R\|\ell\|.\end{aligned}$$

If $\mathcal{M}$ is a closed subspace of $\mathcal{X}$, then

$$\mathrm{q}_{\mathcal{M}}(\ell) = \begin{cases} 0 & \text{if } \ell \in \mathcal{M}^\perp \\ \infty & \text{else} \end{cases}$$

Let $\varphi(x) = \mathrm{I}_E(x)$ for some set $E \subseteq \mathcal{X}$. Then

$$\begin{aligned}\varphi^*(\ell) &= \sup_{x \in \mathcal{X}} (\ell(x) - \mathrm{I}_E(x)) \\ &= \sup_{x \in E} \ell(x) \\ &= \mathrm{q}_E(\ell).\end{aligned}$$

Notice $\varphi^*(\ell) = \mathrm{q}_{\overline{\mathrm{conv}}(E)}(\ell)$. Then

$$\begin{aligned}
\varphi^{**}(x) &= \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \varphi^*(\ell)) \\
&= \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \mathrm{q}_E(\ell)) \\
&= \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \mathrm{q}_{\overline{\mathrm{conv}}(E)}(\ell))
\end{aligned}$$

It can be shown that I_E is convex if and only if E is convex. It can also be shown that I_E is lower semicontinuous if and only if E is closed. So if E is clsoed and convex, then Fenchel-Moreau applies and we get

$$\mathrm{I}_E(x) = \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \mathrm{q}_E(\ell)).$$

In some sense, the gauge (Minkowski) functional p_C plays the role of a norm if we want C convex to play the role of the unit ball. In that sense, the support functional q_C plays the role of the norm in the dual space $\mathcal{X}^*$. In this direct, the Cauchy-Schwarz inequality $|\ell(x)| \leq \|\ell\|\|x\|$ is replaced by

$$|\ell(x)| \leq \mathrm{q}_C(\ell)\mathrm{p}_C(x) \tag{116}$$

for all $\ell \in \mathcal{X}^*$ and $x \in \mathcal{X}$. Indeed, for any $x \in \mathcal{X}$ and $\varepsilon > 0$ we have $x/(\mathrm{p}_C(x) + \varepsilon) \in C$ by definition of $\mathrm{p}_C(x)$, and thus

$$\ell\left(\frac{1}{\mathrm{p}_C(x) + \varepsilon} x\right) \leq \sup_{y \in C} \ell(y) = \mathrm{q}_C(\ell)$$

which implies (116).

Proposition 44.5. *$x \in \overline{\mathrm{conv}}(E)$ if and only if $\ell(x) \leq \mathrm{q}_E(\ell)$ for all $\ell \in \mathcal{X}^*$.*

Proof. Recall that $\mathrm{I}_E^*(\ell) = \mathrm{q}_E(\ell) = \mathrm{q}_{\overline{\mathrm{conv}}(E)}(\ell) = \mathrm{I}^*_{\overline{\mathrm{conv}}(E)}$. We have

$$\begin{aligned}
\mathrm{I}_E^{**}(x) &= \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \mathrm{I}_E^*(\ell)) \\
&= \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \mathrm{q}_E(\ell)).
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
\mathrm{I}_E^{**}(x) &= \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \mathrm{I}^*_{\overline{\mathrm{conv}}(E)}(\ell)) \\
&= \mathrm{I}^{**}_{\overline{\mathrm{conv}}(E)}(x)
\end{aligned}$$

We can apply Fenchel-Moreau to $\mathrm{I}_{\overline{\mathrm{conv}}(E)}$ which is convex and lowersemicontinuous and obtain

$$\mathrm{I}_{\overline{\mathrm{conv}}(E)}(x) = \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \mathrm{q}_E(\ell))$$

So

$$\begin{aligned}
x \in \overline{\mathrm{conv}}(E) &\iff \mathrm{I}_{\overline{\mathrm{conv}}(E)}(x) = 0 \\
&\iff \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \mathrm{q}_E(\ell)) = 0 \\
&\iff \ell(x) \leq \mathrm{q}_E(\ell) \text{ for all } \ell \in \mathcal{X}^*.
\end{aligned}$$

□

44.5 Another Application

For a subspace $\mathcal{M} \subseteq \mathcal{X}$, we define its **annihilator** by

$$\mathcal{M}^{\perp} = \{\ell \in \mathcal{X}^* \mid \ell|_{\mathcal{M}} = 0\}.$$

For a closed subspace $\mathcal{N} \subseteq \mathcal{X}^*$, we define

$$\mathcal{N}_{\perp} = \{x \in \mathcal{X} \mid \ell(x) = 0 \text{ for all } \ell \in \mathcal{N}\}.$$

Proposition 44.6. *If $\mathcal{M} \subseteq \mathcal{X}$ is a closed subspace, then $(\mathcal{M}^{\perp})_{\perp} = \mathcal{M}$.*

Proof. We have $\mathrm{I}_{\mathcal{M}}^*(\ell) = \mathrm{q}_{\mathcal{M}}(\ell) = \mathrm{I}_{\mathcal{M}^{\perp}}(\ell)$. So

$$\begin{aligned}
\mathrm{I}_{\mathcal{M}}(x) &= \mathrm{I}_{\mathcal{M}}^{**}(x) \\
&= \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \mathrm{I}_{\mathcal{M}}^*(\ell)) \\
&= \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \mathrm{I}_{\mathcal{M}^{\perp}}(\ell)) \\
&= \sup_{\ell \in \mathcal{M}^{\perp}} (\ell(x)) \\
&= \mathrm{I}_{(\mathcal{M}^{\perp})_{\perp}}(x)
\end{aligned}$$

□

44.6 Fenchel-Rockafeller

Theorem 44.7. *(Fenchel-Rockafellar) Let $\varphi, \psi \colon \mathcal{X} \to (-\infty, \infty]$ be two convex functions. Suppose there exists $x_0 \in \mathcal{X}$ such that $\varphi(x_0), \psi(x_0) < \infty$ and φ is continuous at x_0. Then*

$$\inf_{x \in \mathcal{X}} (\varphi(x) + \psi(x)) = \sup_{\ell \in \mathcal{X}^*} (-\varphi^*(-\ell) - \psi^*(\ell)) = \max_{\ell \in \mathcal{X}^*} (-\varphi^*(-\ell) - \psi^*(\ell)).$$

Proof. (Sketch) Let $a = \inf_{x \in \mathcal{X}} (\varphi(x) + \psi(x))$ and let $b = \sup_{\ell \in \mathcal{X}^*} (-\varphi^*(-\ell) - \psi^*(\ell))$. It's easy to see that $b \le a$. Indeed,

$$\begin{aligned}
-\varphi^*(\ell) - \psi^*(\ell) &= -\varphi^*(-\ell) - (-\ell(x)) - \psi^*(\ell) - \ell(x) \\
&\le \varphi(x) + \psi(x)
\end{aligned}$$

for all $x \in \mathcal{X}$ and $\ell \in \mathcal{X}^*$. For the reverse direction, let $C = \operatorname{epi} \varphi$, let $B = \{(x, \lambda) \mid \lambda \le a - \psi(x)\}$, and let $A = \operatorname{int} C$. Then A and B are both nonempty convex sets. Furthemore we have $A \cap B = \varnothing$ (otherwise we'll have $(x, \lambda) \in \mathcal{X} \times \mathbb{R}$ such that $\varphi(x) < \lambda \le a - \psi(x)$ which implies $\varphi(x) + \psi(x) < a$, giving a contradiction). Applying Hahn-Banach, we obtain a linear functional $\Phi \colon \mathcal{X} \times \mathbb{R} \to \mathbb{R}$ such that $\overline{C} = \overline{A} \subseteq \{\Phi \ge \alpha\}$ and $B \subseteq \{\Phi \le \alpha\}$. Let $\ell(x) = \Phi(x, 0)$ and $k = \Phi(0, 1) \in \mathbb{R}$. Then

$$\begin{aligned}
\ell(x) + k\lambda &\ge \alpha \text{ for } (x, \lambda) \in \overline{A} = \overline{C} \\
\ell(x) + k\lambda &\le \alpha \text{ for } (x, \lambda) \in B.
\end{aligned}$$

Similarly as before, one can show that $k > 0$. □

44.6.1 Application

Let $C \subseteq \mathcal{X}$ be non-empty and convex. Then

$$\mathrm{d}(x_0, C) = \inf_{x \in C} \|x_0 - x\| = \sup_{\substack{\ell \in \mathcal{X}^* \\ \|\ell\| \le 1}} (\ell(x_0) - \mathrm{q}_C(\ell)).$$

Then $\varphi(x) = \|x - x_0\|$ is convex and $\psi(x) = \mathrm{I}_C(x)$ is convex if C is convex. Then

$$\begin{aligned}
\varphi^*(\ell) &= \sup_{x \in \mathcal{X}} (\ell(x) - \|x - x_0\|) \\
&= \sup_{x \in \mathcal{X}} (\ell(x - x_0) - \|x - x_0\| + \ell(x_0)) \\
&= \varphi^*(\ell) \\
&= \mathrm{I}_{\mathrm{B}_1[0]}(\ell) + \ell(x_0).
\end{aligned}$$

So by Fenchel-Rockafellar, we have

$$\inf_{x\in\mathcal{X}} (\|x-x_0\| + \mathrm{I}_C(x)) = \sup_{\ell\in\mathcal{X}^*} (\ell(x_0) - \mathrm{I}_{\mathrm{B}_1[0]}(-\ell) - \mathrm{q}_C(\ell))$$

Before starting the proof, recall that we proved last time using Fenchel-Rockafellar that if $C \neq \emptyset$ is convex, then

$$\mathrm{d}(x_0, C) = \sup_{\substack{\ell\in\mathcal{X}^* \\ \|\ell\|\leq 1}} (\ell(x_0) - \mathrm{q}_C(\ell)).$$

Note that when $C = \mathcal{M}$ is a subspace, we have

$$\mathrm{d}(x_0, C) = \sup_{\substack{\ell\in\mathcal{X}^* \\ \|\ell\|\leq 1}} (\ell(x_0) - \mathrm{I}_{\mathcal{M}^\perp}(\ell)) = \sup_{\substack{\ell\in\mathcal{M}^\perp \\ \|\ell\|\leq 1}} \ell(x).$$

45 Baire Category Theorem

Theorem 45.1. *Let $\mathcal{X}$ be a Banach space. Then $\mathcal{X}$ cannot be represented as a countable union of nowhere dense sets.*

Recall that a set $E \subseteq \mathcal{X}$ is said to be nowhere dense if $(\overline{E})^\circ = \emptyset$. In other words, $\overline{E}$ doesn't contain any open balls.

Proof. Assume for a contradiction that $\mathcal{X} = \bigcup_{n=1}^\infty E_n$ with every E_n being nowhere dense. In particular, we have $\mathcal{X} = \bigcup_{n=1}^\infty \overline{E}_n$. Let $\mathrm{B}_{r_1}(x_1) \subseteq \mathcal{X}$ be any open ball. Since E_1 is nowhere dense, it follows that $\mathrm{B}_{r_1}(x_1) \cap \overline{E}_1^c$ is a nonempty open set. Thus there exists an open ball, say $\mathrm{B}_{r_2}(x_2)$, such that $\mathrm{B}_{r_2}[x_2] \subseteq \mathrm{B}_{r_1}(x_1) \cap \overline{E}_1^c$ and $r_2 < 2^{-2}$. Since E_2 is nowhere dense, it follows that $\mathrm{B}_{r_2}(x_2) \cap \overline{E}_2^c$ is a nonempty open set. So by the same reason as before, there exists an open ball, say $\mathrm{B}_{r_3}(x_3)$, such that $\mathrm{B}_{r_3}[x_3] \subseteq \mathrm{B}_{r_2}(x_2) \cap \overline{E}_2^c$ and $r_3 < 2^{-3}$. Continuing this process, we obtain a descending sequence of open balls $(\mathrm{B}_{r_n}(x_n))$ such that

$$\mathrm{B}_{r_n}[x_n] \subseteq \mathrm{B}_{r_{n-1}}(x_{n-1}) \cap \overline{E}_{n-1}^c \quad \text{and} \quad r_n < 2^{-n}$$

for all $n \in \mathbb{N}$.

Now let $\varepsilon > 0$ and choose $N \in \mathbb{N}$ such that $2^{-N} < \varepsilon$. Then $n > m \geq N$ implies

$$\begin{aligned}\|x_m - x_n\| &\leq r_m \\ &< 2^{-m} \\ &\leq 2^{-N} \\ &< \varepsilon.\end{aligned}$$

Thus (x_n) is a Cauchy sequence. Being a Cauchy sequence in a Banach space, we see that (x_n) is convergent, say $x_n \to x$. Since $x_n \in \mathrm{B}_{r_k}(x_k)$ for any $n \geq k$, we have $x \in \mathrm{B}_{r_k}[x_k]$. In particular, this implies

$$\begin{aligned}x &\in \bigcap_{n=1}^\infty \mathrm{B}_{r_n}[x_n] \\ &\subseteq \bigcap_{n=1}^\infty \overline{E}_n^c \\ &= \left(\bigcup_{n=1}^\infty \overline{E}_n\right)^c \\ &= \mathcal{X}^c \\ &= \emptyset,\end{aligned}$$

which is a contradiction. $\square$

45.1 Uniform Boundedness Principle

Theorem 45.2. *(Uniform Boundedness Principle) Let $\mathcal{X}$ and $\mathcal{Y}$ be two Banach spaces. Denote by $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ the set of all bounded linear operators from $\mathcal{X}$ to $\mathcal{Y}$. Suppose $\mathcal{A} \subseteq \mathcal{L}(X, \mathcal{Y})$ such that for any $x \in \mathcal{X}$ the set $\{\|Tx\| \mid T \in \mathcal{A}\}$ is bounded above. Then the set $\{\|T\| \mid T \in \mathcal{A}\}$ is bounded above.*

Proof. For each $n \in \mathbb{N}$, let

$$E_n = \{x \in \mathcal{X} \mid \|Tx\| \leq n \text{ for all } T \in \mathcal{A}\}.$$

Observe that (E_n) is an ascending sequence of closed sets. Indeed, it is clearly ascending. To see that each E_n is closed, view it as an infinite intersection of closed sets, namely

$$E_n = \bigcap_{T \in \mathcal{A}} \{x \in \mathcal{X} \mid \|Tx\| \leq n\}.$$

Moreover, for any $x \in \mathcal{X}$ the set $\{\|Tx\| \mid T \in \mathcal{A}\}$ is bounded above, say $\{\|Tx\| \mid T \in \mathcal{A}\} \leq N$ for some $N \in \mathbb{N}$. It follows that $x \in E_N$ and since $x \in \mathcal{X}$ was arbitrary, we see that

$$\mathcal{X} = \bigcup_{n=1}^{\infty} E_n.$$

By the Baire Category Theorem, there must exist some $M \in \mathbb{N}$ such that E_M is not nowhere dense. In other words, E_M contains a nonempty contains a nonempty open ball, say $\mathrm{B}_r(x_0)$. By choosing r small enough, we can assume $\mathrm{B}_r[x_0] \subseteq E_M$. Then for any $x \in \mathrm{B}_1[0]$, we have

$$\begin{aligned}
\|T(rx)\| &\leq \|T(rx + x_0) - Tx_0\| \\
&\leq \|T(rx + x_0)\| + \|Tx_0\| \\
&\leq M + M \\
&= 2M
\end{aligned}$$

for all $T \in \mathcal{A}$. It follows that $\|T\| \leq 2M/r$ for all $T \in \mathcal{A}$. Thus the set $\{\|T\| \mid T \in \mathcal{A}\}$ is bounded above. □

Here is a simple application of the uniform boundedness principle.

Proposition 45.1. *Let (T_n) be a sequence of bounded linear operators $T_n \colon \mathcal{X} \to \mathcal{Y}$ between Banach spaces $\mathcal{X}$ and $\mathcal{Y}$. Assume for each $x \in \mathcal{X}$ the sequence $(T_n x)$ converges in $\mathcal{Y}$. Then the map $T \colon \mathcal{X} \to \mathcal{Y}$ defined by*

$$Tx := \lim_{n \to \infty} T_n x$$

for all $x \in \mathcal{X}$ is a bounded linear operator.

Proof. Since for each $x \in \mathcal{X}$ the sequence $(T_n x)$ is convergent we see that it must be bounded. Let $M_x = \sup_{n \in \mathbb{N}} \|T_n x\| < \infty$. By the uniform boundedness principle, there exists $M > 0$ such that $\sup_{n \in \mathbb{N}} \|T_n\| \leq M < \infty$. Therefore

$$\begin{aligned}
\|Tx\| &= \|\lim_{n \to \infty} T_n x\| \\
&= \lim_{n \to \infty} \|T_n x\| \\
&\leq \sup_{n \in \mathbb{N}} \|T_n x\| \\
&\leq \sup_{n \in \mathbb{N}} \|T_n\| \|x\| \\
&\leq M\|x\|.
\end{aligned}$$

It follows that T is bounded. □

46 Open Mapping Theorem and Closed Graph Theorem

46.1 Main Theorem

Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces. Consider the space $\mathcal{X} \times \mathcal{Y}$ with addition and scalar-multiplication defined pointwise. We endow $\mathcal{X} \times \mathcal{Y}$ with a norm defined by

$$\|(x,y)\|_{\mathcal{X}\times\mathcal{Y}} = \|x\|_{\mathcal{X}} + \|y\|_{\mathcal{Y}}. \tag{117}$$

for all $(x,y) \in \mathcal{X}$. It's easy to prove that $(\mathcal{X} \times \mathcal{Y}, \|\cdot\|_{\mathcal{X}\times\mathcal{Y}})$ is a Banach space. If context is clear, then we drop $\mathcal{X} \times \mathcal{Y}$ from the subscript in $\|\cdot\|_{\mathcal{X}\times\mathcal{Y}}$ in order to clean notation. We'll use the usual projection maps $\pi_1 \colon \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ and $\pi_2 \colon \mathcal{X} \times \mathcal{Y} \to \mathcal{Y}$ defined by $\pi_1(x,y) = x$ and $\pi_2(x,y) = y$. Clearly both π_1 and π_2 are bounded linear operators.

Theorem 46.1. *(Main result) Let $\mathcal{Z} \subseteq \mathcal{X} \times \mathcal{Y}$ be a closed subspace such that $\pi_2(\mathcal{Z}) = \mathcal{Y}$. If $U \subseteq \mathcal{X}$ is open, then $\pi_2(\pi_1^{-1}(U) \cap \mathcal{Z}))$ is an open subset of $\mathcal{Y}$.*

Remark 63. Note that by symmetry if instead of assuming $\pi_2(\mathcal{Z}) = \mathcal{Y}$ we assume $\pi_1(\mathcal{Z}) = \mathcal{X}$, then we have for any open set $V \subseteq \mathcal{Y}$ we have $\pi_1(\pi_2^{-1}(V) \cap \mathcal{Z}))$ is an open subset of $\mathcal{X}$.

46.2 Applications of the Main Theorem

Before we prove Theorem (46.1), let us show how to use it to prove both the open mapping theorem and the closed graph theorem.

46.2.1 Open Mapping Theorem

Theorem 46.2. *(Open mapping theorem) Let $T \colon \mathcal{X} \to \mathcal{Y}$ be a surjective bounded linear operator. Then T is an open map, meaning that for any open subset U of X, the set $T(U)$ is an open subset of $\mathcal{Y}$.*

Proof. Let $\mathcal{Z} = \{(x, Tx) \mid x \in \mathcal{X}\} \subseteq \mathcal{X} \times \mathcal{Y}$ and let U be an open subset of $\mathcal{X}$. Observe that $\mathcal{Z}$ is a closed subspace precisely because T is a bounded linear operator. Furthemore we have $\pi_2(\mathcal{Z}) = \mathcal{Y}$ since T is surjective. Finally, note that $T(U) = \pi_2(\pi_1^{-1}(U) \cap \mathcal{Z}))$. It follows from Theorem (46.1) that $T(U)$ is an open subset of $\mathcal{Y}$. □

46.2.2 Inverse Mapping Theorem

Theorem 46.3. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces and let $T \colon \mathcal{X} \to \mathcal{Y}$ be a bounded linear map which is bijective. Then $T^{-1} \colon \mathcal{Y} \to \mathcal{X}$ is also a bounded linear map.*

Proof. That T^{-1} is linear follows from basic linear algebra. The nontrivial part is that T^{-1} is also bounded. To see why, it suffices to show that T^{-1} is continuous. Let $U \subseteq \mathcal{X}$ be open. Then its preimage under T^{-1} is $T(U)$ since T is bijective. Since T is onto, it follows from the open mapping theorem, that $T(U)$ is open. Thus T^{-1} is continuous. □

46.2.3 Closed Graph Theorem

Theorem 46.4. *Let $T \colon \mathcal{X} \to \mathcal{Y}$ be a linear map such that $x_n \to x$ and $Tx_n \to y$ implies $y = Tx$, or in other words, if the graph of T given by $\{(x, Tx) \mid x \in \mathcal{X}\} \subseteq \mathcal{X} \times \mathcal{Y}$ is a closed set, then T is bounded.*

Proof. Again take $\mathcal{Z} = \{(x, Tx) \mid x \in \mathcal{X}\} \subseteq \mathcal{X} \times \mathcal{Y}$ and let V be an open subset of $\mathcal{Y}$. Since T is linear, $\mathcal{Z}$ is a subspace of $\mathcal{X} \times \mathcal{Y}$. Furthermore, $\mathcal{Z}$ is closed by assumption. Also we clearly have $\pi_1(\mathcal{Z}) = \mathcal{X}$. Finally, note that $T^{-1}(V) = \pi_1(\pi_2^{-1}(V) \cap \mathcal{Z}))$. It follows from Theorem (46.1) that $T^{-1}(V)$ is an open subset of $\mathcal{X}$. Thus T is continuous, and hence bounded. □

46.3 Zabreiko's Lemma

The proof of Theorem (46.1) will depend on the following lemma:

Lemma 46.5. *(Zabreiko) Let $\mathcal{X}$ be a Banach space and let $p \colon \mathcal{X} \to [0, \infty)$ be a seminorm on $\mathcal{X}$. Suppose p is* ***countably subadditive****, that is, suppose for every absolutely convergent series $\sum_{n=1}^{\infty} x_n$ in $\mathcal{X}$, we have*

$$p\left(\sum_{n=1}^{\infty} x_n\right) \leq \sum_{n=1}^{\infty} p(x_n).$$

Then there exists $C > 0$ such that $p(x) \leq C\|x\|$ for every $x \in \mathcal{X}$.

Proof. For each $n \in \mathbb{N}$, let $E_n = \{p \leq n\}$.

Step 1: We will find an $N \in \mathbb{N}$ and $r > 0$ such that $\mathrm{B}_r(0) \subseteq \overline{E}_N$. Observe that E_n is convex and symmetric (here symmetric means $x \in E_n$ implies $-x \in E_n$). From here it is easy to show that $\overline{E}_n$ is convex, symmetric, and closed. Clearly

$$\mathcal{X} = \bigcup_{n=1}^{\infty} \overline{E}_n.$$

So by the Baire category theorem, there exists an $N \in \mathbb{N}$ and an open ball $\mathrm{B}_r(x_0)$ such that $\mathrm{B}_r(x_0) \subseteq \overline{E}_N$. Since $\overline{E}_N$ is symmetric, we have $\mathrm{B}_r(-x_0) \subseteq \overline{E}_N$. Then for each $x \in \mathrm{B}_r(0)$, we have

$$x = \frac{1}{2}(x - x_0) + \frac{1}{2}(x + x_0)$$

where $x - x_0 \in \mathrm{B}_r(-x_0) \subseteq \overline{E}_N$ and $x + x_0 \in \mathrm{B}_r(x_0) \subseteq \overline{E}_N$. Since $\overline{E}_N$ is convex, it follows that $x \in \overline{E}_N$. Therefore $\mathrm{B}_r(0) \subseteq \overline{E}_N$.

Step 2: We will show $\mathrm{B}_r(0) \subseteq E_N$. Let $x \in \mathrm{B}_r(0)$, let $\rho > 0$ such that $\|x\| < \rho < r$, let $q > 0$ such that $q < 1 - \rho/r$, and let $y = (r/\rho)x$. Then observe that $y \in \mathrm{B}_r(0) \subseteq \overline{E}_N$. In particular, this implies $\mathrm{B}_{qr}(y) \cap E_N \neq \emptyset$, so we can choose $y_0 \in \mathrm{B}_{qr}(y) \cap E_N$. Since $y_0 \in \mathrm{B}_{qr}(y)$, we have

$$\|y_0 - y\| < qr$$

In other words, dividing both sides by q give us $(y - y_0)/q \in \mathrm{B}_r(0) \subseteq \overline{E}_N$. In particular, this implies $\mathrm{B}_{qr}((y - y_0)/q) \cap E_N \neq \emptyset$, so we can choose $y_1 \in \mathrm{B}_{qr}((y - y_0)/q) \cap E_N$. Again since $y_1 \in \mathrm{B}_{qr}((y - y_0)/q)$, we have

$$\left\|\frac{y - y_0 - qy_1}{q}\right\| < qr$$

In other words, dividing both sides by q gives us $(y - y_0 - qy_1)/q^2 \in \mathrm{B}_r(0) \subseteq \overline{E}_N$. In particular, this implies $\mathrm{B}_{qr}((y - y_0 - qy_1)/q^2) \cap E_N \neq \emptyset$, so we can choose $y_2 \in \mathrm{B}_{qr}((y - y_0 - qy_1)/q^2) \cap E_N$. More generally, for each $n \geq 2$, we choose

$$y_n \in \mathrm{B}_{qr}\left(\frac{y - y_0 - qy_1 - \cdots - q^{n-1}y_{n-1}}{q^n}\right).$$

In this case, we obtain a sequence $(y_n) \subseteq E_N$ such that

$$\|y - y_0 - qy_1 - q^2y_2 - \cdots - q^n y_n\| < q^n r \tag{118}$$

for all $n \in \mathbb{N}$. Since $\|y_n\| \leq r + qr$ for all $n \in \mathbb{N}$ and $0 < q < 1$, we have $\sum_{n=0}^{\infty} q^n y_n$ is absolutely convergent.

Therefore by (118) we have $y = \sum_{n=1}^{\infty} q^n y_n$. Thus\

$$\begin{aligned}
p(x) &= p\left(\frac{\rho}{r} y\right) \\
&= \frac{\rho}{r} p(y) \\
&= \frac{\rho}{r} p\left(\sum_{n=1}^{\infty} q^n y_n\right) \\
&\leq \frac{\rho}{r} q^n \sum_{n=1}^{\infty} p(y_n) \\
&\leq \frac{\rho}{r} q^n N \\
&= \frac{\rho}{r} \frac{N}{1-q} \\
&\leq N.
\end{aligned}$$

It follows that $\mathrm{B}_r(0) \subseteq E_N$.

Step 3: Let $x \in \mathcal{X}$ be arbitrary nonzero. Then $(r/2)x/\|x\| \in \mathrm{B}_r(0)$ and hence $p((r/2)x/\|x\|) \leq N$. This implies $p(x) \leq (2N/r)\|x\|$. □

Remark 64. We make two remarks.

1. Zabreiko's lemma implies p is continuous. Indeed, suppose $x_n \to x$. Then

$$\begin{aligned}
|p(x_n) - p(x)| &\leq p(x_n - x) \\
&\leq C\|x_n - x\| \\
&\to 0.
\end{aligned}$$

2. Zabreiko's lemma can be used to prove the uniform boundedness principle. Indeed, take $p(x) = \sup_{T \in \mathcal{A}} \|Tx\|$. Then it can be shown that p satisfies the properties from Zabreiko's lemma. Therefore there exist $C > 0$ such that

$$\sup_{T \in \mathcal{A}} \|Tx\| \leq C\|x\|.$$

   Thus for any $T \in \mathcal{A}$ we have $\|Tx\| \leq C\|x\|$ which implies $\|T\| \leq C$ for all $T \in \mathcal{A}$.

46.4 Proof of Main Theorem

We now wish to prove Theorem (46.1).

Proof. Let $p \colon \mathcal{Y} \to [0, \infty)$ be defined by

$$p(y) := \inf\{\|x\| \mid (x, y) \in \mathcal{Z}\}.$$

It's easy to show that p is a seminorm. We claim that it is also countably subadditive. Indeed, let $\sum_{n=1}^{\infty} y_n$ be an absolutely convergent series such that $\sum_{n=1}^{\infty} p(y_n) < \infty$. Let $\varepsilon > 0$ and for each $n \in \mathbb{N}$ choose $x_n \in \mathcal{X}$ such that $\|x_n\| < p(y_n) + \varepsilon/2^n$ and $(x_n, y_n) \in \mathcal{Z}$. Then

$$\sum_{n=1}^{\infty} \|x_n\| \leq \sum_{n=1}^{\infty} p(y_n) + \varepsilon < \infty.$$

Hence $\sum_{n=1}^{\infty} x_n$ is absolutely convergent. Since $\mathcal{Z}$ is a subspace, we have $(\sum_{n=1}^{N} x_n, \sum_{n=1}^{N} y_n) \in \mathcal{Z}$ for all $N \in \mathbb{N}$.

Since $\mathcal{Z}$ is closed, we have $(\sum_{n=1}^{\infty} x_n, \sum_{n=1}^{\infty} y_n) \in \mathcal{Z}$. Then

$$\begin{aligned} p\left(\sum_{n=1}^{\infty} y_n\right) &= \inf\left\{ \|x\| \mid \left(x, \sum_{n=1}^{\infty} y_n\right) \in \mathcal{Z} \right\} \\ &\leq \left\| \sum_{n=1}^{\infty} x_n \right\| \\ &\leq \sum_{n=1}^{\infty} \|x_n\| \\ &\leq \sum_{n=1}^{\infty} p(y_n) + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, it follows that p is countably subadditive. So we can apply Zabreiko's lemma to obtain that p is continuous.

Now let $U = \mathrm{B}_1(0)$ be the open unit ball in $\mathcal{Y}$. Then

$$\begin{aligned} \pi_2(\pi^{-1}(\mathrm{B}_1(0) \cap \mathcal{Z})) &= \pi_2\{(x, y) \mid x \in \mathrm{B}_1(0) \text{ and } (x, y) \in \mathcal{Z}) \\ &= \{y \in \mathcal{Y} \mid p(y) < 1\} \\ &= \{p < 1\}. \end{aligned}$$

Implies $\pi_2(\pi^{-1}(\mathrm{B}_1(0) \cap \mathcal{Z}))$ is open since p is continuous. The general case open sets U can be easily be obtained using linearity and homogeneity. □

47 Riesz Representation Theorem Revisited

In this note, we wish to discuss a proof of a Riesz representation theorem due to Garling. Throughout this note, let X be a compact Hausdorff space and let $C(X)$ be the Banach of space of continuous real-valued functions defined on X equipped with the supremum norm.

47.1 The Baire σ-algebra

Definition 47.1. Let $\mathcal{C}$ be the collection of all subsets of X of the form

$$\{a \leq f \leq b\} := \{x \in X \mid a \leq f(x) \leq b\},$$

where $a, b \in \mathbb{R}$. The **Baire σ-algebra of** X, denoted by $\mathcal{M}_X$, or just $\mathcal{M}$ if X is understood from context, is the σ-algebra generated $\mathcal{C}$, written $\mathcal{M} = \sigma(\mathcal{C})$. The members of $\mathcal{M}$ are called **Baire sets**. A measure defined on $\mathcal{M}$ whose value on every compact Baire set is called a **Baire measure**. Note that in our case, we are already assuming that X is compact, and thus Baire measures correspond to finite measures defined on $\mathcal{M}$.

The are more general definitions of Baire measures in the case where X is not compact, but we will not pursue this direction since we are always assuming X is compact throughout this document. Note that the Baire σ-algebra of X is the *smallest* σ-algebra which makes every continuous function $f\colon X \to \mathbb{R}$ Baire measurable. Recall that every continuous function $f\colon X \to \mathbb{R}$ is Borel measurable. Thus we certainly have $\mathcal{M} \subseteq \mathcal{B}$ where $\mathcal{B}$ is the Borel σ-algebra of X.

47.1.1 The Baire σ-algebra of X is generated by all G_δ-sets

The Baire σ-algebra of X can be generated by another useful collection of special subsets of X, namely the **G_δ-sets**:

Definition 47.2. Let A be a subset of X. We say A is a **G_δ-subset of** X, or just a **G_δ-set** if X is understood from context, if it can be expressed as a countable intersection of open subsets of X.

Proposition 47.1. *Let $\mathcal{D}$ be the collection of all compact G_δ-subsets of X. Then $\mathcal{M} = \sigma(\mathcal{D})$.*

Proof. Let K be a compact G_δ-set and it express it as a countable intersection of open sets, say

$$K = \bigcap_{n=1}^{\infty} U_n.$$

For each $n \in \mathbb{N}$, there exists a continuous function $f_n \colon X \to [0,1]$ such that $f_n|_K = 1$ and $f_n|_{X \setminus U_n} = 0$ by Urysohn's lemma. Clearly the sequence of functions (f_n) converges pointwise to the characteristic function 1_K, and since each f_n is Baire measurable, it follows that 1_K is Baire measurable. In particular, K is a Baire measure since $K = \{1_K = 1\}$. Since $K \in \mathcal{D}$ was arbitrary, it follows that $\sigma(\mathcal{D}) \subseteq \mathcal{M}$.

Conversely, let $f \in C(X)$, let $c \in \mathbb{R}$, and let $A = \{a \leq f \leq b\}$. Then A is a closed subset of X since f is continuous. Note that closed subsets of compact spaces are themselves compact. Indeed, suppose $(U_i)_{i \in I}$ is an open covering of A, that is $A = \bigcup_{i \in I} U_i$. Then $(X \setminus A, U_i)_{i \in I}$ is an open covering of X. Since X is a compact, it can be covered by some finite subcovering, say

$$X = (X \setminus A) \cup U_{i_1} \cup \cdots \cup U_{i_k}.$$

In particular this implies A can be covered by some finite subcovering of (U_i), namely

$$A = U_{i_1} \cup \cdots \cup U_{i_k}.$$

Thus A is a compact subset of X. Moreover, observe that A is a G_δ-set since

$$\begin{aligned} A &= \{a \leq f \leq b\} \\ &= \bigcap_{n=1}^{\infty} \{a + 1/n < f < b + 1/n\}. \end{aligned}$$

Since $f \in C(X)$ and $A \in \mathcal{C}$ were arbitrary, it follows that $\mathcal{M} \subseteq \sigma(\mathcal{D})$. □

The key takeaway from this proposition is that much of this proof involved purely topological arguments. For instance, Urysohn's lemma has nothing to do with measure theory or linear analysis; it's content is only concerned with topological concepts like continuous functions. Essentially it tells us that we have many continous functions $f \colon X \to \mathbb{R}$ to work with.

47.2 Riesz Representation Theorem

Before we state the form of Riesz representation theorem which we will be interested in proving, we consider the following proposition/definition.

Proposition 47.2. *Let μ be a Baire measure. Define $\ell_\mu \colon C(X) \to \mathbb{R}$ by*

$$\ell_\mu(f) = \int_X f \mathrm{d}\mu$$

for all $f \in C(X)$. The map ℓ_μ is a positive linear functional.

Proof. Positivity of ℓ_μ follows from positivity of integration, and linearity of ℓ_μ follows from linearity of integration. To see that ℓ_μ is bounded, note that

$$\begin{aligned} \ell_\mu(f) &= \int_X f \mathrm{d}\mu \\ &\leq \|f\|_\infty \mu(X) \end{aligned}$$

for all $f \in C(X)$. Taking f to be the constant function 1, we see that $\|\ell_\mu\| = \mu(X)$. □

The equality $\|\ell_\mu\| = \mu(X)$ obtained in the proof above seems to suggest that something more is going on than what was stated in the proposition. In fact, there is! Let us denote by $M(X)$ to be the space of signed Baire measures defined on the Baire σ-algebra of X. The sum of two finite signed Baire measures is a finite signed Baire measure, as is the product of a finite signed measure by a real number. Furthermore, the total variation defines a norm, and it turns out that this gives $M(X)$ the structure of a Banach space. The association $\mu \mapsto \ell_\mu$ can be extended to an isomorphism of Banach spaces from $M(X)$ to $C(X)^*$ which is natural in X. For sake of time, we will not pursue this direction too much in this document; however we still wanted to mention it. We are now ready to state the form of the Reisz representation theorem which we will be interested in.

Theorem 47.1. *Let ℓ be a positive linear functional defined on $C(X)$. Then there exists a unique Baire measure μ such that $\ell = \ell_\mu$.*

47.2.1 Extremally Disconnected Spaces

There way that we will prove Theorem (47.1) is by first proving it in the case where X is extremally disconnected. After we do this, we then proceed to the general case. First, let us recall what extremally disconnected means.

Definition 47.3. We say X is **extremally disconnected** if each open subset of X has open closure.

The condition that X be extremally disconnected turns out to be equivalent to the condition that every pair of disjoint open subsets of X has disjoint closures. Indeed, suppose that X is extremally disconnected and let U and U' be two disjoint open subsets of X. Since both $\overline{U}$ and $\overline{U}'$ are open, their intersection $\overline{U} \cap \overline{U}'$ is also open. If $\overline{U} \cap \overline{U}' \neq \emptyset$, then would have $U \cap U' \neq \emptyset$, which is a contradiction. Thus we must have $\overline{U} \cap \overline{U}' \neq \emptyset$. The converse direction is proved in a similar manner.

47.2.2 Proof of the Riesz Representation Theorem when X is Extremally Disconnected

Before we prove Theorem (47.1) in the special case where X is extremally disconnected, we consider the following lemma.

Lemma 47.2. *Suppose X is extremally disconnected compact Hausdorff space. Let $\mathcal{O}$ be the collection of all clopen subsets of X. Then $\mathcal{O}$ forms an algebra. Furthermore, let $\mathcal{V}$ be the space of $\mathcal{O}$-simple functions. Then $\mathcal{V}$ is a uniformly dense subspace of $C(X)$.*

Proof. It is straightforward to check that $\mathcal{O}$ is an algebra, so we will only focus on showing $\mathcal{V}$ is a uniformly dense subspace $C(X)$. First note if $A \in \mathcal{O}$, then the characteristic function $1_A \colon X \to \mathbb{R}$ is continuous since A is clopen. Indeed, suppose U is an open subset of $\mathbb{R}$. Then

$$\{1_A \in U\} = \begin{cases} A & \text{if } 1 \in U \text{ and } 0 \notin U \\ X \backslash A & \text{if } 1 \notin U \text{ and } 0 \in U \\ X & \text{if } 1 \in U \text{ and } 0 \in U \\ \emptyset & \text{if } 1 \notin U \text{ and } 0 \notin U \end{cases}$$

Thus $\mathcal{V}$ really is a subset of $C(X)$. It is straightforward to check that $\mathcal{V}$ is in fact a subspace of $C(X)$. Let us now show that it is uniformly dense in $C(X)$. Let $f \in C(X)$ and let $n \in \mathbb{N}$. For each $k \in \mathbb{Z}$, let

$$A_k = \left\{ \frac{k}{n} < f < \frac{k+1}{n} \right\}.$$

Each A_k is open and are pairwise disjoint from each other. Since X is extremally disconnected, each $\overline{A}_k$ is clopen and pairwise disjoint from each other. Furthermore, since X is compact, only finitely many of the $\overline{A}_k$ are nonempty. Thus the sets $A = \bigcup_{k \in \mathbb{Z}} \overline{A}_k$ and $X \backslash A$ are clopen. Note that $X \backslash A$ is the set of all $x \in X$ such that there exists $k \in \mathbb{Z}$ and an open neighborhood U of x such that $U \subseteq \{f = k/n\}$. For each $k \in \mathbb{Z}$, let

$$B_k = (X \backslash A) \cap \{f = k/n\}.$$

Then each B_k is clopen and pairwise disjoint from each other, and since (B_k) covers the compact space $X \backslash A$, it follows that only finitely many of the B_k are nonempty. Thus we can define the following $\mathcal{O}$-simple function

$$\varphi_n = \sum_{k \in \mathbb{Z}} \frac{k}{n} (1_{\overline{A}_k} + 1_{B_k})$$

which is easily seen to satisfy $|\varphi_n - f| < 1/n$. In other words, the sequence (φ_n) of $\mathcal{O}$-simple functions converges uniformly to f. □

Now we prove Theorem (47.1) in the special case where X is extremally disconnected.

Proposition 47.3. *With the notation as in Theorem (47.1), assume further that X is extremally disconnected. Then $\ell = \ell_\mu$.*

Proof. We first show existence. Define $\mu \colon \mathcal{O} \to [0, \infty)$ by

$$\mu(A) = \ell(1_A)$$

for all $A \in \mathcal{O}$. We claim that $(X, \mathcal{O}, \mu)$ is a finite premeasure space. Indeed, we have

$$\begin{aligned}\mu(\emptyset) &= \ell(1_\emptyset) \\ &= \ell(0) \\ &= 0.\end{aligned}$$

Furthermore, suppose that (A_n) is a disjoint sequence of sets in $\mathcal{O}$ whose union

$$A = \bigcup_{n=1}^{\infty} A_n$$

is also in $\mathcal{O}$. Then A is compact since it is a closed subset of a compact space X, and so it follows that only finitely many of the A_n can be nonempty, say $A_{n_1}, \dots, A_{n_k}$. Thus we have

$$\begin{aligned}\nu\left(\bigcup_{n=1}^{\infty} A_n\right) &= \nu\left(\bigcup_{i=1}^{k} A_{n_i}\right) \\ &= \ell\left(1_{\bigcup_{i=1}^{k} A_{n_i}}\right) \\ &= \ell\left(\sum_{i=1}^{k} 1_{A_{n_i}}\right) \\ &= \sum_{i=1}^{k} \ell(1_{A_{n_i}}) \\ &= \sum_{i=1}^{k} \nu(A_{n_i}) \\ &= \sum_{n=1}^{\infty} \nu(A_n).\end{aligned}$$

It follows that $(X, \mathcal{O}, \mu)$ is a finite premeasure space. Therefore by the Caratheodory Extension Theorem, the premeasure μ extends to a unique measure, which we again denote μ, defined on $\sigma(\mathcal{O})$.

We claim that $\sigma(\mathcal{O}) = \mathcal{M}$ so that μ is in fact a Baire measure. Indeed, we have $\sigma(\mathcal{O}) \subseteq \mathcal{M}$ since each clopen subset of X is a compact G_δ-set. To show $\mathcal{M} \subseteq \sigma(\mathcal{O})$, it suffices to show that each $f \in C(X)$ is a $\sigma(\mathcal{O})$-measurable function. Let $f \in C(X)$ and let $c \in \mathbb{R}$. For each $n \in \mathbb{N}$ set $A_n = \{f < c + 1/n\}$ and set $A = \{f \leq c\}$. Since A_n is open, its closure $\overline{A}_n$ is clopen, and since

$$A_n \subseteq \overline{A}_n \subseteq \{f \leq c + 1/n\},$$

we have $A = \bigcap_{n=1}^{\infty} \overline{A}_n$. Thus $A \in \sigma(\mathcal{O})$, which implies f is $\sigma(\mathcal{O})$-measurable. Therefore $\sigma(\mathcal{O}) = \mathcal{M}$ and hence μ is a Baire measure. From Proposition (47.2), we obtain a corresponding positive linear functional $\ell_\mu \in C(X)^*$.

Now we want to show that $\ell = \ell_\mu$ which will establish existence. First observe that $\ell|_{\mathcal{V}} = \ell_\mu|_{\mathcal{V}}$. Indeed, let φ be an $\mathcal{O}$-simple function and express it in canonical form as $\varphi = \sum_{i=1}^{n} a_i 1_{A_i}$. Then we have

$$\begin{aligned}\ell(\varphi) &= \ell\left(\sum_{i=1}^{n} a_i 1_{A_i}\right) \\ &= \sum_{i=1}^{n} a_i \ell(1_{A_i}) \\ &= \sum_{i=1}^{n} a_i \mu(A_i) \\ &= \int_X \varphi \mathrm{d}\mu. \\ &= \ell_\mu(\varphi).\end{aligned}$$

Nexxt, observe that since $\mathcal{V}$ is uniformly dense in X, we in fact have $\ell = \ell_\mu$ everywhere. Indeed, given $f \in C(X)$, we choose a sequence (φ_n) of $\mathcal{O}$-simple functions which converges uniformly to f. Then since $\mu(X) < \infty$, we have

$$\begin{aligned}
\ell(f) &= \lim_{n\to\infty} \ell(\varphi_n) \\
&= \lim_{n\to\infty} \ell_\mu(\varphi_n) \\
&= \lim_{n\to\infty} \int_X \varphi_n \mathrm{d}\mu \\
&= \int_X f \mathrm{d}\mu \\
&= \ell_\mu(f).
\end{aligned}$$

This establishes existence.

Now we prove uniqueness. Let μ and ν be two Baire measures such that $\ell_\mu = \ell_\nu$, that is, such that

$$\int_X f \mathrm{d}\mu = \int_X f \mathrm{d}\nu$$

for all $f \in C(X)$. We need to show that $\mu = \nu$. Let K be a compact G_δ-set. Express K as an countable intersection of open sets, say

$$K = \bigcap_{n=1}^{\infty} U_n.$$

For each $n \in \mathbb{N}$, there exists a continuous function $f_n \colon X \to [0,1]$ such that $f_n|_K = 1$ and $f_n|_{X\setminus U_n} = 0$ by Urysohn's lemma. Since both μ and ν are finite measures, the characteristic function 1_K is a nonnegative μ-integrable and ν-integrable function which dominates each f_n. Moreover the sequence (f_n) converges pointwise to 1_K. Thus by the Lebesgue dominated convergence theorem, we have

$$\begin{aligned}
\mu(K) &= \int_X 1_K \mathrm{d}\mu \\
&= \lim_{n\to\infty} \int_X f_n \mathrm{d}\mu \\
&= \lim_{n\to\infty} \int_X f_n \mathrm{d}\nu \\
&= \int_X 1_K \mathrm{d}\nu \\
&= \nu(K).
\end{aligned}$$

Since K is an arbitrary compact G_δ-set and $\mathcal{M}$ is generated by the collection of all of these types of sets, we have $\mu = \nu$. This establishes uniqueness. □

47.2.3 Proof of the Riesz Representation Theorem in General

In this last section, we will sketch the proof of Theorem (47.1) in the general case where X is not necessarily extremally disconnected. The idea is at follows: let X_0 be the set X equipped with the discrete topology and let $Y = \beta X_0$ be the Stone–Čech compactification with $\iota \colon X_0 \to Y$ denoting the canonical map. Any function out of a discrete space is continuous and so in particular the identity function $1 \colon X_0 \to X$ is continuous. By the universal properties of the Stone–Čech compactification, there exists a unique continuous map $\pi \colon Y \to X$ such that $\pi \circ \iota = 1$. Now the continuous map π induces an isomorphism $\pi_* \colon C(X) \to C(Y)$ of Banach spaces given by

$$\pi_*(f) = \pi \circ f$$

for all $f \in C(X)$. One check that this is in fact an isomorphism using the universal properties of the Stone–Čech compactification. The isomorphism π_* induces another isomorphism $(\pi_*)^* \colon C(Y)^* \to C(X)^*$ of Banach spaces, given by

$$(\pi_*)^*(\psi) = \psi \circ \pi_*$$

for all $\psi \in C(Y)^*$. In particular, there is a uniuqe $\psi \in C(Y)^*$ such that $\psi \circ \pi_* = \ell$. It turns out that the Stone–Čech compactification of a discrete space is an extremally disconnected space, and so one can apply Proposition (47.3) to get a unique Baire measure ν which represents ψ. Finally one shows that $\nu \circ \pi^{-1}$ is the unique Baire measure which represents ℓ.

Part V
Complex Analysis

Notation

Open and Closed Balls

Throughout these notes, I denotes the closed interval $[0,1]$ in $\mathbb{R}$. For all $r > 0$ and $z \in \mathbb{C}$, we write $B_r(z)$ (respectively $\overline{B}_r(z)$) to denote the open (respectively closed) ball centered at z and of radius r:

$$B_r(z) := \{w \in \mathbb{C} \mid |z-w| < r\} \quad \text{and} \quad \overline{B}_r(z) := \{w \in \mathbb{C} \mid |z-w| \leq r\}.$$

Similarly, for all $r > 0$ and $z \in \mathbb{C}$, we write $C_r(z)$ to denote the circle centered at z and of radius r:

$$C_r(z) := \{w \in \mathbb{C} \mid |z-w| = r\}.$$

Complex Numbers

Let z be a nonzero complex number. A **polar representation** of z is given by

$$z := |z|e^{2\pi i(\mathrm{Arg}(z)+n)}$$

for a unique $|z| \in \mathbb{R}_{\geq 0}$, a unique $\mathrm{Arg}(z) \in [-\pi, \pi)$, and a choice of $n \in \mathbb{Z}$. We call $|z|$ the **modulus** of z. We define the **argument** of z to be the set

$$\arg(z) := \{\mathrm{Arg}(z) + n \mid n \in \mathbb{Z}\}$$

and we call $\mathrm{Arg}(z)$ the **main branch** of the argument of z.

48 Convergence of Sequences of Functions and Power Series

Definition 48.1. Let D be a nonempty subset of $\mathbb{C}$ and let $(f_n \colon D \to \mathbb{C})$ be a sequence of functions. Then

1. The sequence (f_n) converges **pointwise** on D to a function f if for all $z \in D$ and for all $\varepsilon > 0$ there exists $N_{z,\varepsilon} \in \mathbb{N}$ (which depends on $z \in D$ and $\varepsilon > 0$) such that
$$n \geq N_{z,\varepsilon} \text{ implies } |f_n(z) - f(z)| < \varepsilon$$
2. The sequence (f_n) converges **uniformly** on D to a function f if for all $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ (which depends on $\varepsilon > 0$) such that
$$n \geq N_\varepsilon \text{ implies } |f_n(z) - f(z)| < \varepsilon$$
for all $z \in D$.
3. The sequence (f_n) is **uniformly Cauchy** on D if for all $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that
$$m, n \geq N_\varepsilon \text{ implies } |f_m(z) - f_n(z)| < \varepsilon$$
for all $z \in D$.
4. The series $\sum f_n$ converges **pointwise** on D to a function f if the sequence of partial sums $(\sum_{m=1}^n f_m)_{n \in \mathbb{N}}$ converges uniformly on D to f.
5. The series $\sum f_n$ converges **uniformly** on D to a function f if the sequence of partial sums $(\sum_{m=1}^n f_m)_{n \in \mathbb{N}}$ converges uniformly on D to f.

The main advantage in determining whether or not a sequence of functions (f_n) is uniformly Cauchy is that we do not need to know what (f_n) converges to. In contrast, the definition of (f_n) converging uniformly assumes that we already know what it converges to from the outset. Fortunately, since $\mathbb{C}$ is complete, we only need to know that (f_n) is uniformly Cauchy to determine whether it converges uniformly or not.

Theorem 48.1. *Let D be a nonempty subset of $\mathbb{C}$ and let $(f_n\colon D \to \mathbb{C})$ be a sequence of functions.*

1. *The sequence (f_n) converges uniformly on D to a function $f\colon D \to \mathbb{C}$ if and only if (f_n) is uniformly cauchy on D.*
2. *(Weierstrass M-test) Suppose that for each $n \in \mathbb{N}$ there exists $M_n \in [0, \infty)$ such that $\sum_{n=1}^{\infty} M_n < \infty$ and $|f_n(z)| \leq M_n$ for all $z \in D$, then the series $\sum_{n=1}^{\infty} f_n$ converges uniformly on D.*

Proof.

1. First we assume that (f_n) is uniformly cauchy on D. Let $\varepsilon > 0$ and choose $N \in \mathbb{N}$ such that

$$m, n \geq N \text{ implies } |f_m(z) - f_n(z)| < \varepsilon \tag{119}$$

for all $z \in D$. Then for each $z \in D$, the sequence $(f_n(z))$ is a Cauchy sequence in $\mathbb{C}$, and by completeness of $\mathbb{C}$, it must converge to a limit. Let $f(z)$ denote this limit. As we vary $z \in D$, we obtain a function $f\colon D \to \mathbb{C}$, given by

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

Clearly (f_n) converges pointwise to $f\colon D \to \mathbb{C}$. To see that it converges *uniformly* to f, we fix $m \in \mathbb{N}$ and let $n \to \infty$ in (119) and we see that

$$m \geq N \text{ implies } |f_m(z) - f(z)| \leq \varepsilon$$

for all $z \in D$.

Now we assume that (f_n) converges uniformly on D to a function $f\colon D \to \mathbb{C}$. Let $\varepsilon > 0$ and choose $N \in \mathbb{N}$ such that

$$n \geq N \text{ implies } |f_n(z) - f(z)| < \frac{\varepsilon}{2}$$

for all $z \in D$. Then for all $m, n \geq N$, we have

$$\begin{aligned} |f_m(z) - f_n(z)| &\leq |f_m(z) - f(z)| + |f_n(z) - f(z)| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon \end{aligned}$$

for all $z \in D$. Thus, (f_n) is uniformly cauchy.

2. By 1, it suffices to show that the sequence $(\sum_{m=1}^{n} f_m)_{n \in \mathbb{N}}$ of partial sums is uniformly Cauchy on D. Let $\varepsilon > 0$. Since the series $\sum_{n=1}^{\infty} M_n$ converges, the sequence $(\sum_{k=1}^{n} M_k)_{k \in \mathbb{N}}$ of partial sums is necessarily a Cauchy sequence. Therefore, there exists $N \in \mathbb{N}$ such that

$$m, n \geq N \text{ implies } \left| \sum_{k=1}^{m} M_k - \sum_{k=1}^{n} M_k \right| < \varepsilon.$$

In particular, $m, n \geq N$ implies

$$\begin{aligned} \left| \sum_{k=1}^{m} f_k(z) - \sum_{k=1}^{n} f_k(z) \right| &= \left| \sum_{k=m+1}^{n} f_k(z) \right| \\ &\leq \sum_{k=m+1}^{n} |f_k(z)| \\ &\leq \sum_{k=m+1}^{n} M_k \\ &= \left| \sum_{k=m+1}^{n} M_k \right| \\ &= \left| \sum_{k=1}^{m} M_k - \sum_{k=1}^{n} M_k \right| \\ &< \varepsilon. \end{aligned}$$

for all $z \in D$. □

48.1 Uniform Norm

Proposition 48.1. *Let D be a nonempty subset of $\mathbb{C}$ and let $(f_n\colon D \to \mathbb{C})$ be a sequence of continuous functions. If (f_n) converges to f uniformly on D, then f is continuous on D.*

Proof. Choose any $a \in D$. We will show that f is continuous at a. Let $\varepsilon > 0$. Since (f_n) converges to f uniformly, there exists an $N \in \mathbb{N}$ such that $|f_n(z) - f(z)| < \varepsilon/3$ for all $n \geq N$ and $z \in D$. Since f_N is continuous at a, there exists $\delta > 0$ such that $|z - a| < \delta$ implies $|f_N(z) - f_N(a)| < \varepsilon/3$. Combining these together, we see that $|z - a| < \delta$ implies

$$\begin{aligned} |f(z) - f(a)| &\leq |f(z) - f_N(z)| + |f_N(z) - f_N(a)| + |f_N(a) - f(a)| \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \\ &= \varepsilon. \end{aligned}$$

It follows that f is continuous at a. □

Examining the proof in Proposition (48.1) reveals that we can weaken the hypothesis under certain conditions. Let K be a compact subset of $\mathbb{C}$ and let $\mathcal{B}(K, \mathbb{C})$ be the $\mathbb{C}$-vector space set of all bounded functions from D to $\mathbb{C}$. We define the **uniform norm** on $\mathcal{B}(K, \mathbb{C})$ by

$$\|f\|_K = \sup\{|f(x)| \mid x \in K\}$$

for all $f \in \mathcal{B}(K, \mathbb{C})$. The pair $(\mathcal{B}(K, \mathbb{C}), \|\cdot\|_K)$ is easily checked to be a normed vector space. This normed vector space gives rise to a metric space in the usual way. Namely, we define the metric $\mathrm{d}_K\colon \mathcal{B}(K, \mathbb{C}) \times \mathcal{B}(K, \mathbb{C}) \to \mathbb{R}$ by

$$\mathrm{d}_K(f, g) = \|f - g\|_K$$

for all $f, g \in \mathcal{B}(K, \mathbb{C})$. A sequence $(f_n\colon K \to \mathbb{C})$ of bounded functions converges *uniformly* to a function f (which must necessarily be bounded) if and only if it converges to f with respect to metric d_K (check!). This is where the name *uniform* norm comes from.

Proposition 48.2. *Let K be a nonempty compact subset of $\mathbb{C}$ and let (f_n) be a sequence of continuous functions in $(\mathcal{B}(K, \mathbb{C}), \mathrm{d}_K)$. If f is a limit point of (f_n), then f is continuous on K.*

49 Power Series

A **power series centered at** a is a series of the form $\sum a_n(z - a)^n$, where z is a complex variable, a is a given complex number, and (a_n) is a sequence of complex numbers.

49.1 Limit Supremum

To study the convergence of a power series, we study the notion of the limit supremum of a positive real-valued sequence. Let (a_n) be a sequence of positive real numbers. We define the **limit supremum** of (a_n), denoted $\mathrm{limsup}(a_n)$, to be

$$\mathrm{limsup}(a_n) := \lim_{m \to \infty}(\sup\{a_n \mid n \geq m\}).$$

Since $\sup\{a_n \mid n \geq m\}$ is a non-increasing function of m, the limit always exists or equals $+\infty$.

Properties of Limit Supremum

Proposition 49.1. *Let (a_n) be a sequence of positive real-valued numbers.*

1. *If $\mathrm{limsup}(a_n) = A$, then for each $\varepsilon > 0$ and $N \in \mathbb{N}$, there exists some $n \geq N$ such that $a_n > A - \varepsilon$.*
2. *If $\mathrm{limsup}(a_n) = A$, then for each $\varepsilon > 0$, there exists $N \in N$ such that $a_n < A + \varepsilon$ for all $n \geq N$.*
3. *Conversely, if $A \in \mathbb{R}_{\geq 0}$ satisfies 1 and 2, then $\mathrm{limsup}(a_n) = A$.*

Proof.

1. Let $\varepsilon > 0$ and let $N \in \mathbb{N}$. To obtain a contradiction, assume that there does not exist an $n \geq N$ such that $a_n > A - \varepsilon$. Then $A - \varepsilon > a_n$ for all $n > N$. This implies $\sup\{a_n \mid n \geq N\} < A$. This is a contradiction since $\sup\{a_n \mid n \geq m\}$ is a non-increasing function of m.

2. Let $\varepsilon > 0$. To obtain a contradiction, assume that there does not exist an $N \in \mathbb{N}$ such that $a_n < A + \varepsilon$ for all $n \geq N$. Then $\sup\{a_n \mid n \geq N\} \geq A + \varepsilon$ for all $N \in \mathbb{N}$. This implies $\operatorname{limsup}(a_n) \geq A + \varepsilon$, which is a contradiction.

3. Let $A' = \operatorname{limsup}(a_n)$. Assume that $A < A'$. Let $\varepsilon = A' - A$. Then by 2, there exists $N \in \mathbb{N}$ such that $a_n < A + \varepsilon/2$ for all $n \geq N$. So we choose such an $N \in \mathbb{N}$. On the other hand, by 1, there must exist an $n \geq N$ such that $a_n > A' - \varepsilon/2 = A + \varepsilon/2$. Contradiction. An analagous argument gives a contradiction when we assume $A > A'$. Therefore $A = A'$.

□

Lemma 49.1. *Let (a_n) and (b_n) be two sequences of positive real numbers such that limsup$(a_n) = A$ and $\lim(b_n) = B$. Then*

1. *limsup$(a_n b_n) = AB$*

2. *limsup$(a_n + b_n) = A + B$*

Proof.

1. Let $\nu > 0$ and let $\delta > 0$ such that $\delta A + \delta B + \delta^2 < \nu$. Choose $N \in \mathbb{N}$ such that $a_n < A + \delta$ and $b_n < B + \delta$ for all $n \geq N$. Then for all $n \geq N$, we have

$$\begin{aligned} a_n b_n &< (A+\delta)(B+\delta) \\ &= AB + \delta A + \varepsilon B + \delta^2 \\ &< AB + \nu. \end{aligned}$$

Next, let $\varepsilon > 0$, let $N \in \mathbb{N}$, and set $\varepsilon' = \varepsilon/(A+B)$. Choose $n \geq N$ such that $a_n > A - \varepsilon'$ and $b_n > B - \varepsilon'$. Then

$$\begin{aligned} a_n b_n &> (A-\varepsilon')(B-\varepsilon') \\ &= AB - \varepsilon' A - \varepsilon' B + \varepsilon'^2 \\ &> AB - \varepsilon' A - \varepsilon' B \\ &= AB - \varepsilon'(A+B) \\ &= AB - \varepsilon. \end{aligned}$$

Therefore, we must have

$$\operatorname{limsup}(a_n b_n) = AB.$$

2. Let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that $a_n < A + \varepsilon/2$ and $b_n < B + \varepsilon/2$ for all $n \geq N$. Then for all $n \geq N$, we have

$$\begin{aligned} a_n + b_n &< A + \varepsilon/2 + B + \varepsilon/2 \\ &= A + B + \varepsilon. \end{aligned}$$

Next, let $\varepsilon > 0$ and let $N \in \mathbb{N}$. Choose $n \geq N$ such that $a_n > A - \varepsilon/2$ and $b_n > B - \varepsilon/2$. Then

$$\begin{aligned} a_n + b_n &> A - \varepsilon/2 + B - \varepsilon/2 \\ &= A + B - \varepsilon \end{aligned}$$

Therefore, we must have

$$\operatorname{limsup}(a_n + b_n) = A + B.$$

□

Example 49.1. Let (a_n) be a sequence of complex numbers and let $a \in \mathbb{C}$. Suppose that $\operatorname{limsup}(|a_n|^{1/n}) = A$. Then since $\lim(n^{1/n}) = 1$, we have $\operatorname{limsup}(|na_n|^{1/n}) = A$.

Limit Supremum Test of Convergence of Power Series

Theorem 49.2. *Let (a_n) be a sequence of complex numbers and let $a \in \mathbb{C}$. Suppose that* $\limsup(|a_n|^{1/n}) = L$.

1. *If $L = 0$, then the power series $\sum a_n(z-a)^n$ centered at a converges for all $z \in \mathbb{C}$.*
2. *If $L = \infty$, then the the power series $\sum a_n(z-a)^n$ centered at a converges for $z = 0$ only.*
3. *If $0 < L < \infty$, set $R = 1/L$. For any r with $0 < r < R$ the series $\sum a_n(z-a)^n$ converges absolutely and uniformly on the closed disk $\overline{B}_r(a)$ and diverges for $z \notin \overline{B}_R(a)$. In this case, R is called the* ***radius of convergence*** *of the power series.*

Proof. We only prove 3, leaving 1 and 2 as easy exercises. Choose r such that $0 < r < R$. Let $\varepsilon = (R-r)/2rR$ (so $r = 1/(L+2\varepsilon)$). Choose $N \in \mathbb{N}$ such that $|a_n|^{1/n} < L + \varepsilon$ for all $n \geq N$. Then

$$|a_n|^{1/n}|z-a| < \frac{L+\varepsilon}{L+2\varepsilon}$$

for all $z \in \overline{B}_r(a)$. Therefore, letting $M = (L+\varepsilon)/(L+2\varepsilon)$, we see that

$$\begin{aligned}\sum_{n=1}^{\infty} |a_n(z-a)^n| &= \sum_{n=1}^{N} |a_n(z-a)^n| + \sum_{n=N+1}^{\infty} |a_n(z-a)^n| \\ &\leq \sum_{n=1}^{N} |a_n(z-a)^n| + \sum_{n=N+1}^{\infty} M^n \\ &\leq \sum_{n=1}^{N} |a_n(z-a)^n| + \frac{1}{1-M}.\end{aligned}$$

for all $z \in \overline{B}_r(a)$. Thus, the series converges absolutely in $\overline{B}_r(a)$. The series also converges uniformly in $\overline{B}_r(a)$, by Weierstrass M-test, with $M_n = M^n$.

On the other hand, if $z \notin \overline{B}_R(a)$, then

$$\limsup\left(|a_n|^{1/n}|z-a|\right) > 1,$$

so that for infinitely many values of n, $a_n(z-a)^n$ has absolute value greater than 1 and thus $\sum a_n(z-a)^n$ diverges. ☐

Power Series Examples

Example 49.2.

1. The power series $\sum_{n=1}^{\infty} nz^n$ centered at 0 has radius of convergence 1 since $\limsup(n^{1/n}) = 1$.
2. The power series $\sum_{n=0}^{\infty} z^{n^2}$ centered at 0 has radius of convergence 1 since $\limsup(a_n^{1/n}) = 1$, where

$$a_n = \begin{cases} 1 & \text{if } n \text{ is a square} \\ 0 & \text{otherwise} \end{cases}$$

3. The generating function for the Catalan numbers C_n is given by

$$f(z) = (z^2+z)^2+z)^2+z)^2+\cdots) = z + 2z^2 + 5z^3 + 14z^4 + \cdots.$$

Since $\limsup(C_n^{1/n}) = 4$, we see that $\sum C_n z^n$ has radius of convergence $1/4$.

Properties of Sums

Lemma 49.3. *Let $\sum a_n(z-a)^n$ be a power series centered at a and suppose R is its radius of convergence. Then for all r such that $0 < r < R$, we have the estimate*

$$|a_k| \leq r^{-k}\|f(z)\|_{C_r(a)}.$$

for all $k \geq 0$.

Proof. The partial sum f_n of the power series is a polynomial of degree at most n, and hence the coefficient formula tells us that

$$r^k|a_k| \leq \frac{1}{n+1}\sum_{m=0}^{n}|f_n(r\omega^m)| \leq \sup_{|z-a|=r}|f_n(z)|.$$

Now the assertion follows since f_n converges uniformly to f on the disk $|z| \leq r$. □

49.2 Functions Representable by a Power Series

A function f defined on an open set Ω is said to be **representable by a power series in** Ω if, whenever $a \in \Omega$ and $r > 0$ and the disk $B_r(a)$ is included in Ω, there exists a sequence (a_n) of complex numbers such that the equation

$$f(z) = \sum_{n=0}^{\infty} a_n(z-a)^n$$

holds for every $z \in B_r(a)$.

Proposition 49.2. *Let Ω be an open set, let $f\colon \Omega \to \mathbb{C}$ be a function, and let $a \in \Omega$ and $r > 0$ such that $B_r(a) \subset \Omega$. Suppose that*

$$f(z) = \sum_{n=0}^{\infty} a_n(z-a)^n$$

for all $z \in B_r(a)$. Then $f'(z)$ exists for all $z \in B_r(a)$ and

$$f'(z) = \sum_{n=1}^{\infty} na_n(z-a)^{n-1}$$

Proof. Let $z \in B_r(a)$. Choose $\varepsilon > 0$ such that $B_\varepsilon(z) \subset B_r(a)$. Then for all $h \in B_\varepsilon(0)$, we have

$$\begin{aligned}
\lim_{h\to 0}\frac{f(z+h)-f(z)}{h} &= \lim_{h\to 0}\frac{1}{h}\sum_{n=0}^{\infty} a_n\left((z+h-a)^n-(z-a)^n\right)\\
&= \lim_{h\to 0}\lim_{n\to\infty}\frac{1}{h}\sum_{m=1}^{n} a_m\left((z+h-a)^m-(z-a)^m\right)\\
&= \lim_{h\to 0}\lim_{n\to\infty}\sum_{m=1}^{n} a_m\sum_{k=1}^{m}(z-a+h)^{m-k}(z-a)^{k-1}\\
&= \lim_{n\to\infty}\lim_{h\to 0}\sum_{m=1}^{n} a_m\sum_{k=1}^{m}(z-a+h)^{m-k}(z-a)^{k-1}\\
&= \lim_{n\to\infty}\sum_{m=1}^{n} ma_m(z-a)^{m-1}\\
&= \sum_{n=1}^{\infty} na_n(z-a)^{n-1}.
\end{aligned}$$

We need to justify why we were allowed to swap limits. Let $g_m\colon B_\varepsilon(0) \to \mathbb{C}$ be given by

$$g_m(h) = a_m\sum_{k=1}^{m}(z-a+h)^{m-k}(z-a)^{k-1}.$$

We need to show that the series $\sum g_m$ converges uniformly. In fact, this follows from an easy application of Weierstrass M-test. We first observe that

$$\begin{aligned}
|g_m(h)| &= \left|a_m\sum_{k=1}^{m}(z-a+h)^{m-k}(z-a)^{k-1}\right|\\
&< \left|ma_m r^{m-1}\right|.
\end{aligned}$$

Now we just set $M_m = |ma_m r^{m-1}|$ and apply Weierstrass M-test. □

Corollary 17. *Let Ω be an open set, let $f\colon \Omega \to \mathbb{C}$ be a function, let $a \in \Omega$, and let $r > 0$ such that $B_r(a) \subset \Omega$. Suppose that*

$$f(z) = \sum_{n=0}^{\infty} a_n (z-a)^n$$

for all $z \in B_r(a)$. Then $f^{(m)}(z)$ exists for all $m \geq 1$ and $z \in B_r(a)$, and moreover

$$f^{(m)}(z) = \sum_{n=0}^{\infty} \frac{(n+m)!}{n!} a_{n+m} (z-a)^n. \tag{120}$$

In particular, we have $a_m = f^{(m)}(a)/m!$ for all $m \geq 0$.

Proof. The first part follows from an easy induction on m, with Proposition (50.1 giving the base case and the induction step. To get $a_m = f^{(m)}(a)/m!$ for all $m \geq 0$, we set $z = a$ in 120). $\square$

50 Analytic Functions

50.1 Definition of an Analytic Function

Definition 50.1. A function $f\colon D \to \widehat{\mathbb{C}}$ is said to be **analytic at the point** z_0 in D if there exists a nonempty disk $B_r(z_0)$ centered at z_0 such that the restriction of f to $B_r(z_0)$ is the sum of a convergent power series with center z_0, that is,

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n \tag{121}$$

for all $z \in B_r(z_0)$.

50.1.1 Uniqueness of Representation

In principle, an analytic function could have different reprsentations (121) as power series at z_0. In order to prove that this cannot happen, we investigate to which extent the coefficients of a power series are determined by the values of its sums.

Theorem 50.1. *(Uniqueness Principle, Local Identity Theorem) Let f and g be the sums of two power series with center z_0, and assume that both converge in an open disk $B_r(z_0)$, say*

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n \quad \text{and} \quad g(z) = \sum_{n=0}^{\infty} b_n (z - z_0)^n. \tag{122}$$

If there exists a sequence $(z_m) \subset B_r(z_0)\backslash\{z_0\}$ such that $z_m \to z_0$ as $m \to \infty$ and $f(z_m) = g(z_m)$ for all $m \in \mathbb{N}$, then $a_n = b_n$ for all $n \in \mathbb{N}$ and $f(z) = g(z)$ for all $z \in B_r(z_0)$.

Proof. The functions f and g are continuous at z_0, and hence

$$\begin{aligned} a_0 &= f(z_0) \\ &= \lim_{m\to\infty} f(z_m) \\ &= \lim_{m\to\infty} g(z_m) \\ &= g(z_0) \\ &= b_0. \end{aligned}$$

Using the arithmetic rules for convergent sequences, we obtain the representations

$$f_1(z) := \frac{f(z) - a_0}{z - z_0} = \sum_{n=0}^{\infty} a_{n+1} (z - z_0)^n \quad \text{and} \quad g_1(z) := \frac{g(z) - b_0}{z - z_0} = \sum_{n=0}^{\infty} b_{n+1} (z - z_0)^n$$

for all $z \in B_r(z_0)\backslash\{z_0\}$. Because of $a_0 = b_0$ we have $f_1(z_m) = g_1(z_m)$ for all $m \in \mathbb{N}$, which implies $a_1 = b_1$, as just been shown. Proceeding inductively, we get $a_n = b_n$ for all n, and finally $f(z) = g(z)$ for all $z \in B_r(z_0)$. $\square$

50.1.2 Taylor Coefficients

The coefficients a_n of the power series (121) representing a function f analytic at z_0 are referred to as the **Taylor coefficients of** f **at** z_0. The series (121) itself is said to be the **Taylor series of** f **at** z_0. So we can say that a function f is analytic at z_0 if it admits a convergent Taylor series at z_0.

Proposition 50.1. *Let Ω be an open set and let $f\colon \Omega \to \mathbb{C}$ be analytic at a point a in Ω. Then f is holomorphic at a. Moreover, if*

$$f(z) = \sum_{n=0}^{\infty} a_n (z-a)^n$$

for all $z \in B_r(a)$, then f is holomorphic on $B_r(a)$, and we have $f'(z)$

$$f'(z) = \sum_{n=0}^{\infty} (n+1) a_{n+1} (z-a)^n$$

for all $z \in B_r(a)$. In particular, f' is analytic at a.

Proof. Let $z \in B_r(a)$. Choose $\varepsilon > 0$ such that $B_\varepsilon(z) \subset B_r(a)$. Then for all $h \in B_\varepsilon(0)$, we have

$$\begin{aligned}
\lim_{h\to 0} \frac{f(z+h) - f(z)}{h} &= \lim_{h\to 0} \frac{1}{h} \sum_{n=0}^{\infty} a_n \left((z+h-a)^n - (z-a)^n\right) \\
&= \lim_{h\to 0} \lim_{n\to\infty} \frac{1}{h} \sum_{m=1}^{n} a_m \left((z+h-a)^m - (z-a)^m\right) \\
&= \lim_{h\to 0} \lim_{n\to\infty} \sum_{m=1}^{n} a_m \sum_{k=1}^{m} (z-a+h)^{m-k} (z-a)^{k-1} \\
&= \lim_{n\to\infty} \lim_{h\to 0} \sum_{m=1}^{n} a_m \sum_{k=1}^{m} (z-a+h)^{m-k} (z-a)^{k-1} \\
&= \lim_{n\to\infty} \sum_{m=1}^{n} m a_m (z-a)^{m-1} \\
&= \sum_{n=1}^{\infty} n a_n (z-a)^{n-1}.
\end{aligned}$$

We need to justify why we were allowed to swap limits. Let $g_m\colon B_\varepsilon(0) \to \mathbb{C}$ be given by

$$g_m(h) = a_m \sum_{k=1}^{m} (z-a+h)^{m-k} (z-a)^{k-1}.$$

We need to show that the series $\sum g_m$ converges uniformly. In fact, this follows from an easy application of Weierstrass M-test. We first observe that

$$\begin{aligned}
|g_m(h)| &= \left| a_m \sum_{k=1}^{m} (z-a+h)^{m-k} (z-a)^{k-1} \right| \\
&< \left| m a_m r^{m-1} \right|.
\end{aligned}$$

Now we just set $M_m = |m a_m r^{m-1}|$ and apply Weierstrass M-test. □

Corollary 18. *Let Ω be an open set, let $f\colon \Omega \to \mathbb{C}$ be a function, let $a \in \Omega$, and let $r > 0$ such that $B_r(a) \subset \Omega$. Suppose that*

$$f(z) = \sum_{n=0}^{\infty} a_n (z-a)^n$$

for all $z \in B_r(a)$. Then $f^{(m)}(z)$ exists for all $m \geq 1$ and $z \in B_r(a)$, and moreover

$$f^{(m)}(z) = \sum_{n=0}^{\infty} \frac{(n+m)!}{n!} a_{n+m} (z-a)^n. \tag{123}$$

In particular, we have $a_m = f^{(m)}(a)/m!$ for all $m \geq 0$.

Proof. The first part follows from an easy induction on m, with Proposition (50.1 giving the base case and the induction step. To get $a_m = f^{(m)}(a)/m!$ for all $m \geq 0$, we set $z = a$ in 120). □

50.2 Operations Involving Analytic Functions

In this subsection, our goal is to prove the following theorem.

Theorem 50.2. *If f and g are analytic at z_0, then $f + g$, $f - g$, and fg are analytic at z_0. If, moreover, $g(z_0) \neq 0$, then f/g is analytic at z_0. If f is analytic at z_0 and g is analytic at $w_0 := f(z_0)$, then $g \circ f$ is analytic at z_0.*

Note that the composition $g \circ f$ need not exist on the domain of f, but just in a sufficiently small neighborhood of z_0. The analyticity of $f + g$ and $f - g$ are trivial. The proofs of the remaining assertions are more demanding.

50.2.1 Cauchy Product

The next result is a more sophisticated statement about the analyticity of a product fg, which includes an algorithm for computing the Taylor coefficients of fg from the coefficients of the factors f and g.

Theorem 50.3. *(Cauchy Product) Assume that the power series (122) for f and g converge in an open disk $B_R(z_0)$ centered at z_0 and of radius R, and let*

$$c_n := \sum_{m=0}^{n} a_m b_{n-m}.$$

Then the power series $\sum c_n(z - z_0)^n$ converges in $B_R(z_0)$ to the product $f(z)g(z)$, that is

$$f(z)g(z) = \sum_{n=0}^{\infty} c_n(z - z_0)^n \tag{124}$$

for all $z \in B_R(z_0)$.

Proof. Let f_n, g_n, and p_n be the partial sums of the series in (122) and (124) respetively. Then a rearrangement of the finite sums yields

$$\begin{aligned}
f_n(z)g_n(z) &= \sum_{m=0}^{n} a_m(z - z_0)^m \sum_{m=0}^{n} b_m(z - z_0)^m \\
&= \sum_{m=0}^{n} \left(\sum_{k=0}^{m} a_k b_{m-k} \right) (z - z_0)^m + \sum_{m=n+1}^{2n} \left(\sum_{k=m-n}^{n} a_k b_{m-k} \right) (z - z_0)^m \\
&= p_n(z) + \sum_{m=n+1}^{2n} \sum_{k=m-n}^{n} a_k b_{m-k}(z - z_0)^m.
\end{aligned}$$

Fix $z \in B_R(z_0)$. Choose r such that $|z - z_0| < r < R$ and choose a constant c such that $|a_k| \leq cr^{-k}$ and $|b_k| \leq cr^{-k}$ for all k. Setting $q := |z - z_0|/r < 1$, we use the triangle inequality to estimate

$$\begin{aligned}
|f_n(z)g_n(z) - p_n(z)| &\leq \sum_{m=n+1}^{2n} \sum_{k=m-n}^{n} |a_k||b_{m-k}||z - z_0|^m \\
&\leq \sum_{m=n+1}^{2n} \sum_{k=m-n}^{n} c^2 r^{-k}||z - z_0|^k \\
&\leq \sum_{m=n+1}^{2n} (2n - m + 1)c^2 q^k \\
&\leq n^2 c^2 q^{n+1}.
\end{aligned}$$

Since the right-hand side tends to zero as $n \to \infty$, the assertion follows. □

Example 50.1. Let $x, y \in \mathbb{C}$. Computing the Cauchy product of the Taylor series of $\exp x$ and $\exp y$, we obtain the **addition theorem** of the exponential function

$$\begin{aligned} e^x e^y &= \left(\sum_{k=0}^{\infty} \frac{x^k}{k!}\right)\left(\sum_{j=0}^{\infty} \frac{y^j}{j!}\right) \\ &= \sum_{k=0}^{\infty}\left(\sum_{j=0}^{k} \frac{x^j y^{k-j}}{j!(k-j)!}\right) \\ &= \sum_{k=0}^{\infty}\sum_{j=0}^{k} \frac{1}{k!}\binom{k}{j} x^j y^{k-j} \\ &= \sum_{k=0}^{\infty} \frac{(x+y)^k}{k!} \\ &= e^{x+y}. \end{aligned}$$

When this identity is applied to $z = x + iy$ with $x, y \in \mathbb{R}$, it yields a representation of the complex exponential function by familiar real functions,

$$e^{x+iy} = e^x(\cos y + i \sin y),$$

which implies that for all $z \in \mathbb{C}$,

$$|e^z| = e^{\mathrm{Re}(z)}, \quad \arg(e^z) = \mathrm{Im}(z), \quad e^{z+2\pi i} = e^z.$$

In particular, the exponential function has no zeros and is **periodic** with purely imaginary period $2\pi i$.

50.2.2 Reciprocal Functions

Theorem 50.4. *If f is analytic at z_0 and $f(z_0) \neq 0$, then $1/f$ is analytic at z_0. The Taylor coefficients b_k of $1/f$ at z_0 can be computed recursively from the Taylor coefficients a_k of f by $b_0 := 1/a_0$ and*

$$b_k := -\frac{1}{a_0}(a_1 b_{k-1} + a_2 b_{k-2} + \cdots + a_k b_0) \tag{125}$$

for all $k \in \mathbb{N}$.

Proof. In the first step we *assume* that the function $1/f$ is analytic at z_0. Then the Taylor series

$$\frac{1}{f(z)} = \sum_{n=0}^{\infty} b_0 (z - z_0)^n \tag{126}$$

converges in a neighborhood of z_0 and its Cauchy product with the Taylor series of f is the constant function 1. The latter is equivalent to the infinite system of equations

$$\begin{aligned} a_0 b_0 &= 1 \\ a_0 b_1 + a_1 b_0 &= 0 \\ a_0 b_2 + a_1 b_1 + a_2 b_0 &= 0 \\ &\vdots \end{aligned}$$

Since $a_0 \neq 0$, this triangular system can be solved with respect to the coefficients b_k, which yields the recursion (125).

It remains to prove that the series (126), with coefficients b_k given by the recursion (125), indeed has a positive radius of convergence. Choose positive numbers c and r such that $|a_n| \leq c r^{-n}$ for all $n \in \mathbb{N}$. We set $q := 1 + c/|a_0|$ and show that

$$|b_n| \leq \frac{c}{|a_0|^2} \frac{q^{n-1}}{r^n} \tag{127}$$

for all $n \in \mathbb{N}$. For $n = 1$, we have $b_1 = -a_1/a_0^2$ and $|a_1| \leq c/r$, so that indeed

$$\begin{aligned} |b_1| &= \frac{a_1}{a_0^2} \\ &\leq \frac{c}{|a_0|^2} \frac{1}{r}. \end{aligned}$$

Now assume that (127), holds for all $n = 1, 2, \ldots, k-1$ and consider the case where $n = k$. Using $|b_0| = 1/|a_0|$, the recursive definition of b_k, and the triangle inequality, we estimate

$$
\begin{aligned}
|b_k| &\leq \frac{1}{|a_0|}\left(|a_k b_0| + \sum_{j=1}^{k-1} |a_{k-j}||b_j|\right) \\
&\leq \frac{1}{|a_0|}\left(|a_k b_0| + \sum_{j=1}^{k-1} \frac{c}{r^{k-j}} \frac{c}{r^j |a_0|^2} q^{j-1}\right) \\
&\leq \frac{c}{r^k |a_0|^2}\left(1 + \frac{c}{|a_0|} \sum_{j=0}^{k-2} q^j\right) \\
&\leq \frac{c}{r^k |a_0|^2}\left(1 + \frac{c}{|a_0|} \frac{q^{k-1}-1}{q-1}\right) \\
&= \frac{c}{r^k |a_0|^2} q^{k-1},
\end{aligned}
$$

which gives (127) for $n = k$ and thus for all n. Consequently, the power series (126) has radius of convergence not less than r/q. □

Example 50.2. Let the function f be defined on the complex plane by $f(z) := (e^z - 1)/z$ if $z \neq 0$ and $f(0) = 1$. Representing e^z by its Taylor series, we obtain the power series

$$
f(z) = \sum_{k=0}^{\infty} a_k z^k = \sum_{k=0}^{\infty} \frac{z^k}{(k+1)!}
$$

which converges in the entire complex plane and attans the correct value $f(0) = 1$ at $z = 0$. SInce $f(0) \neq 0$, the reciprocal function $1/f$ is also analytic at $z_0 = 0$. Writing the Taylor series of $g := 1/f$ in the form

$$
g(z) = \sum_{k=0}^{\infty} b_k z^k = \sum_{k=0}^{\infty} \frac{B_k}{k!} z^k, \tag{128}
$$

the numbers B_k are determined by the equations $B_0 = b_0 = 1/a_0 = 1$ and

$$
\begin{aligned}
0 &= \sum_{j=0}^{k} a_{k-j} b_j \\
&= \sum_{j=0}^{k} \frac{B_j}{(k-j+1)!j!} \\
&= \frac{1}{(k+1)!} \sum_{j=0}^{k} \binom{k+1}{j} B_j,
\end{aligned}
$$

for $j \in \mathbb{N}$. Solving this system recursively, we get

$$
B_k = -\frac{1}{k+1} \sum_{j=0}^{k} \binom{k+1}{j} B_j
$$

for $j \in \mathbb{N}$. The numbers B_k are called **Bernoulli numbers**. For n odd, all B_n are zero, except B_1 which equals $-1/2$. The first Bernoulli numbers for n ever are

$$
B_2 = \frac{1}{6}, \quad B_4 = -\frac{1}{30}, \quad B_6 = \frac{1}{42}, \quad B_8 = -\frac{1}{30}, \quad B_{10} = \frac{5}{66}, \quad B_{12} = -\frac{691}{2730}.
$$

Note that the series (128) converges for $|z| < 2\pi$.

50.2.3 Composition of Power Series

The final step in proving Theorem (50.2) is concerned with the composition $g \circ f$ of functions given by power series. In order to ensure that the composition makes sense at least locally, we assume that f is analytic at z_0, while g is supposed to be analytic at the image point $w_0 := f(z_0)$. Then, by continuity, f maps a neighborhood of z_0 into the disk of convergence of g. Our goal is to find a convergent power series for $g \circ f$ from the given power series of f and g. The approach is straightforward: we assume that

$$f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k \quad \text{and} \quad g(w) = \sum_{k=0}^{\infty} b_k (w - w_0)^k. \tag{129}$$

substitute $w - w_0 = \sum_{n=1}^{\infty} a_n (z - z_0)^n$ in the series for g, rearrange the double sum according to the powers of $z - z_0$, and show that the resulting series converges to $g \circ f$ in a neighborhood of z_0. The details will be worked out next.

For $n \in \mathbb{N}$, the nth power $(f - a_0)^n$ is analytic at z_0 and the n leading terms of its Taylor series at z_0 vanish. Denoting by a_{nk} the Taylor coefficients of this function, we have

$$(f(z) - a_0)^n = \sum_{k=1}^{\infty} a_{nk} (z - z_0)^k = \sum_{k=n}^{\infty} a_{nk} (z - z_0)^k \tag{130}$$

in some neighborhood of z_0. Substituting the term $w - w_0$ in the power series of $g - b_0$ by the power series of $f - a_0$ (recall that $a_0 = f(z_0) = w_0$), we obtain formally

$$\begin{aligned} \sum_{n=0}^{\infty} b_n (w - w_0)^n &= \sum_{n=1}^{\infty} b_n \left(\sum_{k=n}^{\infty} a_{nk} (z - z_0)^k \right) \\ &= \sum_{k=1}^{\infty} \left(\sum_{n=1}^{k} b_n a_{nk} \right) (z - z_0)^k. \end{aligned}$$

Before we justify that changing the order of summation is possible, we state the result

Theorem 50.5. *If f is analytic at z_0 and g is analytic at $w_0 := f(z_0)$, then $g \circ f$ is analytic at z_0. Let f, g, and $(f - a_0)^n$ be represented by the series (129) and (130) respectively. Then the Taylor coefficients c_k of $g \circ f$ at z_0 are given by*

$$c_0 = b_0, \qquad c_k = \sum_{n=1}^{k} b_n a_{nk}$$

for all $k \in \mathbb{N}$.

50.3 Weierstrass Rearrangement Theorem

Theorem 50.6. *(Weierstrass Rearrangement Theorem) The sum of a power series is analytic at any point in its disk of convergence. If f is given by*

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n \tag{131}$$

for all $z \in B_r(z_0)$, and if $z_1 \in B_r(z_0)$, then

$$f(z) = \sum_{m=0}^{\infty} b_m (z - z_1)^m$$

for all $z \in B_{r_1}(z_1)$, where $r_1 := r - |z_1 - z_0|$ and the coefficients b_k are given by the convergent series

$$b_m = \sum_{n=m}^{\infty} \binom{n}{m} a_n (z_1 - z_0)^{n-m}$$

for all $k \in \mathbb{N}_0$.

Proof. Let $z \in B_{r_1}(z_1)$. Substituting $z - z_0 = (z - z_1) + (z_1 - z_0)$ into (131), we obtain

$$f(z) = \sum_{n=0}^{\infty} a_n((z - z_1) + (z_1 - z_0))^n$$
$$= \sum_{n=0}^{\infty} a_n \sum_{m=0}^{n} \binom{n}{m} (z - z_1)^m (z_1 - z_0)^{n-m}$$

In order to prove the assertion, it only remains to change the order of summation in the double series. It suffices to show that this series converges absolutely. To this end we remark that

$$\sum_{n=0}^{\infty} |a_n| \sum_{m=0}^{n} \binom{n}{m} |z - z_1|^m |z_1 - z_0|^{n-m} = \sum_{n=0}^{\infty} |a_n|(|z - z_1| + |z_1 - z_0|)^n.$$

The last sum converges because $|z - z_1| + |z_1 - z_0| < r$, so that the power series (131) converges absolutely at the point $z = z_0 + |z - z_1| + |z_1 - z_0|$. □

50.4 Definition of Analytic Function

Definition 50.2. A complex function $f\colon D \subseteq \mathbb{C} \to \mathbb{C}$ is said to be **analytic on** A if A is a subset of D and f is analytic at every point of A. We say that f is **analytic** if it is analytic on its domain set. A function which is analytic on the entire complex plane is called **entire**.

Lemma 50.7. *For any complex function $f\colon D \subseteq \mathbb{C} \to \mathbb{C}$ the set A_f of all points in D at which f is analytic is open.*

Proof. If A_f is empty, there is nothing to prove. If $z_0 \in A_f$, then f has a Taylor expansion at z_0 which converges in a open disk D_0 centered at z_0. By Theorem (50.6), $D_0 \subset A_f$. □

50.4.1 Jacobi Theta Function

An interesting family of entire functions are the **Jacobi Theta functions**, given by the series

$$\vartheta(z; q) := \sum_{n \in \mathbb{Z}} q^{n^2} e^{2\pi i n z}$$

for all $z \in \mathbb{C}$, where q is a complex parameter with modulus less than one. In order to show that ϑ is entire, we consider the power series

$$f(z) := \sum_{n=1}^{\infty} q^{n^2} z^n = qz + q^4 z^2 + q^9 z^3 + \cdots .$$

This series converges for all $z \in \mathbb{C}$ because

$$\operatorname{limsup}\left(\left|q^{n^2}\right|^{1/n}\right) = \operatorname{limsup}\left(|q^n|\right) = 0,$$

and thus the function f is entire. The function g defined by $g(z) := e^{2\pi i z}$ is also entire and has no zeros in $\mathbb{C}$, so that its reciprocal $1/g$ is also entire. Finally, $\vartheta(z) = 1 + 2f(g(z))$

$$\vartheta(z; q) = 1 + f(g(z)) + f(1/g(z))$$

for all $z \in \mathbb{C}$.

The function g, and consequently ϑ, is periodic with period 1. The parameter q is said to be the **nome** of the Theta function. It is often represented as $q = e^{\pi i \tau}$, where τ is a complex number with $\mathrm{Im}(\tau) > 0$.

50.4.2 Local Normal Forms

Theorem 50.8. *(Local Normal Form) Let $f\colon D \subseteq \mathbb{C} \to \mathbb{C}$ be analytic on D. If f is not constant in a neighborhood of $z_0 \in D$, then there exist a positive integer m and an analytic function $g\colon D \subset \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ with $g(z_0) \neq 0$ such that*

$$f(z) = f(z_0) + (z - z_0)^m g(z) \tag{132}$$

for all $z \in D$. The integer m and the function g are uniquely determined.

Proof. Assume that the Taylor series $f(z) = \sum a_k(z-z_0)^k$ of f at z_0 converges in a disk D_0. Denoting by a_m the first non-zero coefficient among $a_1, a_2, a_3, \ldots$, we have

$$f(z) = f(z_0) + (z-z_0)^m \sum_{k=m}^{\infty} a_k(z-z_0)^{k-m}$$

for all $z \in D_0$. The sum $g_0(z)$ of the series $\sum_{k=m}^{\infty} a_k(z-z_0)^{k-m}$ is an analytic function in D_0 with $g_0(z_0) = a_m \neq 0$. The function g defined in D by

$$g(z) := \begin{cases} \frac{f(z)-f(z_0)}{(z-z_0)^m} & \text{if } z \in D\backslash\{z_0\} \\ a_m & \text{if } z = z_0 \end{cases}$$

is analytic on $D\backslash\{z_0\}$. Since it coincides with g_0 in D_0 it is also analytic at z_0.

For proving uniqueness we assume that $(z-z_0)^n g_1(z) = (z-z_0)^m g_2(z)$ with $n > m$ for all $z \in D$. Then $(z-z_0)^{n-m} g_1(z) = g_2(z)$, and the left-hand side vanishes at z_0 while $g_2(z_0) \neq 0$. So $m = n$ and then $g_1 = g_2$ is obvious. □

Definition 50.3. The integer m in the representation (132) is called the **order** of the function f at z_0 and is denoted by $\operatorname{ord}(f, z_0)$. If f is constant in a neighborhood of z_0 we set $\operatorname{ord}(f, z_0) = \infty$. If in particular $f(z_0) = 0$, then m is said to be the **order of the zero** z_0.

As an immediate corollary of Theorem (50.8) we get the following result which shows, in particular, that all zeros of non-constant analytic functions are isolated.

Lemma 50.9. *If f is analytic at z_0 and $a := f(z_0)$, then there exists a disk D_0 with center z_0 such that either $f(z) = a$ for all $z \in D_0$ or $f(z) \neq a$ for all $z \in D_0\backslash\{z_0\}$.*

50.5 Analytic Functions in Planar Domain

As we have already seen, it is natural to require that the domain set D of an analytic function is open. From now on, we shall also assume that D is a nonempty connected open subset of $\mathbb{C}$, i.e. that D is a **domain**. This assumption is not too strong, since any open set in $\mathbb{C}$ is the disjoint union of domains, but it simplifies life a lot. In particular it is important when local statements about power series will be "lifted" to global results for analytic functions. This will be demonstrated in the proof of the following theorem.

Theorem 50.10. *(Identity Theorem, Uniqueness Principle) Let f and g be analytic functions in a domain D. If there exists a sequence $(z_n) \subset D\backslash\{z_0\}$ such that $z_n \to z_0 \in D$ and $f(z_n) = g(z_n)$ for all $n \in \mathbb{N}$, then $f(z) = g(z)$ for all $z \in D$.*

Proof. The function $h := f - g$ has a sequence of zeros which converge to $z_0 \in D$. Continuity of h implies that $h(z_0) = 0$, so that z_0 is a zero of h which is not isolated. Since h is analytic in D, we infer from Lemma (50.9) that $h(z) = 0$ in some disk D_0 with center z_0.

We pick any point z_1 in D and show that $h(z_1) = 0$. Since D is open and connected, it must be path-connected. So choose a path $\gamma\colon I \to D$ from z_0 to z_1. Then the set

$$S := \{s \in I \mid h(\gamma(t)) = 0 \text{ for all } t \in [0, s]\}$$

is not empty and we denote by s_0 its supremum. Continuity of h implies that $h(\gamma(s_0)) = 0$. Since $h(\gamma(t)) = 0$ for all $t \in [0, s_0]$, Lemma (50.9) tells us that $h(z) = 0$ in a neighborhood of $\gamma(s_0)$. This is only possible if $s_0 = 1$, because otherwise $h(\gamma(t)) = 0$ for all t in an interval $[0, s_1]$ with $s_1 > s_0$. □

50.5.1 Zeros of Analytic Function

The last theorem establishes the surprising fact that a function which is analytic in a domain is completely determined by its values in an arbitrarily small disk. We state another result concerning the zeros of such a function.

Corollary 19. *If $f \neq 0$ is analytic in a domain D and K is a compact subset of D, then the number of zeros of f in K is finite.*

Proof. If f had infinitely many zeros in K, there would exist a sequence (z_n) of such zeros which converge to a point $z_0 \in K \subset D$. But then $f = 0$ on D by Theorem (50.10). □

Nevertheless an analytic function $f \neq 0$ can have infinitely many zeros in D. If this happens, the zeros must have an accumulation point z_0 on $\widehat{\mathbb{C}}$. Since z_0 cannot lie in D, it must be on the boundary of D (considered as a subset of $\widehat{\mathbb{C}}$). For an entire function, the only possible accumulation point of zeros is the point at infinity.

Example 50.3. The function $\sin(1/z)$ is analytic in $\mathbb{C}\backslash\{0\}$ and has the zeros $z_k = 1/(k\pi)$ with $k = \pm 1, \pm 2, \ldots$, which accumulate at the origin.

50.5.2 Extremal Values

Theorem 50.11. *(Maximum and Minimum Principle) Let $f\colon D \subset \mathbb{C} \to \mathbb{C}$ be a non-constant analytic function. Then $|f|$ has no local maximum in D, and very local minimum of $|f|$ is a zero of f.*

Proof. Assume that $|f|$ attains a maximum or minimum at $z_0 \in D$. By Theorem (50.10) f is not locally constant, so that we can apply Theorem (50.8) and write

$$f(z) = f(z_0) + (z - z_0)^m g(z),$$

where g is analytic in D and $g(z_0) \neq 0$. □

50.6 Analytic Continuation

50.6.1 Direct Analytic Continuation

Theorem 50.12. *(Direct Analytic Continuation) Let the functions $f_1\colon D_1 \to \mathbb{C}$ and $f_2\colon D_2\colon \ \to \mathbb{C}$ be analytic in the domains D_1 and D_2, respectively. Assume that the intersection $D_0 := D_1 \cap D_2$ is nonempty and that $f_1 = f_2$ on D_0. Then there is a unique analytic function f on $D := D_1 \cup D_2$ which coincides with f_1 on D_1, namely*

$$f(z) := \begin{cases} f_1(z) & \text{if } z \in D_1 \\ f_2(z) & \text{if } z \in D_2. \end{cases}$$

Proof. The function f is analytic on D because any point $z \in D$ belongs to D_1 or D_2, so that f coincides with f_1 or f_2 in a neighborhood of z. Since $D_1 \cup D_2$ is a domain, and $D_1 \cap D_2 \neq \emptyset$ is open, uniqueness of f follows from the identity theorem. □

Under the assumptions of Theorem (50.12), the function f is said to be an **analytic continuation of f_1 onto** D. Interchanging the roles of f_1 and f_2, we see that f is also the (unique) analytic extension of f_2 onto D. So direct analytic continuation may extend a function to a larger domain, but this says nothing about how to *find* such an extension. The key to a constructive approach is Weierstrass rearrangement theorem for power series.

50.6.2 Analytic Function Elements

Assume that an analytic function f is given as the sum of a power series which has center z_0 and disk of convergence D_0. It can happen that the rearrangement of that power series to a series centered at a point z_1 in D_0 has a disk of convergence D_1 which protrudes out of D_0. Then by Theorem (50.12), f admits an analytic extension to $D_0 \cup D_1$. In order to explore this further we introduce some notation.

Definition 50.4.

1. An **(analytic) function element** if a pair (f, D) consisting of a disk D and an analytic function $f\colon D \to \mathbb{C}$. The center of the disk D is also referred to as the **center of the function element**.

2. If (f_1, D_1) and (f_2, D_2) are two function elements which satisfy the assumption of Theorem (50.12), we say that (f_2, D_2) is the **direct analytic continuation** of (f_1, D_1) (or vice versa) and write $(f_1, D_1) \bowtie (f_2, D_2)$.

3. A finite sequence $(f_0, D_0), (f_1, D_1), \ldots, (f_n, D_n)$ of funciton elements is said to be a **chain** if any function element (except the first) is the direct analytic continuation of its predecessor,

$$(f_0, D_0) \bowtie (f_1, D_1) \bowtie \cdots \bowtie (f_n, D_n). \tag{133}$$

   We then call (f_n, D_n) an **analytic continuation of** (f_0, D_0) **along the chain**.

4. A function element (f_n, D_n) is an **analytic continuation** of (f_0, D_0) if a chain of function elements satisfying (133) exists. We then write $(f_0, D_0) \sim (f_n, D_n)$.

To understand the procedures that follow better it is essential to recognize some subtleties of these definitions. While it is easy to see that $\sim$ is an **equivalence relation**, the relation $\bowtie$ is reflexive and symmetric, but *not transitive*.

Example 50.4. The binomial series

$$f_0(z) = \sum_{n=0}^{\infty} \binom{1/2}{n} (z-1)^n \tag{134}$$

has radius of convergence one and thus defines a function element (f_0, D_0) with $D_0 := B_1(1) = \{z \in \mathbb{C} \mid |z-1| < 1\}$. If z is real and $0 < z < 1$, we have $f_0(z) = \sqrt{z}$. For $k = 0, 1, \ldots, 8$ we denote by $\omega_k = e^{2\pi i k/9}$ the 9th roots of unity and let $D_k := \{z \in \mathbb{C} \mid |z - \omega_k| < 1\}$. All power series

$$f_k(z) := e^{ik\pi/18} \sum_{n=0}^{\infty} e^{-ik\pi/9} \binom{1/2}{n} (z - \omega_k)^n$$

have radius of convergence one and the nine function elements form a chain

$$(f_0, D_0) \bowtie (f_1, D_1) \bowtie \cdots \bowtie (f_8, D_8)$$

where neighbors are direct analytic continuations of each other. Consequently any two elements (f_j, D_j) and (f_k, D_k) are **analytic continuations** of each other. Moreover, for $k = 1, 2, 3, 4$, the element (f_k, D_k) is a *direct* analytic continuation of (f_0, D_0), but not for $k = 5, 6, 7, 8$. Since (f_6, D_6) is also a direct analytic continuation of (f_3, D_3), we have

$$(f_0, D_0) \bowtie (f_3, D_3) \bowtie (f_6, D_6) \not\bowtie (f_0, D_0),$$

which again shows that the relation $\bowtie$ is not transitive.

Lemma 50.13. *If $D_1 \cap D_2 \cap D_3 \neq \emptyset$, $(f_1, D_1) \bowtie (f_2, D_2)$ and $(f_2, D_2) \bowtie (f_3, D_3)$, then $(f_1, D_1) \bowtie (f_3, D_3)$.*

Proof. The functions f_1 and f_3 are analytic in the domain $D_1 \cap D_3$ and coincide (with f_2) on its open subset $D_1 \cap D_2 \cap D_3$. Thus $f_1 = f_3$ on $D_1 \cap D_3$. □

50.6.3 Analytic Continuation Along a Path

Definition 50.5. Let $\gamma\colon I \to \mathbb{C}$ be a path. A chain of function elements

$$(f_0, D_0) \bowtie (f_1, D_1) \bowtie \cdots \bowtie (f_n, D_n), \tag{135}$$

is said to be a **chain along** γ, if the chain of disks $(D_0, D_1, \ldots, D_n)$ covers γ in the sense of the Path Covering Lemma.

Let (f_0, D_0) and (f, D) be function elements with centers at $\gamma(0)$ and $\gamma(1)$, respectively. We say that (f, D) is an **analytic continuation of** (f_0, D_0) **along** γ, if there exists a chain of function elements (135) along γ such that $(f, D) = (f_n, D_n)$.

It is essential that analytic continuation along a path does not depend on the special choice of the chain of function elements. This statement is made precise in the next lemma.

Lemma 50.14. *Let $(f_0, D_0) \bowtie \cdots \bowtie (f_n, D_n)$ and $(g_0, \widetilde{D}_0) \bowtie \cdots \bowtie (g_m, \widetilde{D}_m)$ be two chains of function elements along a path γ. If $(f_0, D_0) \bowtie (g_0, \widetilde{D}_0)$, then it is also true that $(f_n, D_n) \bowtie (g_m, \widetilde{D}_m)$.*

Proof. Let $\gamma\colon I \to \mathbb{C}$ be a path and let

$$0 = t_0 < t_1 < \cdots < t_n = 1, \qquad 0 = s_0 < s_1 < \cdots < s_m = 1,$$

be partitions of I such that for all $k = 1, \ldots, n$ and $j = 1, \ldots, m$ we have

$$\gamma([t_{k-1}, t_k]) \subset D_k, \qquad \gamma([s_{j-1}, s_j] \subset \widetilde{D}_j.$$

Intuitively, the following procedure can be described as a walk along the path γ, where the left foot is only allowed to step on disks D_k, the right foot is restricted to the disks $\widetilde{D}_j$, and the function elements (f_k, D_k) and $(f_j, \widetilde{D}_j)$ underneath both feet must be direct analytic continuations of each other. We shall show that one can walk step-by-step all the way along γ, following just a simple rule: don't move the foot which is ahead. □

50.6.4 Function Elements and Germs

Though analytic continuation along a path γ is *essentially* independent of the choice of the function elements which cover γ, these elements are by no means uniquely defined. In fact not even the elements at the endpoints of γ are unique, Lemma (50.14) only tells us that the terminal elements of the chain *coincide on some disk* if the initial elements have this property. The redundancy in this process of analytic continuation is sometimes disturbing and makes formulations cumbersome. To eliminate this drawback we utilize the standard technique of forming classes.

Definition 50.6. Two function elements (f_1, D_1) and (f_2, D_2) centered at z_0 are said to be **equivalent** if $f_1(z) = f_2(z)$ in some neighborhood of z_0. A **germ** at z_0 is a class of equivalent function elements centered at z_0. The germ which contains a function element (f, D) is denoted by $f^\star$. We denote by $\mathcal{O}^{\mathrm{an}}_{z_0}$ to be the set of all germs at z_0. One easily checks that $\mathcal{O}^{\mathrm{an}}_{z_0}$ is a $\mathbb{C}$-algebra.

Depending on the situation, one can choose an appropriate **representative** of a germ $f^\star$. The **canonical representative** of a germ $f^\star$ is that function element (f, D) in $f^\star$ which has the disk D of maximal radius (here we allow $D = \mathbb{C}$).

The **value** $f^\star(z_0)$ **of a germ** $f^\star$ at z_0 is the value $f(z_0)$ of any function element (f, D) which represents $f^\star$. Note that the value of a germ is only defined at its center. On the other hand, the germ of a function element (f, D) is *not* determined by the value of f at its center z_0 alone, but by the complete list of its Taylor coefficients. To explain this idea more precisely, let $\mathbb{C}^\infty$ be the set of all sequences $(z_n)_{n\geq 0}$ in $\mathbb{C}$. Then $\mathbb{C}^\infty$ forms a $\mathbb{C}$-algebra, where addition is defined pointwise and where multiplication is defined by the Cauchy product; namely if (a_n) and (b_n) are two sequences in $\mathbb{C}$, then

$$(a_n) + (b_n) = (a_n + b_n) \quad \text{and} \quad (a_n)(b_n) = (c_n),$$

where $c_n = \sum_{m=0}^n a_m b_{n-m}$. Finally, let $\varphi\colon \mathcal{O}^{\mathrm{an}}_{z_0} \to \mathbb{C}^\infty$ be the morphism of $\mathbb{C}$-algebras given by sending a function element (f, D) to the Taylor sequence $(f^{(n)}(z_0))$. The identity theorem implies that this morphism is well-defined and injective.

The concept of germs is not restricted to function elements. If the function f is analytic at a point z, it is analytic in a neighborhood of z, and thus it induces a germ at z which we denote by $f^\star_z$.

50.6.5 Analytic Continuation of Germs

Definition 50.7. We say that a germ $f^\star$ at b is an analytic continuation of a germ $f^\star_a$ at a along a path γ from a to b if there exists a chain of function elements

$$(f_0, D_0) \bowtie \cdots \bowtie (f_n, D_n)$$

along γ such that (f_0, D_0) represents $f^\star_a$ and (f_n, D_n) represents $f^\star$, respectively.

Whenever an analytic continuation of a germ along a path γ exists, Lemma (50.14) tells us that the terminal germ is uniquely determined and does not depend on the specific choice of the function element along γ. We thus can speak of *the* analytic continuation $f^\star(\gamma)$ of a germ $f^\star$ along a path γ.

50.6.6 The Monodromy Principle

In the next step we study analytic continuation of a germ along *different paths* with the same endpoints.

Theorem 50.15. *(Monodromy Principle I) Let γ_s, with $s \in I$, be a family of homotopic paths with fixed endpoints. If the germ $f^\star$ admits an analytic continuation $f^\star(\gamma_s)$ along any parth γ_s, then $f^\star(\gamma_0) = f^\star(\gamma_1)$.*

Example 50.5. (The Complex Logarithm) Our starting point is the function element (f_0, D_0) in the disk $D_0 := B_1(1)$ with

$$f_0(z) = \sum_{k=1}^\infty \frac{(-1)^{k+1}}{k}(z-1)^k = \log|z| + i\mathrm{Arg}z. \tag{136}$$

In order to prove that this function element admits an unrestricted analytic continuation in $\mathbb{C}\backslash\{0\}$, we consider any path $\gamma\colon I \to \mathbb{C}\backslash\{0\}$ with initial point $z_0 = 1$ and arbitrary terminal point z_1.

In order to construct function elements of an analytic continuation of (f_0, D_0) along γ, we first pick a point $z_t := \gamma(t)$ on γ and denote by $D_t := B_{|z_t|}(z_t)$ for all $t \in I$ (the largest disk around z_t contained in $\mathbb{C}\backslash\{0\}$). To find an appropriate argument of z_t, we denote by $t \mapsto a(t)$ the continuous branch of the argument along γ which is equal to the principle value $\mathrm{Arg} 1 = 0$ at its initial point and set $\arg_\gamma z_t := a(t)$. Finally, we define the function element (f_t, D_t) by

$$f_t(z) := \log |z_t| + i \arg_\gamma z_t + \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k z_t^k} (z - z_t)^k$$

for all $z \in D_t$. The series on the right-hand side results from substituting z by z/z_t in (136), so that D_t is indeed its disk of convergence.

51 Holomorphic Functions

51.1 Definition of Holomorphic Function

Let Ω be an open subset of $\mathbb{C}$ and let f be a complex-valued function defined on Ω. The function f is said to be **holomorphic at the point** $z \in \Omega$ if the quotient

$$\frac{f(z+h) - f(z)}{h}$$

converges to a limit when $h \to 0$. Here $h \in \mathbb{C}$ and $h \neq 0$ with $z + h \in \Omega$ so that the quotient is well defined. The limit of the quotient, when it exists, is denoted by $f'(z)$, and is called the **derivative of** f **at** z:

$$f'(z) := \lim_{h \to 0} \frac{f(z+h) - f(z)}{h}$$

The function f is said to be **holomorphic on** Ω if it is holomorphic at every point of Ω. If C is a closed subset of $\mathbb{C}$, we say that f is **holomorphic on** C if f is holomorphic in some open set containing C. Note that if f is holomorphic at a point $z \in \mathbb{C}$, then it is not necessarily holomorphic on $\{z\}$. If f is holomorphic on all of $\mathbb{C}$ we say that f is **entire**.

51.2 Examples of Holomorphic Functions

Example 51.1.

1. Let $f \colon \mathbb{C} \to \mathbb{C}$ be given by $f(z) = z$ for all $z \in \mathbb{C}$. Then f is entire. Indeed, let $z \in \mathbb{C}$. Then

$$\begin{aligned} \lim_{h \to 0} \frac{f(z+h) - f(z)}{h} &= \lim_{h \to 0} \frac{z + h - z}{h} \\ &= \lim_{h \to 0} \frac{h}{h} \\ &= 1. \end{aligned}$$

2. Let $f \colon \mathbb{C} \to \mathbb{C}$ be given by $f(z) = \overline{z}$ for all $z \in \mathbb{C}$. Then f is continuous everywhere in $\mathbb{C}$ but is not holomorphic at any point in $\mathbb{C}$. To see this, let $z \in \mathbb{C}$. Then

$$\begin{aligned} \lim_{h \to 0} \frac{f(z+h) - f(z)}{h} &= \lim_{h \to 0} \frac{\overline{z+h} - \overline{z}}{h} \\ &= \lim_{h \to 0} \frac{\overline{z} + \overline{h} - \overline{z}}{h} \\ &= \lim_{h \to 0} \frac{\overline{h}}{h}. \end{aligned}$$

But this limit doesn't exist. Indeed, assume it did exist (to obtain a contradiction). Then setting $h = \varepsilon$, where $\varepsilon \in \mathbb{R}$, and taking $\varepsilon \to 0$, we see that

$$\begin{aligned} \lim_{h \to 0} \frac{\overline{h}}{h} &= \lim_{\varepsilon \to 0} \frac{\overline{\varepsilon}}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \frac{\varepsilon}{\varepsilon} \\ &= 1. \end{aligned}$$

On the other hand, setting $h = i\varepsilon$, where $\varepsilon \in \mathbb{R}$, and taking $\varepsilon \to 0$, we see that

$$\begin{aligned}\lim_{h\to 0}\frac{\overline{h}}{h} &= \lim_{i\varepsilon\to 0}\frac{\overline{i\varepsilon}}{i\varepsilon}\\ &= \lim_{i\varepsilon\to 0}\frac{-i\varepsilon}{i\varepsilon}\\ &= -1.\end{aligned}$$

This is a contradiction. We conclude that the limit does not exist.

3. Let $f\colon \mathbb{C}\to\mathbb{C}$ be given by $f(z) = |z|^2 = z\overline{z}$ for all $z \in \mathbb{C}$. Then f is holomorphic at the point 0 but nowhere else. Indeed,

$$\begin{aligned}\lim_{h\to 0}\frac{f(0+h)-f(0)}{h} &= \lim_{h\to 0}\frac{|h|^2}{h}\\ &= \lim_{h\to 0}\frac{h\overline{h}}{h}\\ &= \lim_{h\to 0}\overline{h}\\ &= 0,\end{aligned}$$

implies that f is holomorphic at 0. Now assume (to obtain a contradiction) that f is holomorphic at some $w \neq 0$. Let $g\colon \mathbb{C}\to\mathbb{C}$ be the identity function, given by $g(z) = z$ for all $z \in \mathbb{C}$. Then since g is holomorphic at w and $g(w) \neq 0$, the quotient f/g must be holomorphic at w as well. But this is a contradiction since the quotient is the complex-conjugation function, given by $f(z)/g(z) = \overline{z}$ for all $z \in \mathbb{C}$, which we know is not holomorphic anywhere in $\mathbb{C}$. This example demonstrates that a function being holomorphic at a point does *not* imply it being holomorphic in some neighborhood of that point.

4. Let $f\colon \mathbb{C}\backslash\{0\}\to\mathbb{C}$ be given by $f(z) = 1/z$ for all $z \in \mathbb{C}$. The f is holomorphic in its domain $\mathbb{C}\backslash\{0\}$. Indeed, let $z \in \mathbb{C}\backslash\{0\}$. Then

$$\begin{aligned}\lim_{h\to 0}\frac{f(z+h)-f(z)}{h} &= \lim_{h\to 0}\frac{\frac{1}{z+h}-\frac{1}{z}}{h}\\ &= \lim_{h\to 0}\frac{\frac{-h}{(z+h)z}}{h}\\ &= \lim_{h\to 0}\frac{-1}{(z+h)z}\\ &= \frac{-1}{z^2}.\end{aligned}$$

z_{0}

51.3 Holomorphic Functions form a $\mathbb{C}$-Vector Space

Proposition 51.1. *Let f and g be complex-valued functions defined in a neighborhood of a point z in the complex plane. Then for all $a, b \in \mathbb{C}$, the function $af + bg$ is holomorphic at z. Moreover, we have*

$$(af+bg)'(z) = af'(z) + bg'(z).$$

Proof. This follows from linearity of the limit operator:

$$\begin{aligned}(af+bg)'(z) &= \lim_{h\to 0}\frac{(af+bg)(z+h)-(af+bg)(z)}{h}\\ &= \lim_{h\to 0}\frac{af(z+h)+bg(z+h)-af(z)-bg(z)}{h}\\ &= a\lim_{h\to 0}\frac{f(z+h)-f(z)}{h} + b\lim_{h\to 0}\frac{g(z+h)-g(z)}{h}\\ &= af'(z)+bg'(z).\end{aligned}$$

□

51.4 Chain Rule and Product Rule

Let Ω be an open subset of $\mathbb{C}$ and f a complex-valued function on Ω. Recall that f is continuous at $z_0 \in \Omega$ if and only if there exists a small neighborhood $U \subseteq \Omega$ of z_0 and a function $\psi : U \to \mathbb{C}$ where $\psi(z) \to 0$ as $z \to z_0$ such that

$$f(z) = f(z_0) + \psi(z), \tag{137}$$

for all $z \in U$. Indeed, we can pick $U = \Omega$ and define $\psi : U \to \mathbb{C}$ by $\psi(z) = f(z) - f(z_0)$ for all $z \in U$. Then continuity of f at z_0 implies $\psi(z) \to 0$ as $z \to z_0$, and conversely the expression (137) together with the fact that $\psi(z) \to 0$ as $z \to z_0$ implies f is continuous at z_0.

When f is holomorphic at $z_0 \in \Omega$, there is an even better approximation of f at z_0:

Lemma 51.1. *Let Ω be an open subset of $\mathbb{C}$ and f a complex-valued function on Ω. Then f is holomorphic at $z_0 \in \Omega$ if and only if there exists a complex number a (which is necessarily equal to $f'(z_0)$ as limits are unique) such that*

$$f(z_0 + h) = f(z_0) + ah + \psi(h)h, \tag{138}$$

where ψ is a function defined for all small h and $\lim_{h\to 0} \psi(h) = 0$.

Remark 65. By setting $h = z - z_0$, we can rewrite (138) as

$$f(z) = f(z_0) + a(z - z_0) + \psi_0(z)(z - z_0),$$

where $\psi_0(z) = \psi(z - z_0)$ (and so $\psi_0(z) \to 0$ as $z \to z_0$).

Proof. Assume (138) holds. Then

$$\lim_{h\to 0} \frac{f(z_0 + h) - f(z_0)}{h} = \lim_{h\to 0} \frac{ah + \psi(h)h}{h} = a$$

implies f is holomorphic at z_0. Conversely, assume that f is holomorphic at z_0. Define ψ as

$$\psi(h) = \frac{f(z_0 + h) - f(z_0) - f'(z_0)h}{h}.$$

Then ψ satisfies the desired properties and (138) holds. □

Let us now show how we can use Lemma (51.1) to prove the Chain Rule and Product Rule. First we prove the Chain Rule:

Proposition 51.2. *Let f be holomorphic at z_0 and g be holomorphic at $f(z_0)$. Then $g \circ f$ is holomorphic at z_0 and, moreover, the Chain Rule holds*

$$(g \circ f)'(z_0) = g'(f(z_0))f'(z_0).$$

Proof. Since f is holomorphic at z_0, we can express f locally at z_0 as

$$f(z_0 + h) = f(z_0) + f'(z_0)h + \psi_1(h)h$$

where ψ_1 is a function defined for all small h and $\psi_1(h) \to 0$ as $h \to 0$. Since g is holomorphic at $f(z_0)$, we can express g locally at $f(z_0)$ as

$$g(f(z_0) + h) = g(f(z_0)) + g'(f(z_0))h + \psi_2(h)h$$

where ψ_2 is a function defined for all small h and $\psi_2(h) \to 0$ has $h \to 0$. Using these local expressions, we can now express $g \circ f$ locally at z_0:

$$\begin{aligned}
(g \circ f)(z_0 + h) &= g(f(z_0 + h)) \\
&= g(f(z_0) + f'(z_0)h + \psi_1(h)h) \\
&= g(f(z_0)) + g'(f(z_0))(f'(z_0)h + \psi_1(h)h) + \psi_2(h)(f'(z_0)h + \psi_1(h)h) \\
&= g(f(z_0)) + g'(f(z_0))f'(z_0)h + \psi_3(h)h
\end{aligned}$$

where $\psi_3(h) = g'(f(z_0))\psi_1(h)h + \psi_2(h)f'(z_0)h + \psi_1(h)\psi_2(h)h$. Since ψ_3 is a function defined for all small h and $\psi_3(h) \to 0$ has $h \to 0$, it follows from uniqueness of limits that

$$(g \circ f)'(z_0) = g'(f(z_0))f'(z_0).$$

□

Corollary 20. *Let f be holomorphic at z_0. If $f(z_0) \neq 0$, then $1/f$ is holomorphic at z_0*

Proof. The function $1/f$ can be viewed as the composition of $g \circ f$, where g is given by $g(z) = 1/z$. Then $1/f$ is holomorphic at z_0 since f is holomorphic at z_0 and g is holomorphic at $f(z_0)$ (because $f(z_0) \neq 0$). □

Now we will prove the Product Rule:

Proposition 51.3. *Let f and g be holomorphic at z_0. Then fg is holomorphic at z_0 and, moreover, the Product Rule holds*

$$(fg)'(z_0) = f'(z_0)g(z_0) + f(z_0)g'(z_0).$$

Proof. Since f is holomorphic at z_0, we can express f locally at z_0 as

$$f(z_0 + h) = f(z_0) + f'(z_0)h + \psi_1(h)h$$

where ψ_1 is a function defined for all small h and $\psi_1(h) \to 0$ as $h \to 0$. Since g is holomorphic at z_0, we can express f locally at z_0 as

$$g(z_0 + h) = g(z_0) + g'(z_0)h + \psi_2(h)h$$

where ψ_2 is a function defined for all small h and $\psi_2(h) \to 0$ as $h \to 0$. Then

$$\begin{aligned}(fg)(z_0 + h) &= f(z_0 + h)g(z_0 + h) \\ &= (f(z_0) + f'(z_0)h + \psi_1(h)h)(g(z_0) + g'(z_0)h + \psi_2(h)h) \\ &= f(z_0)g(z_0) + (f(z_0)g'(z_0) + f'(z_0)g(z_0))h + \psi_3(h)h,\end{aligned}$$

where $\psi_3(h) = f(z_0)\psi_2(h) + f'(z_0)g'(z_0)h + g(z_0)\psi_1(h)$. Since ψ_3 is a function defined for all small h and $\psi_3(h) \to 0$ as $h \to 0$, it follows from uniqueness of limits that

$$(fg)'(z_0) = f'(z_0)g(z_0) + f(z_0)g'(z_0)$$

□

Since the function $f : \mathbb{C} \to \mathbb{C}$, given by $f(z) = z$, is holomorphic, it follows from Proposition (51.1), Proposition (**??**) and the fact that the function $f : \mathbb{C} \to \mathbb{C}$, given by $f(z) = z$, is entire, that polynomials are entire.

51.5 Analytic Functions are Holomorphic

Let Ω be an open set and let $f : \Omega \to \mathbb{C}$. We say f is **analytic** if at each $a \in \Omega$, there exists an open neighborhood U of a and a power series $\sum a_n(z-a)^n$ centered at a such that $U \subseteq \Omega$ and

$$f(z) = \sum_{n=0}^{\infty} a_n(z-a)^n$$

for all $z \in U$.

Proposition 51.4. *Let Ω be an open set and let $f : \Omega \to \mathbb{C}$ be analytic. Then f is holomorphic.*

Proof. Let $a \in \Omega$. Choose $r > 0$ and a power series $\sum a_n(z-a)^n$ centered at a such that $B_r(a) \subset \Omega$ and

$$f(z) = \sum_{n=0}^{\infty} a_n(z-a)^n$$

for all $z \in B_r(a)$. We claim that f is holomorphic in $B_r(a)$. Let $\varepsilon > 0$ such that $B_\varepsilon(z) \subset B_r(a)$ and let $z \in B_r(a)$.

Then for all $h \in B_\varepsilon(0)$, we have

$$\begin{aligned}
f'(z) &= \lim_{h\to 0} \frac{f(z+h)-f(z)}{h} \\
&= \lim_{h\to 0} \frac{1}{h} \sum_{n=0}^{\infty} a_n \left((z+h-a)^n - (z-a)^n\right) \\
&= \lim_{h\to 0} \lim_{n\to\infty} \frac{1}{h} \sum_{m=1}^{n} a_m \left((z+h-a)^m - (z-a)^m\right) \\
&= \lim_{h\to 0} \lim_{n\to\infty} \sum_{m=1}^{n} a_m \sum_{k=1}^{m} (z-a+h)^{m-k}(z-a)^{k-1} \\
&= \lim_{n\to\infty} \lim_{h\to 0} \sum_{m=1}^{n} a_m \sum_{k=1}^{m} (z-a+h)^{m-k}(z-a)^{k-1} \\
&= \lim_{n\to\infty} \sum_{m=1}^{n} m a_m (z-a)^{m-1} \\
&= \sum_{n=1}^{\infty} n a_n (z-a)^{n-1}.
\end{aligned}$$

We need to justify why we were allowed to swap limits. Let $g_m \colon B_\varepsilon(0) \to \mathbb{C}$ be given by

$$g_m(h) = a_m \sum_{k=1}^{m} (z-a+h)^{m-k}(z-a)^{k-1}.$$

We need to show that the series $\sum g_m$ converges uniformly. In fact, this follows from an easy application of Weierstrass M-test. We first observe that

$$\begin{aligned}
|g_m(h)| &= \left| a_m \sum_{k=1}^{m} (z-a+h)^{m-k}(z-a)^{k-1} \right| \\
&< \left| m a_m r^{m-1} \right|.
\end{aligned}$$

Now we just set $M_m = |m a_m r^{m-1}|$ and apply Weierstrass M-test. □

One of the great triumphs of complex analysis is that the converse to Proposition (51.4) is also true, namely holomorphic functions are analytic.

51.6 Cauchy-Riemann Equations

Throughout this subsection, let f be a complex-valued function defined on some open subset Ω of $\mathbb{C}$ and fix a point $z_0 = x_0 + iy_0$ in Ω. Since $\mathbb{C}$ is a 2-dimensional $\mathbb{R}$-vector space, there is a unique decomposition of a complex number z as

$$z = x + iy,$$

where x and y are real numbers. Similarly, there is a unique decomposition of f as

$$f = u + iv,$$

where u and v are real-valued functions defined on Ω.

We define a map $\widetilde{\cdot} : \mathbb{C} \to \mathbb{R}^2$, given by mapping the complex number $z = x + iy$ to the vector $\widetilde{z} = (x, y)$. We also define $\widetilde{u} : \mathbb{R}^2 \to \mathbb{R}$ and $\widetilde{v} : \mathbb{R}^2 \to \mathbb{R}$ by the formulas $\widetilde{u}(x,y) = u(x+iy)$ and $\widetilde{v}(x,y) = v(x+iy)$ respectively. Similarly, we define $\widetilde{f} : \mathbb{R}^2 \to \mathbb{R}^2$ by the formula $\widetilde{f}(x,y) = (\widetilde{u}(x,y), \widetilde{v}(x,y))$.

We say $\widetilde{f}$ is **differentiable** at $\widetilde{z}_0$ if there exists a linear transformation $J\widetilde{f}(\widetilde{z}_0) : \mathbb{R}^2 \to \mathbb{R}^2$ such that

$$\frac{\|\widetilde{f}(\widetilde{z}_0 + \widetilde{h}) - \widetilde{f}(\widetilde{z}_0) - J\widetilde{f}(\widetilde{z}_0)(\widetilde{h})\|}{\|\widetilde{h}\|} \to 0$$

as $\widetilde{h} \to 0$ where $\widetilde{h} = (h_1, h_2) \in \mathbb{R}^2$. Equivalently, we can write

$$\widetilde{f}(\widetilde{z}_0 + \widetilde{h}) = \widetilde{f}(\widetilde{z}_0) + J\widetilde{f}(\widetilde{z}_0)(\widetilde{h}) + \|\widetilde{h}\|\psi(\widetilde{h}),$$

where $\widetilde{\psi}(\widetilde{h}) \to 0$ as $\widetilde{h} \to 0$. The linear transformation $J\widetilde{f}(\widetilde{z}_0)$ is unique and is called the **derivative** of $\widetilde{f}$ at $\widetilde{z}_0$. If $\widetilde{f}$ is differentiable, then the partial derivatives of its component functions exist, and the linear transformation $J\widetilde{f}(\widetilde{z}_0)$ is described in the standard basis of $\mathbb{R}^2$ by the Jacobian matrix of $\widetilde{f}$:

$$J\widetilde{f}(\widetilde{z}_0) = \begin{pmatrix} \partial_x\widetilde{u}(\widetilde{z}_0) & \partial_y\widetilde{u}(\widetilde{z}_0) \\ \partial_x\widetilde{v}(\widetilde{z}_0) & \partial_y\widetilde{v}(\widetilde{z}_0) \end{pmatrix}.$$

In the case of complex-differentiation the complex-derivative is a complex number $f'(z_0)$, while in the case of real-derivatives, it is a matrix. There is, however, a connection between these two notions, which is given in terms of special relations that are satisfied by the entries of the Jacobian matrix, that is, the partials of u and v.

Theorem 51.2. *If f is holomorphic at z_0, then the partial derivatives $\partial_x\widetilde{u}$, $\partial_y\widetilde{u}$, $\partial_x\widetilde{v}$, and $\partial_y\widetilde{v}$ exist and satisfy the **Cauchy-Riemann Equations** at z_0:*

$$\begin{aligned} \partial_x\widetilde{u}(\widetilde{z}_0) &= \partial_y\widetilde{v}(\widetilde{z}_0) \\ -\partial_y\widetilde{v}(\widetilde{z}_0) &= \partial_x\widetilde{v}(\widetilde{z}_0). \end{aligned}$$

Moreover, $\widetilde{f}$ is real-differentiable and its Jacobian at the point $\widetilde{z}_0$ satisfies

$$det\ J\widetilde{f}(\widetilde{z}_0) = |f'(z_0)|^2.$$

Proof. Let $\varepsilon > 0$. Then

$$\begin{aligned} f'(z_0) &= \lim_{\varepsilon\to 0} \frac{f(z_0+\varepsilon) - f(z_0)}{\varepsilon} \\ &= \lim_{\varepsilon\to 0} \frac{u((x_0+\varepsilon)+iy_0) + iv((x_0+\varepsilon)+iy_0) - u(x_0+iy_0) - iv(x_0+iy_0)}{\varepsilon} \\ &= \lim_{\varepsilon\to 0} \frac{u((x_0+\varepsilon)+iy_0) - u(x_0+iy_0)}{\varepsilon} + i\lim_{\varepsilon\to 0} \frac{v((x_0+\varepsilon)+iy_0) - v(x_0+iy_0)}{\varepsilon} \\ &= \partial_x\widetilde{u}(\widetilde{z}_0) + i\partial_x\widetilde{v}(\widetilde{z}_0). \end{aligned}$$

Similarly,

$$\begin{aligned} f'(z_0) &= \lim_{i\varepsilon\to 0} \frac{f(z_0+i\varepsilon) - f(z_0)}{i\varepsilon} \\ &= \lim_{\varepsilon\to 0} \frac{u(x_0+i(y_0+\varepsilon)) + iv(x_0+i(y_0+\varepsilon)) - u(x_0+iy_0) - v(x_0+iy_0)}{i\varepsilon} \\ &= \lim_{\varepsilon\to 0} \frac{v(x_0+i(y_0+\varepsilon)) - v(x_0+iy_0)}{\varepsilon} - i\lim_{\varepsilon\to 0} \frac{u(x_0+i(y_0+\varepsilon)) - u(x_0+iy_0)}{\varepsilon} \\ &= -i\partial_y\widetilde{u}(\widetilde{z}_0) + \partial_y\widetilde{v}(\widetilde{z}_0). \end{aligned}$$

Equating the two formulas for $f'(z_0)$ above yields the Cauchy-Riemann equations.

To see that $\widetilde{f}$ is differentiable, it suffices to observe that if $\widetilde{h} = (h_1, h_2)$ and $h = h_1 + ih_2$, then the Cauchy-Riemann equations imply

$$J\widetilde{f}(\widetilde{z}_0)(\widetilde{h}) = \left(\partial_x\widetilde{u} - i\partial_y\widetilde{u}\right)(h_1 + ih_2) =$$

complex-differentiability of f at z_0 implies

$$f(z_0+h) - f(z_0) = f'(z_0)(h) + h\psi(h),$$

note that

$$\widetilde{f}(\widetilde{z}_0+\widetilde{h}) - \widetilde{f}(\widetilde{z}_0) = J\widetilde{f}(\widetilde{z}_0)(\widetilde{h}) + \|\widetilde{h}\|\widetilde{\psi}(\widetilde{h}),$$

$$\widetilde{f}(\widetilde{z}_0+\widetilde{h}) - \widetilde{f}(\widetilde{z}_0) = J\widetilde{f}(\widetilde{z}_0)(\widetilde{h}) + \|\widetilde{h}\|\widetilde{\psi}(\widetilde{h}),$$

$$\det J\widetilde{f}(\widetilde{z}_0) = |f'(z_0)|^2.$$

□

Define the two differential operators

$$\partial_z = \frac{1}{2}\left(\partial_x - i\partial_y\right) \text{ and } \partial_{\overline{z}} = \frac{1}{2}\left(\partial_x + i\partial_y\right).$$

Proposition 51.5. *If f is holomorphic at $z_0 = x_0 + iy_0$, then*

$$\partial_{\overline{z}} f(z_0) = 0 \textit{ and } f'(z_0) = \partial_z f(z_0) = 2\partial_z u(z_0).$$

Also, if we consider f as a function in two real variables x, y, then f is real-differentiable $p = (x_0, y_0)$ and

$$\det J\widetilde{f}(p) = |f'(z_0)|^2.$$

Proof. First note that the Cauchy-Riemann equations at z_0 are equivalent to $\partial_{\overline{z}} f(z_0) = 0$. Indeed,

$$\begin{aligned}
0 &= \partial_{\overline{z}} f(z_0) \\
&= \frac{1}{2}\left(\partial_x + i\partial_y\right) f(z_0) \\
&= \frac{1}{2}\left(\partial_x f(z_0) + i\partial_y f(z_0)\right) \\
&= \frac{1}{2}\left(\partial_x u(z_0) + i\partial_x v(z_0) + i\partial_y u(z_0) - \partial_y v(z_0)\right) \\
&= \frac{1}{2}\partial_x u(z_0) - \partial_y v(z_0) + \frac{i}{2}\left(\partial_y u(z_0) + \partial_x v(z_0)\right),
\end{aligned}$$

and equating the real and imaginary parts gives us the Cauchy-Riemann equations.

Moreover, we have

$$\begin{aligned}
f'(z_0) &= \frac{1}{2}\left(\partial_x f(z_0) - i\partial_y f(z_0)\right) \\
&= \partial_x f(z_0),
\end{aligned}$$

and the Cauchy-Riemann equations give $\partial_z f = 2\partial_z u$.

To prove that f is real-differentiable, it suffices to observe that if $H = (h_1, h_2)$ and $h = h_1 + ih_2$, then the Cauchy-Riemann equations imply

$$Jf(p)(h) = (\partial_x u - i\partial_y u)(h_1 + ih_2) = f'(z_0)h$$

□

We can clarify the situation further by defining two differential operators

$$\partial_z := \frac{1}{2}\left(\partial_x - i\partial_y\right) \quad \text{and} \quad \partial_{\overline{z}} := \frac{1}{2}\left(\partial_x + i\partial_y\right)$$

Example 51.2. Let $f(z) = \overline{z} = x - iy$. Then

$$\partial_x u = 1 \neq -1 = \partial_y v,$$

so f is not differentiable.

Corollary 21.

1. *If f is holomorphic on $D_r(z_0)$ with $f'(z) = 0$ for all $z \in D_r(z_0)$, then f is constant on $D_r(z_0)$.*
2. *If f is holomorphic on $D_r(z_0)$ and real-valued on $D_r(z_0)$, then f is constant.*

Proof.

1. Since $f'(z) = 0$, we have $\partial_x u + i\partial_x v = 0 = \partial_y v - i\partial_y u$. This implies $\partial_x u = \partial_y u = \partial_x v = \partial_y v = 0$. Therefore u and v are constant functions, and hence, f is constant.
2. Let $f = u + iv$. Then $v = 0$ on $D_r(z_0)$. So $\partial_x v = \partial_y v = 0$ implies $\partial_x u = \partial_y u = 0$. Therefore u and v are constant, and so f is constant.

$\square$

Example 51.3. The function $f(z) = |z|^2$ is not analytic anywhere because it is real-valued and non-constant.

Theorem 51.3. *Let $f(z) = \sum_{n=0}^{\infty} a_n(z-z_0)^n$ have radius of convergence $R > 0$. Then f is analytic on $D_R(z_0)$ with*

$$f'(z) = \sum_{n=1}^{\infty} na_n(z-z_0)^{n-1}$$

for $|z-z_0| < R$.

52 Complex Integration

52.1 Paths

Throughout this section, let D be a nonempty subset of $\mathbb{C}$.

52.2 Definition of a Path

Definition 52.1. A **path** is a continuous function of the form $\gamma\colon [a,b] \to \mathbb{C}$ where $[a,b]$ is a closed interval in $\mathbb{R}$. The image set $[\gamma] := \gamma([a,b])$ is said to be the **trace** (or **trajectory**) of γ. If $[\gamma] \subseteq D$, then we say γ is a **path in** D. The space of all paths in D is denoted $\mathcal{P}(D)$. The points $\gamma(a)$ and $\gamma(b)$ are called the **source** and **target** of γ, respectively. We say γ is **simple** if $\gamma(s) = \gamma(t)$ with $s < t$ implies $s = a$ and $t = b$. We say γ is a **loop in** D **based at** z if $\gamma(a) = z = \gamma(b)$. The space of all loops in D is denoted $\mathcal{L}(D)$ and the space of all loops in D based at z is denoted $\mathcal{L}(D,z)$.

52.3 Reparametrization

Let $\gamma\colon [a,b] \to \mathbb{C}$ be a path. Suppose $\varphi\colon [c,d] \to [a,b]$ is a continuous function from the closed interval $[c,d]$ to the closed interval $[a,b]$. Then $\gamma \circ \varphi\colon [c,d] \to \mathbb{C}$ is a path and is called a **reparametrization** of γ. More specifically,

1. if $\varphi(c) = a$ and $\varphi(d) = b$, then we call $\gamma \circ \varphi\colon [c,d] \to \mathbb{C}$ a **positive reparametrization** of γ.
2. if $\varphi(c) = b$ and $\varphi(d) = a$, then we call $\gamma \circ \varphi\colon [c,d] \to \mathbb{C}$ a **negative reparametrization** of γ.
3. if φ is a linear map, then we call $\gamma \circ \varphi\colon [c,d] \to \mathbb{C}$ a **linear reparametrization** of γ.
4. if $\gamma \circ \varphi$ is a positive linear reparametrization whose domain is I, then the map $\varphi\colon I \to [a,b]$ is uniquely determined: it is given by the formula $\varphi(t) = a(1-t) + tb$ for all $t \in I$. In this case, we call $\gamma \circ \varphi\colon I \to \mathbb{C}$ *the* **normalized form** of γ. Any path whose domain is I is called a **normal path**.

As it turns out, the choice of the parameter interval $[a,b]$ is not that essential for the definition of a path. Thus we will usually only work with normal paths. Any contstruction we describe which uses normal paths can easily be extended to all paths by taking their normalized forms.

52.4 Standard Examples

Example 52.1. Let $w, z \in D$, $r > 0$, and let $\gamma\colon I \to D$ be a path in D.

1. The **constant path at** z is the path $c_z\colon I \to \mathbb{C}$ defined by $c_z(t) = z$ for all $t \in I$.
2. The **oriented line segment from** w **to** z is the path $[w,z]\colon I \to D$ defined by $[w,z](t) = w(1-t) + tz$ for all $t \in I$.
3. The **standard parametrization of** $C_r(z)$ is the path $\gamma_r(z)\colon I \to \mathbb{C}$ defined by $\gamma_r(z)(t) = z + re^{2\pi it}$ for all $t \in I$.
4. The **reversed path** (or **negative path**) γ^- of γ is the path $\gamma^-\colon I \to \mathbb{C}$ defined by the formula $\gamma^-(t) = \gamma(1-t)$ for all $t \in I$.

52.5 Concatenation of Paths

let $\gamma_1 \colon I \to \mathbb{C}$ and $\gamma_2 \colon I \to \mathbb{C}$ be two paths such that $\gamma_1(1) = \gamma_2(0)$. Then their **concatenation** $\gamma_2 \oplus \gamma_1$ is a path $\gamma_2 \oplus \gamma_1 \colon I \to \mathbb{C}$ defined by the formula

$$(\gamma_2 \oplus \gamma_1)(t) = \begin{cases} \gamma_1(2t) & 0 \leq t \leq \frac{1}{2} \\ \gamma_2\left(2\left(t - \frac{1}{2}\right)\right) & \frac{1}{2} < t \leq 1. \end{cases}$$

for all $t \in [0,1]$. The idea is that we traverse γ_1 and then γ_2 all in one day. The $2t$ in $\gamma_1(2t)$ comes from the fact that we need to traverse γ_1 twice as fast, the $t - \frac{1}{2}$ in $\gamma_2\left(2\left(t - \frac{1}{2}\right)\right)$ comes from the fact that we need to wait half a day before we start traversing γ_2, and the $2\left(t - \frac{1}{2}\right)$ in $\gamma_2\left(2\left(t - \frac{1}{2}\right)\right)$ comes from the fact that we need to traverse γ_2 twice as fast.

52.6 Polygonal Paths and Paraxial Paths

A **polygonal path** is the sum $\gamma = \gamma_1 \oplus \cdots \oplus \gamma_n$ of segments $\gamma_k = [z_{k-1}, z_k]$ A polygonal path is called **paraxial** if each of its segments is parallel to the real or imaginary axis.

53 Homotopy

Throughout this subsection, let D be a nonempty subset of $\mathbb{C}$.

53.1 Homotopy of Paths

Let $\alpha \colon I \to D$ and $\beta \colon I \to D$ be two paths in D. We say that α is **homotopic** to β as paths, denoted $\alpha \sim \beta$, if there exists a continuous function $H \colon I \times I \to D$ such that $H(0,t) = \alpha(t)$ and $H(1,t) = \beta(t)$ for all $t \in I$. The map H is called a **homotopy** joining α to β. It is easy to check that $\sim$ is an equivalence relation in the set $\mathcal{P}(D)$.

53.2 Homotopy of Paths with Fixed Endpoints

Let $\alpha \colon I \to D$ and $\beta \colon I \to D$ be two paths in D with the same source and target, say $\alpha(0) = w = \beta(0)$ and $\alpha(1) = z = \beta(1)$. Then a homotopy H from α to β is called a **homotopy with fixed endpoints** if we additionally have $H(s,0) = w$ and $H(s,1) = z$ for all $s \in I$.

53.3 Homotopy of Loops

Let α and β be two loops in D based at z. We say α is homotopic to β as loops based at z, denoted $\alpha \sim_z \beta$, if there exists a homotopy $H \colon I \times I \to D$ with fixed endpoints from α to β. The space of all loops in D based at z is denoted $\mathcal{L}(D,z)$. It is easy to check that $\sim_z$ is an equivalence relation in the set $\mathcal{L}(D,z)$.

53.4 Free Homotopy of Loops

There is a slightly more general notion of homotopy in the context of loops which we will also consider. Let α and β be two loops in D based at z. We say α is **freely homotopic** to β, denoted $\alpha \sim_{\text{free}} \beta$ if there exists a homotopy $H \colon I \times I \to D$ from α to β such that $H(s,0) = H(s,1)$ for all $s \in I$. Such a homotopy is called a **free homotopy** from α to β. The added condition "$H(s,0) = H(s,1)$ for all $s \in I$" ensures that for each $s \in I$, the function $H(s,-) \colon I \to D$ is a loop. So the base point is allowed to move freely as s goes from 0 to 1.

53.5 The Fundamental Group $\pi(D,z)$

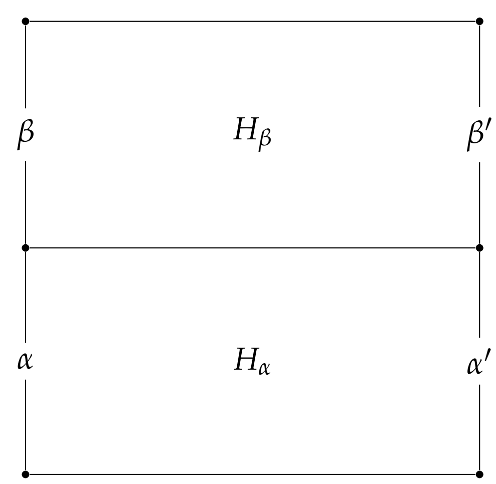
Let $z \in D$. Concatenation serves as a natural binary operation on $\mathcal{L}(D,z)$. However, this binary operation is rather complicated. For example, it isn't associative, it has no identity, and it has no inverses. It turns out however, if we consider loops up to homotopy with fixed endpoints, then we do get these properties. In fact, we will show that $(\mathcal{L}(D,z)/\sim_z, \oplus)$ forms a group, called the **fundamental group of D based at z**. Let us show this in the following sequence of steps:

53.5.1 Concatenation Passes to Quotient

First we need to show that concatenation passes to the quotient $\mathcal{L}(D,z)/\sim_z$, thus giving a well-defined binary operation on $\mathcal{L}(D,z)/\sim_z$. Let $\alpha\colon I \to D$ and $\beta\colon I \to D$ be two loops based at z. Suppose α' and β' are two loops based at z such that $\alpha' \sim_z \alpha$ and $\beta' \sim_z \beta$. Let H_α and H_β be their respective homotopies with endpoints. Define $H\colon I \times I \to X$ by the formula

$$H(s,t) = \begin{cases} H_\alpha(s,2t) & 0 \leq t \leq \frac{1}{2} \\ H_\beta(s,2t-1) & \frac{1}{2} < t \leq 1. \end{cases}$$

for all $(s,t) \in I \times I$. Then H is easily seen to be a homotopy with fixed endpoints from $\beta \oplus \alpha$ to $\beta' \oplus \alpha'$. One may visualize this homotopy as below:

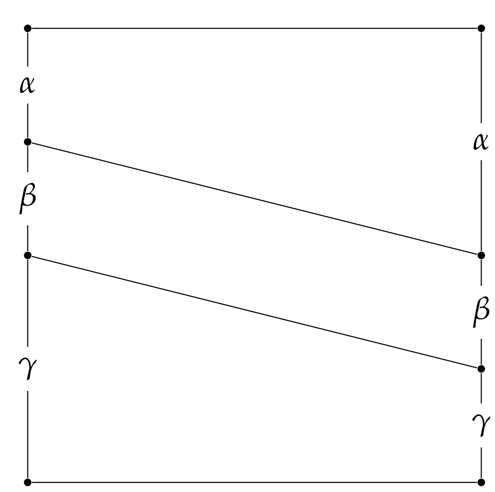


53.5.2 Associativity

Next we show that $(\mathcal{L}(D,z)/\sim_z, \oplus)$ is associative. Suppose α, β, and γ are loops based at z. Define $H\colon I \times I \to X$ by the formula

$$H(s,t) = \begin{cases} \gamma\left(\left(\frac{4}{2-s}\right)\cdot t\right) & 0 \leq t \leq \frac{2-s}{4} \\ \beta\left(4\cdot\left(t-\left(\frac{2-s}{4}\right)\right)\right) & \frac{2-s}{4} < t \leq \frac{3-s}{4} \\ \alpha\left(\left(\frac{4}{1+s}\right)\cdot\left(t-\left(\frac{3-s}{4}\right)\right)\right) & \frac{3-s}{4} < t \leq 1 \end{cases}$$

for all $(s,t) \in I \times I$. Then H is easily seen to be a homotopy with fixed endpoints from $(\alpha \oplus \beta) \oplus \gamma$ to $\alpha \oplus (\beta \oplus \gamma)$. One may visualize this homotopy as below:

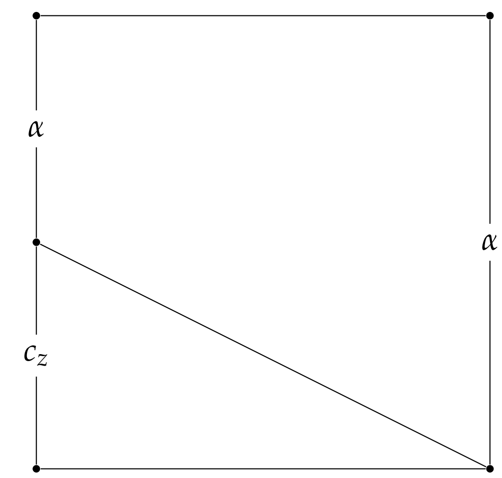


53.5.3 Identity

Next, we want to show that c_z represents the identity element in $(\mathcal{L}(D,z)/\sim_z, \oplus)$. Let α be a loop in D based at z. Define $H\colon I \times I \to X$ by the formula

$$H(s,t) = \begin{cases} c_z\left(\left(\frac{2}{1-s}\right)t\right) & 0 \leq t < \frac{1-s}{2} \\ \alpha\left(\left(\frac{2}{1+s}\right)\left(t-\left(\frac{1-s}{2}\right)\right)\right) & \frac{1-s}{2} \leq t \leq 1 \end{cases}$$

for all $(s,t) \in I \times I$. Then H is easily seen to be a homotopy fixed at z from $\alpha \oplus c_z$ to c_z. One may visualize this homotopy as below:



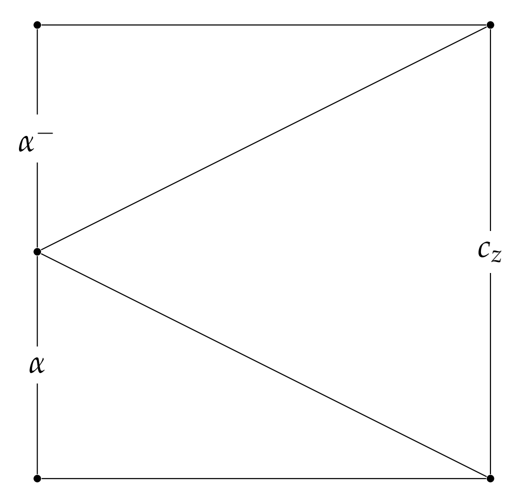
A similar argument gives a homotopy fixed at z from $c_z \oplus \alpha$ to α.

53.5.4 Inverses

Finally, we want to show that $(\mathcal{L}(D,z)/\sim_z, \oplus)$ has inverses. Suppose α is a loop based at z. Define $H\colon I \times I \to X$ by

$$H(s,t) = \begin{cases} \alpha\left(\left(\frac{2}{1-s}\right)t\right) & 0 \le t < \frac{1-s}{2} \\ c_z\left(\left(\frac{1}{s}\right)\left(t - \left(\frac{1-s}{2}\right)\right)\right) & \frac{1-s}{2} \le t < \frac{1+s}{2} \\ \alpha^-\left(\left(\frac{2}{1-s}\right)\left(t - \frac{2-s}{2}\right)\right) & \frac{1+s}{2} \le t \le 1 \end{cases}$$

for all $(s,t) \in I \times I$. Then H is easily seen to be a homotopy fixed at z from $\alpha^- \oplus \alpha$ to c_z. One may visualize this homotopy as below:



A similar argument gives a homotopy fixed at z from $\alpha \oplus \alpha^-$ to c_z.

53.5.5 Changing the Base Point

We have shown that $(\mathcal{L}(D,z)/\sim_z, \oplus)$ forms a group. This group is called the **fundamental group of** D **based at** z and is denoted as $\pi_1(D,z)$. We now what to consider what happens if we change the base point $z \in D$ to another basepoint, say $w \in D$.

Proposition 53.1. *If D is path connected, then $\pi_1(D,z) \cong \pi_1(D,w)$.*

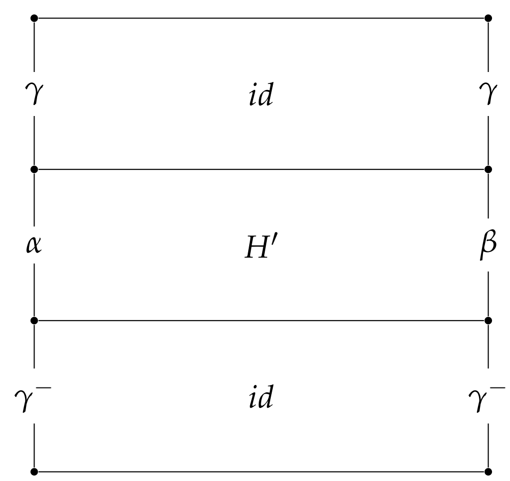
Proof. Choose a path γ from z to w. Define $-^\gamma\colon \pi_1(D,z) \to \pi_1(D,w)$ by the formula

$$\overline{\alpha}^\gamma = \overline{\gamma \oplus \alpha \oplus \gamma^-}$$

for all $\overline{\alpha} \in \pi_1(D,z)$. We need to show that this is a well-defined map, so choose another representative of the equivalence class $\overline{\alpha}$, say β, and let H' be a homotopy with fixed endpoints from α to β. Define $H\colon I \times I \to D$ by the formula

$$H(s,t) = \begin{cases} \gamma^-(3t) & 0 \leq t < \frac{1}{3} \\ H'(s,t) & \frac{1}{3} \leq t < \frac{2}{3} \\ \gamma\,(3t-2) & \frac{2}{3} \leq t \leq 1 \end{cases}$$

Then H is easily checked to be a homotopy with fixed endpoints from $\gamma \circ \alpha \circ \gamma^-$ to $\gamma \circ \beta \circ \gamma^-$. Thus, $-^{\gamma}$ is well-defined. One may visualize this homotopy as in the diagram below:



The map $-^{\gamma}$ is easily checked to be a group isomorphism with inverse being given by $-^{\gamma^-}$. $\square$

53.5.6 Simply Connected Domains

A domain D is called **simply connected** if for any two paths $\alpha\colon I \to D$ and $\beta\colon I \to D$ which share the same source and target are homotopic with fixed endpoints in D.

53.5.7 Null-Homotopic

A loop which is freely homotopic in D to a constant path is said to be **null-homotopic** in D.

Lemma 53.1. *For any path γ in D, the loop $\gamma \oplus \gamma^-$ is null-homotopic in D to its base point $c_{\gamma(0)}$.*

Proof. Define the function $H\colon I \times I \to D$ by the formula

$$H(s,t) = \begin{cases} \gamma(2t) & \text{if } 0 \leq t \leq \frac{1-s}{2} \\ c_{\gamma(1-s)}(t) & \text{if } \frac{1-s}{2} < t \leq \frac{1+s}{2} \\ \gamma^-(2t-1-s) & \text{if } \frac{1+s}{2} < t \leq 1 \end{cases}$$

Then H is easily checked to be a homotopy with fixed endpoints from $\gamma^- \oplus \gamma$ to $c_{\gamma(0)}$. $\square$

Lemma 53.2. *If a loop with base point z_0 is null-homotopic in D, then it is also homotopic with fixed endpoints to the constant path c_{z_0}.*

Proof. Let H be a homotopy from the given path γ_0 to a point z_1. We define γ_s and γ_s^+ by the formulas

$$\gamma_s(t) = H(s,t) \quad \text{and} \quad \gamma_s^+(t) = H(st,0)$$

for all $s,t \in I$, and we set $\gamma_s^- = (\gamma_s^+)^-$. Then the path γ_s^+ lies in D and connects z_0 with the moving basepoint $z_s = H(s,0) = H(s,1)$ of the loop γ_s. The family of loops

$$\gamma_s^* = \gamma_s^+ \oplus \gamma_s \oplus \gamma_s^-$$

has fixed base point z_0 and all paths in this family are homotopic in D. Now γ_0 is homotopic to γ_0^*, γ_0^* is homotopic to γ_1^*, and γ_1^* is homotopic to $\gamma_1^+ \oplus \gamma_1^-$, and the latter is homotopic to the base point z_0. $\square$

54 Smooth Paths

54.1 Definition of Smooth Path

Let $\gamma\colon [a,b] \to \mathbb{C}$ be a path. We say that γ is **smooth** if it is continuously differentiable on $[a,b]$ (i.e. $\gamma'(t)$ exists for $t \in [a,b]$ and the function $t \mapsto \gamma'(t)$ is continuous). At the points $t = a$ and $t = b$, the quantities $\gamma'(a)$ and $\gamma'(b)$ are interpreted as one-sided limits

$$\gamma'(a) = \lim_{\substack{h\to 0 \\ h>0}} \frac{\gamma(a+h) - \gamma(a)}{h} \quad \text{and} \quad \gamma'(b) = \lim_{\substack{h\to 0 \\ h<0}} \frac{\gamma(b+h) - \gamma(b)}{h}.$$

In general, these quantities are called the right-hand derivative of $\gamma(t)$ at a, and the left-hand derivative of $\gamma(t)$ at b, respectively. More generally, we say that γ is **piecewise smooth** if there exists a partition

$$a = t_0 < t_1 < \cdots < t_n = b,$$

of the interval $[a,b]$ such that the restriction $\gamma|_{[t_k,t_{k+1}]}\colon [t_k, t_{k+1}] \to \mathbb{C}$ is smooth for each $k = 0, \dots, n-1$. In particular, the right-hand derivative of γ at t_k may differ from the left-hand derivative of γ at t_k, for $k = 1, \dots, n-1$.

Example 54.1. The paths described in Example (52.1) are all smooth.

54.2 Integrating along a Smooth Path

Let $\gamma\colon [a,b] \to \mathbb{C}$ be a smooth path and suppose f is a continuous function defined on $[\gamma]$. Then the **integral of** f **along** γ is defined by the formula

$$\int_\gamma f(z)\mathrm{d}z = \int_a^b f(\gamma(t))\gamma'(t)\mathrm{d}t.$$

If γ is piecewise smooth, then we choose a partition

$$a = t_0 < t_1 < \cdots < t_n = b,$$

of the interval $[a,b]$ such that the restriction $\gamma|_{[t_k,t_{k+1}]}\colon [t_k, t_{k+1}] \to \mathbb{C}$ is smooth for each $k = 0, \dots, n-1$, and we define

$$\int_\gamma f(z)\mathrm{d}z = \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} f(\gamma(t))\gamma'(t)\mathrm{d}t.$$

54.3 Reparametrizing a Smooth Path

Definition 54.1. Let $\gamma\colon [a,b] \to \mathbb{C}$ be a smooth path. Suppose $\varphi\colon [c,d] \to [a,b]$ is a continuously differentiable function from the closed interval $[c,d]$ to the closed interval $[a,b]$. Then $\gamma \circ \varphi\colon [c,d] \to \mathbb{C}$ is a path and is called a **smooth reparametrization** of γ.

Remark 66. Note that a linear reparametrization is a smooth reparametrization. Thus the normalized form of a smooth path is also a smooth path.

As in the continuous case, the parameter interval $[a,b]$ is not that essential for the definition of a smooth path. For example, if $\gamma \circ \varphi\colon [c,d] \to \mathbb{C}$ is a smooth positive reparametrization of $\gamma\colon [a,b] \to \mathbb{C}$, then the change of variables formula with $t = \varphi(s)$ implies

$$\begin{aligned}
\int_\gamma f(z)\mathrm{d}z &= \int_a^b f(\gamma(t))\gamma'(t)\mathrm{d}t \\
&= \int_c^d f(\gamma(\varphi(s)))\gamma'(\varphi(s))\varphi'(s)\mathrm{d}s \\
&= \int_c^d f((\gamma\circ\varphi)(s))(\gamma\circ\varphi)'(s)\mathrm{d}s \\
&= \int_{\gamma\circ\varphi} f(z)\mathrm{d}z.
\end{aligned}$$

Thus we will usually only work with smooth normal paths. Any construction we describe which uses smooth normal paths can easily be extended to all paths by taking their normalized forms.

54.4 Defining the Length of a Path

Let $\gamma\colon [a,b] \to \mathbb{C}$ be a smooth path. Then the **length** of γ, denoted $\text{length}(\gamma)$, is defined by the formula

$$\text{length}(\gamma) := \int_a^b |\gamma'(t)|\, \mathrm{d}t.$$

If γ is piecewise-smooth, then we choose a partition

$$a = t_0 < t_1 < \cdots < t_n = b,$$

of the interval $[a,b]$ such that the restriction $\gamma|_{[t_k,t_{k+1}]}\colon [t_k, t_{k+1}] \to \mathbb{C}$ is smooth for each $k = 0, \dots, n-1$, and we define

$$\text{length}(\gamma) = \sum_{k=0}^{n-1} \text{length}\left(\gamma|_{[t_k,t_{k+1}]}\right).$$

54.5 Length of a Smooth Path equals the Length of its Normalized Form

Let $\gamma\colon [a,b] \to \mathbb{C}$ be a smooth path and let $\varphi\colon I \to [a,b]$ be given by $\varphi(t) = a(1-t) + bt$ for all $t \in I$ (so $\gamma \circ \varphi\colon I \to \mathbb{C}$ is the normalized form of $\gamma\colon [a,b] \to \mathbb{C}$). Then by the change of variables formula, we have

$$\begin{aligned}
\text{length}(\gamma) &= \int_a^b |\gamma'(t)|\, \mathrm{d}t. \\
&= \int_0^1 |\gamma'(\varphi(s))|\, \varphi'(s)\mathrm{d}s \\
&= \int_0^1 |\gamma'(\varphi(s))|\, |\varphi'(s)|\, \mathrm{d}s \\
&= \int_0^1 |\gamma'(\varphi(s))\varphi'(s)|\, \mathrm{d}s \\
&= \int_0^1 |(\gamma \circ \varphi)'(s)|\, \mathrm{d}s \\
&= \text{length}(\gamma \circ \varphi),
\end{aligned}$$

where we used the fact that for all $s \in I$, we have $\varphi'(s) = b - a > 0$.

54.6 Properties of Integration

Throughout this subsection, let Ω be an open subset of $\mathbb{C}$.

54.6.1 Linearity of Integration

Proposition 54.1. *Let $\gamma\colon I \to \Omega$ be a smooth path in Ω, let f, g be complex-valued functions defined on Ω, and let $\alpha, \beta \in \mathbb{C}$. Then*

$$\int_\gamma (\alpha f + \beta g)(z) dz = \alpha \int_\gamma f(z) dz + \beta \int_\gamma g(z) dz.$$

Proof. This follows from linearity of the Riemann integral. Indeed, we have

$$\begin{aligned}
\int_\gamma (\alpha f + \beta g)(z) dz &= \int_0^1 (\alpha f + \beta g)(\gamma(t))\gamma'(t) dt \\
&= \int_0^1 (\alpha f(\gamma(t))\gamma'(t) + \beta g(\gamma(t))\gamma'(t)) dt \\
&= \alpha \int_0^1 f(\gamma(t))\gamma'(t) dt + \beta \int_0^1 g(\gamma(t))\gamma'(t) dt \\
&= \alpha \int_\gamma f(z) dz + \beta \int_\gamma g(z) dz.
\end{aligned}$$

□

54.6.2 Additivity of Concatenation of Smooth Paths

Proposition 54.2. *Let $\gamma_1 \colon I \to \Omega$ and $\gamma_2 \colon I \to \Omega$ be smooth paths in Ω such that $\gamma_1(1) = \gamma_2(0)$, and let $f \colon \Omega \to \mathbb{C}$ be a function. Then*

$$\int_{\gamma_2 \oplus \gamma_1} f(z)dz = \int_{\gamma_2} f(z)dz + \int_{\gamma_1} f(z)dz.$$

Proof. This follows from additivity of the Riemann Integral. Indeed, we have

$$\begin{aligned}
\int_{\gamma_2 \oplus \gamma_1} f(z)dz &= \int_0^1 f((\gamma_2 \oplus \gamma_1)(t))(\gamma_2 \oplus \gamma_1)'(t)dt \\
&= 2\int_0^{1/2} f(\gamma_1(2t))\gamma_1'(2t)dt + 2\int_{1/2}^1 f(\gamma_2(2t-1))\gamma_2'(2t-1)dt \\
&= \int_0^1 f(\gamma_1(u))\gamma_1'(u)du + \int_0^1 f(\gamma_2(v))\gamma_2'(v)dv \\
&= \int_{\gamma_1} f(z)dz + \int_{\gamma_2} f(z)dz,
\end{aligned}$$

where we used the change of variables $u = 2t$ and $v = 2t - 1$.

□

54.6.3 Negativity of Reverse Orientation

Proposition 54.3. *Let $\gamma \colon I \to \Omega$ be a smooth paths in Ω, and let $f \colon \Omega \to \mathbb{C}$ be a function. Then*

$$\int_{\gamma^-} f(z)dz = -\int_{\gamma} f(z)dz$$

Proof. This follows from a straightforward calculation:

$$\begin{aligned}
\int_{\gamma^-} f(z)dz &= -\int_0^1 f(\gamma(1-t))\gamma'(1-t)dt \\
&= \int_1^0 f(\gamma(s))\gamma'(s)ds \\
&= -\int_0^1 f(\gamma(s))\gamma'(s)ds \\
&= -\int_{\gamma} f(z)dz,
\end{aligned}$$

where we used the change of variable $s = 1 - t$. □

54.6.4 Useful Inequality

Proposition 54.4. *Let $\gamma \colon I \to \Omega$ be a smooth path in Ω, and let $f \colon \Omega \to \mathbb{C}$ be a function. Then one has the inquality*

$$\left| \int_{\gamma} f(z)dz \right| \leq \|f(z)\|_{[\gamma]} \cdot \mathit{length}(\gamma).$$

Proof. This follows from a straightforward calculation:

$$\begin{aligned}
\left| \int_{\gamma} f(z)dz \right| &= \left| \int_0^1 f(\gamma(t))\gamma'(t)dt \right| \\
&\leq \int_0^1 \left| f(\gamma(t))\gamma'(t) \right| dt \\
&\leq \sup_{t \in I} |f(\gamma(t))| \cdot \int_0^1 \left| \gamma'(t) \right| dt \\
&= \sup_{z \in [\gamma]} |f(z)| \cdot \mathrm{length}(\gamma)
\end{aligned}$$

□

54.6.5 Primitives

Definition 54.2. Let $f\colon \Omega \to \mathbb{C}$ be a function. A **primitive** for f on Ω is a function $F\colon \Omega \to \mathbb{C}$ that is holomorphic on Ω and such that $F'(z) = f(z)$ for all $z \in \Omega$.

Theorem 54.1. *Let $\gamma\colon I \to \Omega$ be a piecewise smooth path in Ω from z_1 to z_2, and let $f\colon \Omega \to \mathbb{C}$ be a function. Suppose that $F\colon \Omega \to \mathbb{C}$ is a primitive for f in Ω. Then*

$$\int_\gamma f(z)dz = F(z_2) - F(z_1).$$

In particular the integral vanishes if γ is a loop.

Proof. We first assume that γ is smooth. Then by the chain rule and the fundamental theorem of calculus, we have

$$\begin{aligned}
\int_\gamma f(z)dz &= \int_0^1 f(\gamma(t))\gamma'(t)dt \\
&= \int_0^1 F'(\gamma(t))\gamma'(t)dt \\
&= \int_0^1 (F\circ\gamma)'(t)dt \\
&= (F\circ\gamma)(b) - (F\circ\gamma)(a) \\
&= F(\gamma(b)) - F(\gamma(a)). \\
&= F(z_2) - F(z_1).
\end{aligned}$$

Now we assume γ is piecewise smooth. Choose a partition

$$0 = t_0 < t_1 < \cdots < t_n = 1,$$

of the interval I such that the restriction $\gamma_k := \gamma|_{[t_k,t_{k+1}]}\colon [t_k, t_{k+1}] \to \mathbb{C}$ is smooth for each $k = 0, \dots, n-1$. Then

$$\begin{aligned}
\int_\gamma f(z)dz &= \sum_{k=0}^{n-1} \int_{\gamma_k} f(z)dz \\
&= \sum_{k=0}^{n-1} F(\gamma(t_{k+1})) - F(\gamma(t_k)) \\
&= F(\gamma(t_n)) - F(\gamma(t_0)) \\
&= z_2 - z_1.
\end{aligned}$$

□

Corollary 22. *Assume that Ω is path-connected. Let $f\colon \Omega \to \mathbb{C}$ be a holomorphic function and suppose that $f'(z) = 0$ for all $z \in \Omega$. Then f is constant on Ω.*

Proof. Fix a point $z_0 \in \Omega$. It suffices to show that $f(z) = f(z_0)$ for all $z \in \Omega$. Let $z \in \Omega$. Since Ω is connected, there exists a path $\gamma\colon I \to \Omega$ from z_0 to z. Since f is clearly a primitive for f', we have

$$\int_\gamma f'(w)dw = f(z) - f(z_0).$$

By assumption, $f' = 0$, so the integral on the left is 0, and we conclude that $f(z) = f(z_0)$ as desired. □

Integral Representation of the Taylor Coefficients

Theorem 54.2. *Let f be the sum of the power series $\sum a_n(z-a)^n$ with radius of convergence R. Then for any $n \geq 0$ and r such that $0 < r < R$, we have*

$$a_m = \frac{1}{r^m}\int_0^1 f(a + re^{2\pi it})e^{-2\pi imt}dt$$

Proof. By uniform convergence of the power series $\sum a_n(z-a)^n$ on $C_r(a)$, we have

$$\begin{aligned}\int_0^1 f(a+re^{2\pi it})e^{-2\pi imt}dt &= \int_0^1 \sum_{n=0}^{\infty} a_n r^n e^{2\pi i(n-m)t}dt \\ &= \sum_{n=0}^{\infty} a_n r^n \int_0^1 e^{2\pi i(n-m)t}dt \\ &= a_m r^m.\end{aligned}$$

□

55 More on Paths

55.1 Path Covering Lemma

Let $\gamma\colon I \to \mathbb{C}$ be a path. A **chain of disks covering** γ is a finite sequence $(D_0, D_1, \ldots, D_n)$ of open disks D_k with the following properties

1. There exists a partition $0 = t_0 < t_1 < \cdots < t_n = 1$ of the interval I such that $\gamma(t_k)$ is the center of D_k for $k = 0, 1, \ldots, n$.

2. The section of γ between $\gamma(t_{k-1})$ and $\gamma(t_{k+1})$ is contained in D_k, more precisely,

$$\begin{array}{lll}\gamma(t) \subset D_0, & t_0 \le t \le t_1, & \\ \gamma(t) \subset D_k, & t_{k-1} \le t \le t_{k+1}, & (k = 1, \ldots, n-1) \\ \gamma(t) \subset D_n, & t_{n-1} \le t \le t_n. & \end{array}$$

Lemma 55.1. *(Path Covering Lemma) Let Ω be a nonempty open connected subset of $\mathbb{C}$ and let $\gamma\colon I \to \Omega$ be a path in Ω. Then there exists a chain of disks which is contained in Ω and covers γ. Moreover, the radii of all disks can be chosen to be of the same size and arbitrarily small.*

Proof. Since γ is continuous on the compact interval I, its trace $[\gamma]$ is a compact subset of D. The complement of Ω in $\mathbb{C}$ is closed, and hence the distance d between $[\gamma]$ and $\mathbb{C}\backslash\Omega$ is positive. If $0 < r < d$, then all disks with radius r and centers on $[\gamma]$ are contained in Ω. Because γ is uniformly continuous, there exists a positive number δ such that $s, t \in I$ and $|s-t| < \delta$ imply that $|\gamma(s) - \gamma(t)| < r$. So all requirements are satisfied if the partition $0 = t_0 < t_1 < \cdots < t_n = 1$ is chosen such that $t_k - t_{k-1} < \delta$. □

55.2 Homotopic Paths with Specific Properties

Technically it is of great importance that any path in D can be approximated by homotopic paths with specific properties.

Lemma 55.2. *Let $\gamma\colon I \to D$ be a path in an open set $D \subseteq \mathbb{C}$. Then there exists a smooth path $\widetilde{\gamma}\colon I \to D$ and a paraxial path $\widehat{\gamma}\colon I \to D$ which are homotopic to γ in D. For each $\varepsilon > 0$ both paths can be chosen such that*

$$\|\gamma - \widetilde{\gamma}\|_I < \varepsilon \quad \textit{and} \quad \|\gamma - \widehat{\gamma}\|_I < \varepsilon.$$

Proof. By the path covering lemma, γ can be covered by a sequence of disks D_k with radii less than $\varepsilon/2$. Let

$$0 = t_0 < t_1 < \cdots < t_n = 1$$

be a subdivision of the parameter interval I, and denote by $z_k = \gamma(t_k)$ the centers of the covering disks D_k. Then the restriction γ_k of γ to $[t_{k-1}, t_k]$ is homotopic in D_k (and hence in D) to the line segment $[z_{k-1}, z_k]$ for all $k = 1, \ldots, n$. Indeed, D_k is convex, so the map $H\colon I \times I \to D_k$ given by

$$H(s,t) = \gamma_k(t)(1-s) + [z_{k-1}, z_k](t)s$$

for all $s, t \in I$ serves as a homotopy from γ_k to $[z_{k-1}, z_k]$.

This induces a homotopy between γ and the polygonal path $\widehat{\gamma} := [z_0, z_1] \oplus \cdots \oplus [z_{n-1}, z_n]$. Smoothing the function $\widehat{\gamma}$ at the points t_k appropriately, we also obtain a smooth path $\widetilde{\gamma}$ which is homotopic to $\widehat{\gamma}$ and hence to γ. Finally, the segments $[z_{k-1}, z_k]$ are homotopic in D_k to the sum $[z_{k-1}, \mathrm{Re}(z_k) + i\mathrm{Im}(z_{k-1})] \oplus [\mathrm{Re}(z_k) + i\mathrm{Im}(z_{k-1}), z_k]$ of two segments which are parallel to the real and imaginary axis, respectively. □

55.3 Winding Numbers

We now introduce a geometric characteristic of loops which describes how many times they "wind around" some point in the plane.

Lemma 55.3. *Let $\gamma\colon I \to \mathbb{C}\backslash\{0\}$ be a path. Then there exist continuous functions $a\colon I \to \mathbb{R}$ and $r\colon I \to \mathbb{R}_+$ such that*

$$\gamma(t) = r(t)e^{ia(t)} \tag{139}$$

for all $t \in I$.

Proof. The function $r(t) := |\gamma(t)|$ is continuous and positive. So the proof amounts to finding an appropriate argument $a(t)$ of $\gamma(t)$ such that $t \mapsto a(t)$ is continuous. For this purpose, we use the path covering lemma with $D := \mathbb{C}\backslash\{0\}$.

At the initial point of γ we choose the principal branch of the argument, $a(0) := \operatorname{Arg}(\gamma(0))$. If $t \in [t_0, t_1]$, all points $\gamma(t)$ lie in the disk D_0. Since D_0 does not contain the origin, it is contained in a sector with vertex at 0 and opening angle less than π. Consequently the argument $a(t) = \arg(\gamma(t))$ can be chosen such that $|a(t) - a(0)| < \pi/2$, which yields a continuous function a on $[0, t_1]$.

Suppose that such a function has already been constructed on some interval $[0, t_k]$. Then it can be prolongated to $[0, t_{k+1}]$ by choosing $a(t) = \arg(\gamma(t))$ on $[t_k, t_{k+1}]$ such that $|a(t) - a(t_k)| < \pi/2$, which is possible since $\gamma(t) \in D_k$ and $0 \notin D_k$. By induction, a can be extended to all of I. □

Any continuous function a satisfying (139) is called a **continuous branch** of the argument along the path γ. The difference of two such functions a_1 and a_2 on I is a constant integral multiple of 2π. If a is continuous branch of the argument along a loop, then $a(1) - a(0)$ is an integral multiple of 2π which does not depend on the special choice of the branch a.

Definition of Winding Numbers

Let γ be a loop in $\mathbb{C}\backslash\{0\}$ and denote by a a continuous branch of the argument along γ. Then the integer

$$\operatorname{wind}(\gamma) := \frac{1}{2\pi}(a(1) - a(0))$$

is called the **winding number** (or **index**) **of** γ. If $z_0 \in \mathbb{C}$ and γ is a loop in $\mathbb{C}\backslash\{z_0\}$, the **winding number of** γ **about** z_0 is defined by

$$\operatorname{wind}(\gamma, z_0) := \operatorname{wind}(\gamma - z_0).$$

Stability of Winding Numbers

Next we prove the intuitive fact that small perturbations of a loop do not change its winding number. Recall that the distance between two paths γ and γ_0 is defined in terms of the uniform norm:

$$\|\gamma - \gamma_0\|_I := \max_{t\in I} |\gamma(t) - \gamma_0(t)|.$$

Lemma 55.4. *Let $\gamma_0\colon I \to \mathbb{C}\backslash\{0\}$ be a loop, and denote by d the distance of its trace $[\gamma_0]$ from the origin. Then for all loops $\gamma\colon I \to \mathbb{C}$ with $\|\gamma - \gamma_0\|_I < d$,*

$$\mathit{wind}(\gamma) = \mathit{wind}(\gamma_0).$$

Proof. Since $[\gamma_0]$ is a compact subset of $\mathbb{C}\backslash\{0\}$, its distance from the origin is positive. Then $|\gamma(t) - \gamma_0(t)| < d$ implies that $\gamma(t)/\gamma_0(t)$ lies in the right half-plane.

Let a_0 be a continuous branch of the argument along γ_0. If we choose a continuous branch of the argument along γ such that $|a(0) - a_0(0)| < \pi/2$, then $|a(t) - a_0(t)| < \pi/2$ for all $t \in I$. Invoking the triangle inequality we see that

$$|(a(1) - a(0)) - (a_0(1) - a_0(0))| < \pi,$$

and since this number is an integral multiple of 2π, it must be zero. □

Lemma 55.5. *Let $D \subseteq \mathbb{C}$ be a simply connected domain and $z_0 \in D$. Then two loops γ_0 and γ_1 are homotopic in the punctured domain $D\backslash\{z_0\}$ if and only if they have the same winding number about z_0.*

56 Cauchy's Theorem and its Applications

Roughly speaking, Cauchy's Theorem states that if f is holomorphic in an open set Ω and γ is a whose interior is also contained in Ω, then

$$\int_\Gamma f(z)dz = 0.$$

Many results that follow, and in particular the calculus of residues, are related in one way or another to this fact. We first prove this in the special case that our curve Γ is a triangle:

56.1 Goursat's Theorem

Theorem 56.1. *(Goursat's Theorem) If Ω is an open set in $\mathbb{C}$, and $T \subset \Omega$ is a triangle whose interior is also contained in Ω, then*

$$\int_T f(z)dz = 0,$$

whenever f is holomorphic in Ω.

Proof. We call $T^{(0)}$ our original triangle (with a fixed orientation which we choose to be positive), and let $d^{(0)}$ and $p^{(0)}$ denote the diameter and perimeter of $T^{(0)}$, respectively. The first step in our construction consists of bisecting each side of the triangle and connected the midpoints. This creates four new smaller triangles, denote $T_1^{(1)}$, $T_2^{(1)}$, $T_3^{(1)}$, and $T_4^{(1)}$ that are similar to the original triangle. The orientation is chosen to be consistent with that of the original triangle, and so after cancellations arising from integration over the same side in two opposite directions, we have

$$\int_{T^{(0)}} f(z)dz = \int_{T_1^{(1)}} f(z)dz + \int_{T_2^{(1)}} f(z)dz + \int_{T_3^{(1)}} f(z)dz + \int_{T_4^{(1)}} f(z)dz.$$

By triangle inequality, we must have

$$\left|\int_{T^{(0)}} f(z)dz\right| \leq 4 \left|\int_{T_j^{(1)}} f(z)dz\right|$$

for some $j \in \{1, 2, 3, 4\}$. Without loss of generality, assume $j = 1$. Observe that if $d^{(1)}$ and $p^{(1)}$ denote the diameter and perimeter of $T^{(1)}$, respectively, then $d^{(1)} = (1/2)d^{(0)}$ and $p^{(1)} = (1/2)p^{(0)}$. We now repeat this process for the triangle $T^{(1)}$, bisecting into four smaller triangles. Continuing this process, we obtain a sequence of triangles

$$T^{(0)}, T^{(1)}, \dots, T^{(n)}, \dots$$

with the properties that

$$\left|\int_{T^{(0)}} f(z)dz\right| \leq 4^n \left|\int_{T^{(n)}} f(z)dz\right|$$

and

$$d^{(n)} = 2^{-n}d^{(0)}, \quad p^{(n)} = 2^{-n}p^{(0)},$$

where $d^{(n)}$ and $p^{(n)}$ denote the diameter and perimeter of $T^{(n)}$, respectively. We also denote $\mathcal{T}^{(n)}$ the *solid* closed triangle with boundary $T^{(n)}$, and observe that our construction yields a sequence of nested compact sets

$$\mathcal{T}^{(0)} \supset \mathcal{T}^{(1)} \supset \cdots \supset \mathcal{T}^{(n)} \supset \cdots$$

whose diameter goes to 0. Thus, by Proposition (**??**), there exists a unique point z_0 that belongs to all the solid triangles $\mathcal{T}^{(n)}$. Since f is holomorphic at z_0, we can write

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + \psi(z)(z - z_0),$$

where $\psi(z) \to 0$ as $z \to z_0$. Since the constant $f(z_0)$ and the linear function $f'(z_0)(z - z_0)$ have primitives, we can integrate the above equality and, using Corollary (**??**), we obtain

$$\int_{T^{(n)}} f(z)dz = \int_{T^{(n)}} \psi(z)(z - z_0)dz.$$

Now z_0 belongs to the closure of the triangle $\mathcal{T}^{(n)}$ and z to its boundary, so we must have $|z - z_0| \leq d^{(n)}$, and so we estimate

$$\left| \int_{T^{(n)}} f(z) dz \right| \leq \varepsilon_n d^{(n)} p^{(n)},$$

where $\varepsilon_n = \|\psi\|_{z \in T^{(n)}} \to 0$ as $n \to \infty$. Therefore

$$\left| \int_{T^{(n)}} f(z) dz \right| \leq \varepsilon_n 4^{-n} d^{(0)} p^{(0)},$$

which yields our final estimate

$$\left| \int_{T^{(0)}} f(z) dz \right| \leq 4^n \left| \int_{T^{(n)}} f(z) dz \right| \leq \varepsilon_n d^{(0)} p^{(0)}.$$

Letting $n \to \infty$ concludes the proof since $\varepsilon_n \to 0$. □

Corollary 23. *If f is holomorphic in an open set Ω that contains a rectangle R and its interior, then*

$$\int_R f(z) dz = 0.$$

Proof. We simply decompose the rectangle into two triangles, and integrate over the two triangles. □

56.2 Local existence of primitives and Cauchy's theorem in a disc

Theorem 56.2. *A holomorphic function in an open disc has a primitive in that disc.*

Proof. After a translation, we may assume without loss of generality that the disc, say D, is centered at the origin. Given a point $z \in D$, consider the piecewise-smooth curve that joins 0 to z first by moving in the horizontal direction form 0 to $\mathrm{Re}(z)$, and then in the vertical direction from $\mathrm{Re}(z)$ to z. We choose the orientation from 0 to z, and denote this polygonal line by Γ_z. Define

$$F(z) = \int_{\Gamma_z} f(w) dw.$$

The choice of Γ_z gives an unambigious definition of the function F. We contend that F is holomorphic in D and that $F'(z) = f(z)$ for all $z \in D$. To prove this, fix $z \in D$ and let $h \in \mathbb{C}$ be sufficiently small so that $z + h$ also belongs to the disc. Now consider the difference

$$F(z+h) - F(z) = \int_{\Gamma_{z+h}} f(w) dw - \int_{\Gamma_z} f(w) dw.$$

The function f is first integrated along Γ_{z+h} with the original orientation, and then along Γ_z with the reverse orientation. Since we integrate f over the line segment starting at the origin in two opposite directions, it cancels. Then we complete the square and triangle, so that after an application of Goursat's theorem for triangles and rectangles, we are left with the line segment from z to $z + h$. Hence, we have

$$F(z+h) - F(z) = \int_\eta f(w) dw,$$

where η is the straight line segment from z to $z + h$. Since f is continuous at z we can write

$$f(w) = f(z) + \psi(w)$$

where $\psi(w) := f(w) - f(z) \to 0$ as $w \to z$. Therefore

$$F(z+h) - F(z) = \int_\eta f(z) dw + \int_\eta \psi(w) dw = f(z) \int_\eta dw + \int_\eta \psi(w) dw.$$

On the one hand, the constant 1 has w as a primitive, so the first integral is simply h. On the other hand, we have

$$\left| \int_\eta \psi(w) dw \right| \leq \sup_{w \in \eta} |\psi(w)| \cdot |h|.$$

Since the supremum above goes to 0 as h tends to 0, we conclude that

$$\lim_{h \to 0} \frac{F(z+h) - F(z)}{h} = f(z),$$

thereby proving that F is a primitive for f on the disc. $\square$

Theorem 56.3. *If f is holomorphic in a disc, then*

$$\int_{\Gamma} f(z) dz = 0$$

for any closed curve Γ in that disc.

Proof. Since f has a primitive, we can apply Corollary (**??**). $\square$

Example 56.1. We show that if $y \in \mathbb{R}$, then

$$e^{-\pi y^2} = \int_{\mathbb{R}} e^{-\pi x^2} e^{-2\pi i x y} dx.$$

This gives a new proof of the fact that $e^{-\pi x^2}$ is its own Fourier transform. If $y = 0$, then the equality becomes

$$1 = \int_{\mathbb{R}} e^{-\pi x^2} dx.$$

Now suppose that $y > 0$ and consider the function $f(z) = e^{-\pi z^2}$, which is entire, and in particular holomorphic in the interior of the toy contour Γ_R, where Γ_r consists of a rectangle with vertices R, $R + iy$, $-R + iy$, $-R$ and the positive counterclockwise orientation. By Cauchy's theorem,

$$\int_{\Gamma_R} f(z) dz = 0.$$

The integral over the real segment is simply

$$\int_{-R}^{R} e^{-\pi x^2} dx,$$

which converges to 1 as $R \to \infty$. The integral on the vertical side on the right is

$$I(R) = \int_0^y f(R + it) i dt = \int_0^y e^{-\pi(R^2 + 2iRt - t^2)} i dt.$$

This integral goes to 0 as $R \to \infty$ since y is fixed and we may estimate it by

$$|I(R)| \leq C e^{-\pi R^2}.$$

Similarly, the integral over the vertical segment on the left also goes to 0 as $R \to \infty$ for the same reasons. Finally, the integral over the horizontal segment on top is

$$\int_{R}^{-R} e^{-\pi(x+iy)^2} dx = e^{-\pi x^2} \int_{-R}^{R} e^{-\pi x^2} e^{-2\pi i x y} dx.$$

Therefore we find in the limit as $R \to \infty$ that

$$0 = 1 - e^{-\pi x^2} \int_{-R}^{R} e^{-\pi x^2} e^{-2\pi i x y} dx,$$

and our desired formula is established. In the case $y < 0$, we then consider the symmetric rectangle in the lower half-plane.

56.3 Differentiable and Analytic Functions

Lemma 56.4. *Let $f\colon B_R(z_0) \to \mathbb{C}$ be a holomorphic function and let $r > 0$ such that $0 < r < R$. If F is a primitive of f, then for all points $z \in B_r(z_0)$, we have*

$$F(z) = \frac{1}{2\pi i} \int_{\gamma_r(z_0)} \frac{F(w)}{w - z} dw.$$

Proof. 1. We begin with an auxiliary result. Let $z_0 \in B_R(a)$ be fixed. Define a function $\varphi\colon B_R(a - z_0) \to \mathbb{C}$ by the formula

$$\varphi(h) := F(z_0 + h) - F(z_0) - f(z_0)h - \frac{1}{2} f'(z_0)h^2$$

for all $h \in B_R(a - z_0)$. The function φ is differentiable in $B_R(a - z_0)$ and its derivative with respect to h can be computed by the chain rule,

$$\begin{aligned} \varphi'(h) &= F'(z_0 + h) - f(z_0) - f'(z_0)h \\ &= f(z_0 + h) - f(z_0) - f'(z_0)h. \end{aligned}$$

Since f is differentiable at z_0, the right-hand side is of order $o(h)$ as $h \to 0$, that is, for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $|\varphi'(h)| \le \varepsilon |h|$ whenever $|h| < \delta$.

The function φ' is continuous, whence the mapping $I \to \mathbb{C}$ given by $t \mapsto \varphi(th)$ is continuously differentiable (with respect to the real variable t), so that $\varphi(h)$ can be represented by the fundamental theorem of calculus,

$$\varphi(h) = \int_0^1 \frac{d}{dt}(\varphi(th))dt = \int_0^1 \varphi'(th)h dt.$$

Using the standard estimate for integrals in combination with $|\varphi'(h)| \le \varepsilon |h|$, we conclude that $|\varphi(h)| \le h^2 \varepsilon$ for all h with $|h| < \delta$. Since ε can be chosen arbitrarily small, we have

$$\lim_{h \to 0} \varphi(h)/h^2 = 0. \tag{140}$$

2. The function G defined by

$$G(z) := \begin{cases} \frac{F(z) - F(z_0)}{z - z_0} & \text{if } z \in B_R(a) \backslash \{z_0\} \\ f(z_0) & \text{if } z = z_0 \end{cases}$$

is differentiable in $z \in B_R(a) \backslash \{z_0\}$. In order to prove that G is also differentiable at z_0, we consider its difference quotient at z_0. By definition of G and φ, we have

$$\begin{aligned} \frac{G(z) - G(z_0)}{z - z_0} &= \frac{(F(z) - F(z_0) - (z - z_0)f(z_0))}{(z - z_0)^2} \\ &= \frac{1}{2} f'(z_0) + \frac{\varphi(z - z_0)}{(z - z_0)^2}, \end{aligned}$$

and using (140), we find $G'(z_0) = (1/2)f'(z_0)$.

3. Because G is differentiable in the disk $B_R(a)$, we can apply Goursat's lemma, which tells us that the integral of G along the closed path $\gamma_r(a)$ vanishes. Hence

$$\begin{aligned} 0 &= \int_{\gamma_r(a)} G(z)dz \\ &= \int_{\gamma_r(a)} \frac{F(z) - F(z_0)}{z - z_0} dz \\ &= \int_{\gamma_r(a)} \frac{F(z)}{z - z_0} dz - F(z_0) \int_\gamma \frac{dz}{z - z_0} \\ &= \int_{\gamma_r(a)} \frac{F(z)}{z - z_0} dz - F(z_0) \cdot 2\pi i \cdot \operatorname{wind}(\gamma, z_0). \end{aligned}$$

□

56.3.1 Cauchy Integrals

Definition 56.1. Let γ be a piecewise smooth path in $\mathbb{C}$ and assume that the function $\varphi\colon [\gamma] \to \mathbb{C}$ is continuous. The function $f\colon \mathbb{C}\backslash[\gamma] \to \mathbb{C}$ defined by

$$f(z) := \frac{1}{2\pi i}\int_\gamma \frac{\varphi(w)}{w-z}dw \tag{141}$$

for all $z \in \mathbb{C}\backslash[\gamma]$ is said to be the **Cauchy integral with density φ along γ.**

Theorem 56.5. *Let γ be a piecewise smooth path in $\mathbb{C}$ and assume that $\varphi\colon [\gamma] \to \mathbb{C}$ is continuous. Then the function f defined by the Cauchy integral (141) is analytic on $D := \mathbb{C}\backslash[\gamma]$ and tends to zero as $z \to \infty$. For any disk $D_0 \subset D$ with center z_0, the taylor series*

$$f(z) = \sum_{n=0}^{\infty} a_n (z-z_0)^n$$

of f at z_0 converges in D_0 and its coefficients satisfy

$$a_n = \frac{1}{2\pi i}\int_\gamma \frac{\varphi(z)}{(z-z_0)^{n+1}}dz.$$

Proof. Fix $z \in D_0$. Invoking a compactness argument, we know of the existence of a constant $q < 1$ such that

$$\left|\frac{z-z_0}{w-z_0}\right| \le q < 1$$

for all $w \in [\gamma]$. Consequently,

$$\begin{aligned}\frac{\varphi(w)}{w-z} &= \frac{\varphi(w)}{(w-z_0)-(z-z_0)}\\ &= \sum_{n=0}^{\infty}\frac{\varphi(w)}{w-z_0}\left(\frac{z-z_0}{w-z_0}\right)^n.\end{aligned}$$

The function $w \mapsto \varphi(w)/(w-z_0)$ is bounded on the compact set $[\gamma]$, the series converges uniformly with respect to $w \in [\gamma]$ (apply Weierstrass M-test with $M_n = Mq^n$, where M is a bound for the function $w \mapsto \varphi(w)/(w-z_0)$). Interchanging the order of summation and integration, we obtain

$$\begin{aligned}2\pi i f(z) &= \int_\gamma \frac{\varphi(w)}{w-z}dw\\ &= \sum_{n=0}^{\infty}\left(\int_\gamma \frac{\varphi(w)}{(w-z_0)^{n+1}}dw\right)(z-z_0)^n\end{aligned}$$

for all $z \in D_0$, which proves the claim. Finally, the standard integral estimate yields that $f(z) \to 0$ as $z \to \infty$. □

56.4 Cauchy's Integral Formula

Theorem 56.6. *(Cauchy's Integral Formula) Let Ω be an open set, let $f\colon \Omega \to \mathbb{C}$ be a holomorphic function, let $a \in \Omega$, and let $r > 0$ such that $\overline{B}_r(a) \subset \Omega$. For every $z \in B_r(a)$, we have*

$$f(z) = \frac{1}{2\pi i}\int_{\gamma_r(a)} \frac{f(w)}{w-z}dw.$$

To get a feel for how this theorem works, let us assume that f is analytic at a. Then we can choose $\varepsilon > 0$ such that $\overline{B}_\varepsilon(a) \subset B_r(z_0)$ and

$$f(z) = \sum_{n=0}^{\infty} a_n(z-a)$$

for all $z \in \overline{B}_\varepsilon(a)$. Then we'd have

$$\begin{aligned}\int_{\gamma_r(z_0)} \frac{f(z)}{z-a}dz &= \int_{\gamma_\varepsilon(a)} \frac{f(z)}{z-a}dz \\ &= \int_{\gamma_\varepsilon(a)} \sum_{n=0}^{\infty} a_n (z-a)^{n-1} dz \\ &= \sum_{n=0}^{\infty} a_n \int_{\gamma_\varepsilon(a)} (z-a)^{n-1} dz \\ &= 2\pi i f(a).\end{aligned}$$

Theorem 56.7. *(Cauchy's Integral Formula) Let Ω be an open subset of $\mathbb{C}$, let $f\colon \Omega \to \mathbb{C}$ be a holomorphic function, let $a \in \Omega$, and let $r > 0$ such that $C_r(a) \subset \Omega$. For every $z \in B_r(a)$, we have*

$$f(z) = \frac{1}{2\pi i} \int_{\gamma_r(a)} \frac{f(w)}{w-z} dw.$$

Proof. Let $z \in B_r(a)$. Define $\gamma\colon I \to \Omega$ to be the path

$$\gamma := \oplus e^{2\pi i \operatorname{Arg}(z-a)} \gamma_r(a)$$

$$\zeta$$

If $a \neq z$, let $\zeta = \min\{w - z \mid w \in C_r(a)\}$, be the and let $\gamma_{\delta,\varepsilon}$ be the loop in Ω be given by

$$\gamma_{\delta,\varepsilon} = -$$

and consider the "keyhole" $\gamma_{\delta,\varepsilon}$ which omits the point z. Here δ is the width of the corridor, and ε is the radius of the small circle centered at z. Since the function $f(w)/(w-z)$ is holomorphic away from the point $w = z$, we have

$$\int_{\Gamma_{\delta,\varepsilon}} \frac{f(w)}{w-z} dw = 0$$

by Cauchy's theorem for the chosen toy contour. Now we make the corridor narrower by letting δ tend to 0, and use the continuity of $f(w)/(w-z)$ to see that in the limit, the integrals over the two sides of the corridor cancel out. The remaining part consists of two curves, the large boundary circle C with the positive orientation, and a small circle C_ε centered at z of radius ε and oriented negatively, that is, clockwise. To see what happens to the integral over the small circle, we write

$$\frac{f(w)}{w-z} = \frac{f(w) - f(z)}{w-z} + \frac{f(z)}{w-z}$$

and note that since f is holomorphic, the first term on the right-hand is bounded so that its integral over C_ε tends to 0 as $\varepsilon \to 0$. To conclude the proof, it suffices to observe that

$$\begin{aligned}\int_{C_\varepsilon} \frac{f(z)}{w-z} dw &= f(z) \int_{C_\varepsilon} \frac{dw}{w-z} \\ &= -f(z) \int_0^{2\pi} \frac{\varepsilon i e^{-it}}{\varepsilon e^{-it}} dt \\ &= -f(z) 2\pi i\end{aligned}$$

so that in the limit we find

$$0 = \int_C \frac{f(w)}{w-z} dw - 2\pi i f(z),$$

as was to be shown. □

Integral Representation of the Taylor Coefficients

Theorem 56.8. *Let f be the sum of the power series $\sum a_n(z-a)^n$ with radius of convergence R. Then for any $n \geq 0$ and r such that $0 < r < R$, we have*

$$a_m = \frac{1}{r^m}\int_0^1 f(a+re^{2\pi i t})e^{-2\pi i m t}dt$$

Proof. By uniform convergence of the power series $\sum a_n(z-a)^n$ on $C_r(a)$, we have

$$\begin{aligned}
\int_0^1 f(a+re^{2\pi i t})e^{-2\pi i m t}dt &= \int_0^1 \sum_{n=0}^{\infty} a_n r^n e^{2\pi i(n-m)t}dt \\
&= \sum_{n=0}^{\infty} a_n r^n \int_0^1 e^{2\pi i(n-m)t}dt \\
&= a_m r^m.
\end{aligned}$$

□

Corollary 24. *If f is holomorphic in an open set Ω, then f has infinitely many complex derivatives in Ω. Moreover, if $C \subset \Omega$ is a circle whose interior is also contained in Ω, then*

$$f^{(n)}(z) = \frac{n!}{2\pi i}\int_C \frac{f(w)}{(w-z)^{n+1}}dw$$

for all z in the interior of C.

Proof. Let f be the sum of the power series $\sum a_n(z-a)^n$ with radius of convergence R. Then

$$\begin{aligned}
f^{(m)}(a) &= m! a_m \\
&= \frac{m!}{r^m}\int_0^1 f(a+re^{2\pi i t})e^{-2\pi i m t}dt \\
\\
&= \frac{m!}{2\pi i}\int_0^1 \frac{f(z)}{(re^{2\pi i t})^{n+1}}2\pi i r e^{2\pi i t}dz \\
&= \frac{m!}{2\pi i}\int_{\gamma_r(a)} \frac{f(z)}{(z-a)^{n+1}}dz
\end{aligned}$$

□

Corollary 25. *If f is holomorphic in an open set Ω, then f has infinitely many complex derivatives in Ω. Moreover, if $C \subset \Omega$ is a circle whose interior is also contained in Ω, then*

$$f^{(n)}(z) = \frac{n!}{2\pi i}\int_C \frac{f(w)}{(w-z)^{n+1}}dw$$

for all z in the interior of C.

Proof. The proof is by induction on n, the case $n = 0$ being simply the Cauchy integral formula. Suppose that f has up to $n-1$ complex derivatives and that

$$f^{(n-1)}(z) = \frac{(n-1)!}{2\pi i}\int_C \frac{f(w)}{(w-z)^n}dw.$$

Now for small h, the difference quotient for $f^{(n-1)}$ takes the form

$$\begin{aligned}
\frac{f^{(n-1)}(z+h) - f^{(n-1)}(z)}{h} &= \frac{(n-1)!}{2\pi i}\int_C \frac{f(w)}{h}\left(\left(\frac{1}{w-(z+h)}\right)^n - \left(\frac{1}{w-z}\right)^n\right)dw \\
&= \frac{(n-1)!}{2\pi i}\int_C \frac{f(w)}{h}\left(\frac{1}{(w-(z+h))} - \frac{1}{(w-z)}\right)\left(\sum_{m=0}^{n-1}\left(\frac{1}{w-(z+h)}\right)^{n-m-1}\left(\frac{1}{w-z}\right)^m\right)dw \\
&= \frac{(n-1)!}{2\pi i}\int_C f(w)\left(\frac{1}{(w-(z+h))(w-z)}\right)\left(\sum_{m=0}^{n-1}\left(\frac{1}{w-(z+h)}\right)^{n-m-1}\left(\frac{1}{w-z}\right)^m\right)dw
\end{aligned}$$

Now observe that if h is small, then $z+h$ and z stay at a finite distance from the boundary circle C, so in the limit as h tends to 0, we find that the quotient converges to

$$\frac{(n-1)!}{2\pi i}\int_C f(w)\left(\frac{1}{(w-z)^2}\right)\left(\frac{n}{(w-z)^{n-1}}\right)dw = \frac{n!}{2\pi i}\int_C \frac{f(w)}{(w-z)^{n+1}}dw,$$

which completes the induction argument and proves the theorem. $\square$

56.4.1 Taylor's Theorem

Theorem 56.9. *Let Ω be an open set, let $f\colon \Omega \to \mathbb{C}$ be a holomorphic function, and let $a \in \Omega$. Then there exists $r > 0$ and a power series $\sum a_n(z-a)^n$ centered at a such that $\overline{B}_r(a) \subset \Omega$*

$$f(z) = \sum_{n=0}^{\infty} a_n(z-a)^n$$

for all $z \in \overline{B}_r(a)$. Furthermore, the coefficients are given by

$$a_n = \frac{1}{2\pi i}\int_{\gamma_r(a)} \frac{f(w)}{(w-a)^{n+1}}dw$$

for all $n \geq 0$.

Proof. Let $z \in B_r(a)$. Then

$$\begin{aligned} f(z) &= \frac{1}{2\pi i}\int_{\gamma_r(a)} \frac{f(w)}{w-z}dw \\ &= \frac{1}{2\pi i}\int_{\gamma_r(a)} \frac{f(w)}{w-a}\left(\frac{1}{1-\left(\frac{z-a}{w-a}\right)}\right)dw \\ &= \frac{1}{2\pi i}\int_{\gamma_r(a)} \frac{f(w)}{w-a}\sum_{n=0}^{\infty}\left(\frac{z-a}{w-a}\right)^n dw \\ &= \sum_{n=0}^{\infty}\left(\frac{1}{2\pi i}\int_{\gamma_r(a)} \frac{f(w)}{(w-a)^{n+1}}dw\right)(z-a)^n. \\ &= \sum_{n=0}^{\infty} a_n(z-a)^n. \end{aligned}$$

where we are allowed to interchange the integral with the sum since the series $\sum_{n=0}^{\infty}\left(\frac{z-a}{w-a}\right)^n$ converges uniformly in $w \in C_r(a)$. $\square$

56.4.2 Limit of Holomorphic Functions Converging Uniformly on Compact Subsets is Holomorphic

Theorem 56.10. *Let Ω be a nonempty open subset of $\mathbb{C}$ and let (f_n) be a sequence of analytic functions on Ω that converges uniformly to f on each compact subset of Ω. Then f is holomorphic and (f_n') converges uniformly to f' on each compact subset.*

Proof. Let $a \in \Omega$. Choose $z_0 \in \Omega$ and $r > 0$ such that $a \in \overline{B}_r(z_0) \subset \Omega$. Then

$$f_n(a) = \frac{1}{2\pi i}\int_{\gamma_r(z_0)} \frac{f_n(z)}{z-a}dz$$

for all $n \in \mathbb{N}$. Since (f_n) converges to f uniformly on $\overline{B}_r(z_0)$, the function f is continuous on $\overline{B}_r(z_0)$, and so $f(z)/(z-a)$ is integrable along $\gamma_r(z_0)$. Thus

$$\begin{aligned} \left|\frac{1}{2\pi i}\int_{\gamma_r(z_0)} \frac{f_n(z)}{z-a}dz - \frac{1}{2\pi i}\int_{\gamma_r(z_0)} \frac{f(z)}{z-a}dz\right| &= \left|\frac{1}{2\pi i}\int_{\gamma_r(z_0)} \frac{f_n(z)-f(z)}{z-a}dz\right| \\ &\leq \frac{1}{2\pi}\|f_n - f\|_{\overline{B}_r(z_0)}\left|\int_{\gamma_r(z_0)} \frac{dz}{z-a}\right| \\ &= \|f_n - f\|_{\overline{B}_r(z_0)}, \end{aligned}$$

which tends to 0 as n tends to ∞. Therefore

$$\begin{aligned}\frac{1}{2\pi i}\int_{\gamma_r(z_0)}\frac{f(z)}{z-a}dz &= \lim_{n\to\infty}\frac{1}{2\pi i}\int_{\gamma_r(z_0)}\frac{f_n(z)}{z-a}dz\\ &= \lim_{n\to\infty} f_n(a)\\ &= f(a).\end{aligned}$$

This implies f is holomorphic in Ω.

To show $f_n' \to f'$ uniformly on compact subsets of Ω, it suffices to work with closed discs. Let $\overline{D}$ be a closed disc in Ω with radius $r > 0$. Choose a in the interior of $\overline{D}$. Then

$$f_n'(a) = \frac{1}{2\pi i}\int_\gamma \frac{f_n(z)}{(z-a)^2}dz, \quad \text{and} \quad f'(a) = \frac{1}{2\pi i}\int_\gamma \frac{f(z)}{(z-a)^2}dz.$$

Therefore

$$\begin{aligned}|f_n'(a) - f'(a)| &\le \frac{\|f_n - f\|_{\overline{D}}}{r}\\ &\to 0.\end{aligned}$$

□

56.4.3 Cauchy's Inequalities

Corollary 26. *(Cauchy's inequality) Let f is holomorphic in a given set that contains the closure of a disc D centered at z_0 and of radius R, then*

$$\left|f^{(n)}(z_0)\right| \le \frac{n!}{R^n}\sup_{z\in C}|f(z)|.$$

Proof. Applying Cauchy's Integral Formula for $f^{(n)}(z_0)$, we have

$$\begin{aligned}\left|f^{(n)}(z_0)\right| &= \left|\frac{n!}{2\pi i}\int_C \frac{f(\zeta)}{(\zeta - z)^{n+1}}d\zeta\right|\\ &= \frac{n!}{2\pi}\left|\int_0^{2\pi}\frac{f(z_0+Re^{it})}{R^n e^{int}}d\zeta\right|\\ &\le \frac{n!}{R^n}\sup_{z\in C}|f(z)|.\end{aligned}$$

□

56.4.4 Louiville's Theorem

Theorem 56.11. *(Louiville's Theorem) Every bounded entire function must be constant.*

Proof. Let f be a bounded entire function. Suppose $\sum a_n z^n$ is the power series representation of f at 0 and choose a constant M such that $|f(z)| \le M$ for all $z \in \mathbb{C}$. Then for every $r > 0$ and $n \ge 1$, we have

$$\begin{aligned}|a_n| &= \left|\frac{1}{2\pi i}\int_{\gamma_r(0)}\frac{f(z)}{z^{n+1}}dz\right|\\ &= \left|\int_0^1 \frac{f(re^{2\pi i t})}{r^n e^{2\pi i n t}}dt\right|\\ &\le \int_0^1 \left|\frac{f(re^{2\pi i t})}{r^n e^{2\pi i n t}}\right| dt\\ &\le \frac{M}{r^n}.\end{aligned}$$

This implies $a_n = 0$ for every $n \ge 1$. Thus $f(z) = a_0$, which proves the theorem. □

56.4.5 Fundamental Theorem of Algebra

Corollary 27. *Every non-constant polynomial $P(z) = a_n z^n + \cdots + a_1 z + a_0$ with complex coefficients has a root in $\mathbb{C}$.*

Proof. If $P(z)$ has no roots, then $Q(z) := 1/P(z)$ is a bounded holomorphic function. To see this, we can of course assume that $a_n \neq 0$ and write

$$Q(z) = \frac{1}{a_n z^n + \cdots + a_1 z + a_0} = \left(\frac{1}{z^n}\right)\left(\frac{1}{\frac{a_0}{z^n} + \frac{a_1}{z^{n-1}} + \cdots + a_n}\right).$$

As $z \to \infty$, the denominator of the second term in the round brackets converges to $a_n \neq 0$, hence the second term itself goes to $1/a_n$. But the first term tends to zero, hence

$$\lim_{z\to\infty} Q(z) = 0.$$

In particular, $|Q(z)|$ is bounded by 1 outside of some circle $|z| = r$. Inside this circle, $|Q(z)|$ is continuous, hence bounded. Thus $|Q(z)|$, and therefore $Q(z)$ itself is bounded on the whole complex plane. By Liousville's theorem, we then conclude that $Q(z)$ is constant. This contradicts our assumption that $P(z)$ is nonconstant and proves the corollary. □

Theorem 56.12. *(Identity Theorem) Let f, g be holomorphic functions on a connected open set D of $\mathbb{C}$. If $f = g$ on a nonempty open subset of D, then $f = g$ on D.*

Remark 67. This says that a holomorphic function is completely determined by its values on a (possibly quite small) neighborhood in D. This is not true for real-differentiable functions. In comparison, holomorphy is a much more rigid notion.

Proof. Let S be the set of all $z \in D$ such that $f(z) = g(z)$. We show that S is open and closed, and hence must be D. Since $f - g$ is continuous, and $S = (f-g)^{-1}\{0\}$, we see that S is closed. To show, that S is open, suppose w lies in S. Then, because the Taylor series of f and g at w have non-zero radius of convergence, the open disk $B_r(w)$ also lies in S for some r. □

56.5 Further Applications

56.5.1 Morera's Theorem

Theorem 56.13. *Suppose f is a continuous function in the open disc D such that for any triangle T cotained in D,*

$$\int_T f(z)dz = 0,$$

then f is holomorphic.

Proof. By the proof of Theorem (56.2), the function f has a primitive F in D that satisfies $F' = f$. By the regularity theorem, we know that F is indefinitely (and hence twice) complex differentiable, and therefore f is holomorphic. □

Theorem 56.14. *If $\{f_n\}_{n=1}^{\infty}$ is a sequence of holomorphic functions that converges uniformly to a function f in every compact subset of Ω, then f is holomorphic in Ω.*

Proof. Let D be any disc whose closure is contained in Ω and T any triangle in that disc. Then, since each f_n is holomorphic, Goursat's theorem implies

$$\int_T f_n(z)dz = 0$$

for all n. By assumption, $f_n \to f$ uniformly in the closure of D, so f is continuous and

$$\int_T f_n(z)dz \to \int_T f(z)dz.$$

As a result, we find $\int_T f(z)dz = 0$ and by Morera's theorem, we conclude that f is holomorphic in D. Since this conclusion is true for every D whose closure is contained in Ω, we find that f is holomorphic in all of Ω. □

56.5.2 Sequence of Holomorphic Functions

Theorem 56.15. *Let $F(z,s)$ be defined for $(z,s) \in \Omega \times [0,1]$ where Ω is an open set in $\mathbb{C}$. Suppose F satisfies the following properties:*

1. *$F(z,s)$ is holomorphic in z for each s.*
2. *F is continuous on $\Omega \times [0,1]$.*

Then the function f defined on Ω by

$$f(z) = \int_0^1 F(z,s)ds$$

is holomorphic.

Remark 68. The second condition says that F is jointly continuous in both arguments. To prove this result, it suffices to prove that f is holomorphic in any disc D contained in Ω, and by Morera's theorem this could be achieved by showing that for any triangle T contained in D we have

$$\int_T \int_0^1 F(z,s)dsdz = 0.$$

Interchanging the order of integration, and using property (1) would then yield the desired result. We can, however, get around the issue of justifying the change in the order of integration by arguing differently. The idea is to interpret the integral as a "uniform" limit of Riemann sums, and then apply the results of the previous section.

Proof. For each $n \geq 1$, we consider the Riemann sum

$$f_n(z) = (1/n) \sum_{k=1}^{n} F(z, k/n).$$

Then f_n is holomorphic in all of Ω by property (1), and we claim that on any disc D whose closure is contained in Ω, the sequence $\{f_n\}_{n=1}^{\infty}$ converges uniformly to f. To see this, we recall that a continuous function on a compact set is uniformly continuous, so if $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\sup_{z \in D} |F(z,s_1) - F(z,s_2)| < \varepsilon \quad \text{whenever } |s_1 - s_2| < \delta.$$

Then if $n > 1/\delta$, and $z \in D$ we have

$$\begin{aligned}
|f_n(z) - f(z)| &= \left| \sum_{k=1}^{n} \int_{(k-1)/n}^{k/n} F(z,k/n) - F(z,s)ds \right| \\
&\leq \sum_{k=1}^{n} \int_{(k-1)/n}^{k/n} |F(z,k/n) - F(z,s)|\, ds \\
&< \sum_{k=1}^{n} \frac{\varepsilon}{n} \\
&< \varepsilon.
\end{aligned}$$

This proves the claim, and by Theorem (56.14), we conclude that f is holomorphic in D. As a consequence, f is holomorphic in Ω, as was to be shown. □

56.5.3 Scwarz reflection principle

Let Ω be an open subset of $\mathbb{C}$ that is symmetric with respect to the real line, that is $z \in \Omega$ if and only if $\bar{z} \in \Omega$. Let Ω^+ denote the part of Ω that lies in the upper half-plane and Ω^- that part that lies in the lower half-plane. Also, let $I = \Omega \cap \mathbb{R}$ so that I denotes the interior of that part of the boundary of Ω^+ and Ω^- that lies on the real axis. Then we have

$$\Omega^+ \cup I \cup \Omega^- = \Omega$$

and the only interesting case of the next theorem occurs, of course, when I is nonempty.

Theorem 56.16. *If f^+ and f^- are holomorphic functions in Ω^+ and Ω^- respectively, that extend continuously to I and $f^+(x) = f^-(x)$ for all $x \in I$, then the function f defined on Ω by*

$$f(z) = \begin{cases} f^+(z) & \text{if } z \in \Omega^+, \\ f^+(z) = f^-(z) & \text{if } z \in I, \\ f^-(z) & \text{if } z \in \Omega^- \end{cases}$$

is holomorphic on all of Ω.

Proof. One first notes that f is continuous on Ω. The only difficulty is to prove that f is holomorphic at points of I. Suppose D is a disc centered at a point on I and entirely contained in Ω. We prove that f is holomorphic in D by Morera's theorem. Suppose T is a triangle in D. If T does not intersect I, then

$$\int_T f(z)dz = 0$$

since f is holomorphic in the upper and lower half-discs. Suppose now that one side or vertex of T is contained in I, and the rest of$T$ is in, say the upper half-disc. If T_ε is the triangle obained from T by slightly raising the edge or vertex which lies on I, we have

$$\int_{T_\varepsilon} f(z)dz = 0$$

since T_ε is entirely contained in the upper half-disc. We then let $\varepsilon \to 0$, and by continuity we conclude that

$$\int_T f(z)dz = 0.$$

If the interior of T intersects I, we can reduce the siutation to the previous one by writing T as the union of triangles each of which has an edge or vertex on I. By Morera's theorem, we conclude that f is holomorphic in D, as was to be shown. □

Theorem 56.17. *(Schwarz reflection principle) Suppose that f is a holomorphic function in Ω^+ that extends continuously to I and such that f is real-valued on I. Then there exists a function F holomorphic in all of Ω such that $F = f$ on Ω^+.*

Proof. The idea is simply to define $F(z)$ for $z \in \Omega^-$ by $F(z) = \overline{f(\overline{z})}$. To prove that F is holomorphic in Ω^- we note that if $z, z_0 \in \Omega^-$, then $\overline{z}, \overline{z}_0 \in \Omega^+$ and hence, the power series expansion of f near $\overline{z}_0$ gives

$$f(\overline{z}) = \sum_{n=0}^{\infty} a_n \left(\overline{z} - \overline{z}_0\right)^n.$$

As a consequence we see that

$$F(z) = \sum_{n=0}^{\infty} \overline{a}_n \left(z - z_0\right)^n$$

and F is holomorphic in Ω^-. Since f is real valued on I, we have $\overline{f(x)} = f(x)$ whenever $x \in I$ and hence F extends continuously up to I. The proof is complete once we invoke the symmetry principle. □

Theorem 56.18. *Suppose f is a holomorphic function in a region Ω that vanishes on a sequence of distinct points with a limit point in Ω. Then f is identically 0.*

Proof. Suppose that $z_0 \in \Omega$ is a limit point for the sequence $\{w_k\}_{k=1}^{\infty}$ and that $f(\omega_k) = 0$. First, we show that f is idenitcally zero in a small disc containing z_0. For that, we choose a disc D centered at z_0 and contained in Ω, and consider the power series expansion of f in that disc

$$\sum_{n=0}^{\infty} a_n (z - z_0)^n.$$

If f is not identically zero, there exists a smallest integer m such that $a_m \neq 0$. But then we can write

$$f(z) = a_m (z - z_0)^m (1 + g(z - z_0)),$$

where $g(z - z_0)$ converges to 0 as $z \to z_0$. Taking $z = w_k \neq z_0$ for a sequence of points converging to z_0, we get a contradiction since $a_m (w_k - z_0)^m \neq 0$ and $1 + g(w_k - z_0) \neq 0$, but $f(w_k) = 0$.

We conclude the proof using the fact that Ω is connected. Let U denote the interior of the set of points where $f(z) = 0$. Then U is open by definition and nonempty by the argument just given. The set U is also closed since if $z_n \in U$ and $z_n \to z$, then $f(z) = 0$ by continuity, and f vanishes in a neighborhood of z by the argument above. Hence, $z \in U$. Now if we let V denote the complement of U in Ω, we conclude that U and V are both open, disjoint, and

$$\Omega = U \cup V.$$

Since Ω is connected we conclude that either U or V is empty. Since $z_0 \in U$, we find that $U = \Omega$ and the proof is complete.

$\square$

Corollary 28. *Suppose f and g are holomorphic in a region Ω and $f(z) = g(z)$ for all z in some nonempty open subset of Ω (or more generally for z in some sequence of distinct points with limit point in Ω). Then $f(z) = g(z)$ througout Ω.*

Suppose we are given a pair of functions f and F analytic in regions Ω and Ω', respectively, with $\Omega \subset \Omega'$. If the two functions agree on the smaller set Ω, we say that F is an **analytic continuiation** of f into the region Ω' The corollary then guarantees that there can be only one such analytic continuation, since F is uniquely determined by f.