Christopher Heil

Introduction to Real Analysis II

Abstract Measures, Operator Theory, and Functional Analysis

April 12, 2017

Springer

Contents

Preliminaries

We use the symbol □ to denote the end of a proof, and the symbol ♢ to denote the end of a definition, remark, example, or exercise. We also use ♢ to indicate the end of the statement of a theorem whose proof will be omitted. More challenging Problems are marked with an asterisk *. A detailed index of symbols employed in the text can be found at the end of the volume.

The text contains both "Exercises" and "Problems." The Exercises are incorporated into the development of the theory in each section. A list of additional Problems appears at the end of most sections. The Problems not only provide additional practice, but also help to further develop understanding of the material of the section. Hints for selected Exercises and Problems appear at the end of the volume. Certain especially challenging problems are marked with the symbol "*".

Numbers

The set of natural numbers is denoted by $\mathbb{N} = \{1, 2, 3, \dots\}$. The set of integers is $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$, $\mathbb{Q}$ denotes the set of rational numbers, $\mathbb{R}$ is the set of real numbers, and $\mathbb{C}$ is the set of complex numbers.

The *real part* of a complex number $z = a+ib$ (where $a, b \in \mathbb{R}$) is $\mathrm{Re}(z) = a$, and its *imaginary part* is $\mathrm{Im}(z) = b$. We say that z is *rational* if both its real and imaginary parts are rational numbers. The *complex conjugate* of z is $\overline{z} = a - ib$. The *polar form* of z is $z = re^{i\theta}$ where $r \ge 0$ and $\theta \in [0, 2\pi)$. The *modulus*, or *absolute value*, of z is $|z| = \sqrt{z\overline{z}} = \sqrt{a^2 + b^2} = r$. If $z \ne 0$ then its *argument* is $\arg(z) = \theta$. Given any $z \in \mathbb{C}$, there is a complex number α such that $|\alpha| = 1$ and $z\alpha = |z|$. If $z \ne 0$ then $\alpha = e^{-i\theta} = \overline{z}/|z|$, while if $z = 0$ then α can be any complex number that has unit modulus.

The set of *extended real numbers* is

$$\mathbb{R} \cup \{-\infty, \infty\} = [-\infty, \infty].$$

We extend many of the normal arithmetic operations to $[-\infty, \infty]$. For example, if $-\infty < a \le \infty$ then we set $a + \infty = \infty$. However, $\infty - \infty$ and $-\infty + \infty$ are undefined, and are referred to as *indeterminate forms*. Given a strictly positive extended real number $0 < a \le \infty$ we define

$$a \cdot \infty = \infty, \quad (-a) \cdot \infty = -\infty, \quad a \cdot (-\infty) = -\infty, \quad (-a) \cdot (-\infty) = \infty.$$

We also adopt the following conventions:

$$0 \cdot (\pm\infty) = 0, \qquad \frac{1}{\pm\infty} = 0.$$

Given an extended real number p in the range $1 \le p \le \infty$, its *dual index* is the extended real number p' that satisfies

$$\frac{1}{p} + \frac{1}{p'} = 1.$$

We have $1 \le p' \le \infty$, and $(p')' = p$. If $1 < p < \infty$, then $p' = p/(p-1)$. Some examples are $1' = \infty$, $\left(\frac{3}{2}\right)' = 3$, $2' = 2$, $3' = \frac{3}{2}$, and $\infty' = 1$.

Given i, j in an index set I (typically $I = \mathbb{N}$), the *Kronecker delta* of i and j is the number δ_{ij} defined by the rule

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \ne j. \end{cases}$$

Sets

We write $A \subseteq B$ to denote that A is a subset of a set B. If $A \subseteq B$ and $A \ne B$ then we say that A is a *proper subset* of B, and we write $A \subsetneq B$.

The *empty set* is denoted by $\varnothing$.

A collection of sets $\{X_i\}_{i \in I}$ is *disjoint* if $X_i \cap X_j = \varnothing$ whenever $i \ne j$. The collection $\{X_i\}_{i \in I}$ is a *partition* of X if it is disjoint and $\cup_{i \in I} X_i = X$.

If X is a set, then the *complement* of $S \subseteq X$ is $X \setminus S = \big\{x \in X : x \notin S\big\}$. We sometimes abbreviate $X \setminus S$ as S^{C} if the set X is understood. If A, B are subsets of X, then the *relative complement* of A in B is

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

The *power set* of X is $\mathcal{P}(X) = \big\{S : S \subseteq X\big\}$, the set of all subsets of X.

The *Cartesian product* of sets X and Y is $X \times Y = \{(x, y) : x \in X, y \in Y\}$, the set of all ordered pairs of elements of X and Y. The Cartesian product of finitely many sets $X_1, \dots, X_N$ is

$$\prod_{j=1}^{N} X_j \;=\; X_1 \times \cdots \times X_N \;=\; \bigl\{(x_1, \dots, x_N) : x_k \in X_k,\, k = 1, \dots, N\bigr\}.$$

A set is *countable* if it is either finite or has the same cardinality as the natural numbers $\mathbb{N}$. We sometimes say that an infinite countable set is *countably infinite.* A set that is not countable is said to be *uncountable.*

Intervals

An *interval* in the real line $\mathbb{R}$ is any one of the following sets:

- (a, b), $[a, b)$, $(a, b]$, $[a, b]$ where $a,\, b \in \mathbb{R}$ and $a < b$, or
- (a, ∞), $[a, \infty)$, $(-\infty, a)$, $(-\infty, a]$ where $a \in \mathbb{R}$, or
- $\mathbb{R} = (-\infty, \infty)$.

An *open interval* is an interval of the form (a, b), (a, ∞), $(-\infty, a)$, or $(-\infty, \infty)$. A *closed interval* is an interval of the form $[a, b]$, $[a, \infty)$, $(-\infty, a]$, or $(-\infty, \infty)$. We refer to $[a, b]$ as a *finite closed interval* or a *compact interval.*

The empty set $\varnothing$ and a singleton $\{a\}$ are not intervals, but even so we adopt the notational conventions

$$[a, a] \;=\; \{a\} \qquad \text{and} \qquad (a, a) \;=\; [a, a) \;=\; (a, a] \;=\; \varnothing.$$

We also consider *extended intervals*, which are

- $(a, \infty] = (a, \infty) \cup \{\infty\}$ or $[a, \infty] = [a, \infty) \cup \{\infty\}$, where $a \in \mathbb{R}$,
- $[-\infty, b) = (-\infty, b) \cup \{-\infty\}$ or $[-\infty, b] = (-\infty, b] \cup \{-\infty\}$, where $b \in \mathbb{R}$,
- $[-\infty, \infty] = (-\infty, \infty) \cup \{-\infty, \infty\}$.

An extended interval is not an interval—whenever we refer to an "interval" without qualification we implicitly *exclude* the extended intervals.

Euclidean Space

We let $\mathbb{R}^d$ denote d-dimensional real Euclidean space, the set of all ordered d-tuples of real numbers. Similarly, $\mathbb{C}^d$ is d-dimensional complex Euclidean space, the set of all ordered d-tuples of complex numbers.

The *zero vector* is $0 = (0, \dots, 0)$. We use the same symbol "0" to denote the zero vector and the number zero; the intended meaning should be clear from context.

The *dot product* of vectors $x = (x_1, \dots, x_d)$ and $y = (y_1, \dots, y_d)$ in $\mathbb{R}^d$ or $\mathbb{C}^d$ is

$$x \cdot y \;=\; x_1\overline{y_1} + \cdots + x_d\overline{y_d},$$

and the *Euclidean norm* of x is

$$\|x\| \;=\; (x \cdot x)^{1/2} \;=\; \bigl(|x_1|^2 + \cdots + |x_d|^2\bigr)^{1/2}.$$

The *translation* of $E \subseteq \mathbb{R}^d$ by a vector $h \in \mathbb{R}^d$ (or $E \subseteq \mathbb{C}^d$ by $h \in \mathbb{C}^d$) is $E + h = \{x + h : x \in E\}$.

Sequences

Let I be a fixed set. Given a set X and points $x_i \in X$ for $i \in I$, we write $\{x_i\}_{i \in I}$ to denote the sequence of elements x_i indexed by the set I. We call I an *index set* in this context, and refer to x_i as the *ith component* of the sequence $\{x_i\}_{i \in I}$. If we know that the x_i are *scalars* (real or complex numbers), then we often write $(x_i)_{i \in I}$ instead of $\{x_i\}_{i \in I}$. Technically, a sequence $\{x_i\}_{i \in I}$ is shorthand for the mapping $x \colon I \to X$ given by $x(i) = x_i$, and therefore the components x_i of a sequence need not be distinct. If the index set I is understood then we may write $\{x_i\}$ or $\{x_i\}_i$, of if the x_i are scalars then we may write (x_i) or $(x_i)_i$.

Countable Sequences

Often the index set I of a sequence is countable. If $I = \{1, \dots, d\}$ then we may write a sequence in list form as

$$\{x_n\}_{n=1}^d \;=\; \{x_1, \dots, x_d\},$$

or if the x_n are scalars then we often write $(x_n)_{n=1}^d = (x_1, \dots, x_d)$. Similarly, if $I = \mathbb{N}$ then we may write

$$\{x_n\}_{n \in \mathbb{N}} \;=\; \{x_1, x_2, \dots\},$$

or if each x_n is a scalar then we usually write $(x_n)_{n \in \mathbb{N}} = (x_1, x_2, \dots)$.

A *subsequence* of a countable sequence $\{x_n\}_{n \in \mathbb{N}} = \{x_1, x_2, \dots\}$ is a sequence of the form $\{x_{n_k}\}_{n \in \mathbb{N}} = \{x_{n_1}, x_{n_2}, \dots\}$ where $n_1 < n_2 < \cdots$.

We say that a countable sequence of real numbers $(x_n)_{n \in \mathbb{N}}$ is *monotone increasing* if $x_n \le x_{n+1}$ for every n, and *strictly increasing* if $x_n < x_{n+1}$ for every n. We define monotone decreasing and strictly decreasing sequences similarly.

Functions

Let X and Y be sets. We write $f\colon X \to Y$ to mean that f is a function with domain X and codomain Y. We usually write $f(x)$ to denote the image of x under f, but if $L\colon X \to Y$ is a linear map from one vector space X to another vector space Y then we may write Lx instead of $L(x)$.

- The *direct image* of a set $A \subseteq X$ under f is $f(A) = \{f(x) : x \in A\}$.
- The *inverse image* of a set $B \subseteq Y$ under f is

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

- The *range* of f is $\mathrm{range}(f) = f(X) = \{f(x) : x \in X\}$.
- f is *injective*, or *1-1*, if $f(x) = f(y)$ implies $x = y$.
- f is *surjective*, or *onto*, if $\mathrm{range}(f) = Y$.
- f is *bijective* if it is both injective and surjective. The inverse function of a bijection $f\colon X \to Y$ is the function $f^{-1}\colon Y \to X$ defined by $f^{-1}(y) = x$ if $f(x) = y$.
- Given $S \subseteq X$, the *restriction* of a function $f\colon X \to Y$ to the domain S is the function $f|_S\colon S \to Y$ defined by $(f|_S)(x) = f(x)$ for $x \in S$.
- The *zero function on* X is the function $0\colon X \to \mathbb{R}$ defined by $0(x) = 0$ for every $x \in X$. We use the same symbol 0 to denote the zero function and the number zero.
- The *characteristic function* of $A \subseteq X$ is the function $\chi_A\colon X \to \mathbb{R}$ given by

$$\chi_A(x) \;=\; \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

- If the domain of a function f is $\mathbb{R}^d$, then the *translation* of f by a vector $a \in \mathbb{R}^d$ is the function $T_a f$ defined by $T_a f(x) = f(x-a)$ for $x \in \mathbb{R}^d$.

Extended Real-Valued Functions

A function that maps a set X into the real line $\mathbb{R}$ is called a *real-valued function*, and a function that maps X into the extended real line $[-\infty, \infty]$ is an *extended real-valued function.* Every real-valued function is extended real-valued, but an extended real-valued function need not be real-valued. An extended real-valued function f is *nonnegative* if $f(x) \ge 0$ for every x, where we use the convention that $a < \infty$ for every real number a.

Let $f\colon X \to [-\infty, \infty]$ be an extended real-valued function. We associate to f the two extended real-valued functions f^+, f^- defined by

$$f^+(x) \;=\; \max\{f(x),0\} \qquad \text{and} \qquad f^-(x) \;=\; \max\{-f(x),0\}.$$

We call f^+ the *positive part* and f^- the *negative part* of f. They are each nonnegative, and $f(x) = f^+(x) - f^-(x)$ while $|f(x)| = f^+(x) + f^-(x)$.

Given $f\colon X \to [-\infty,\infty]$, to avoid multiplicities of parentheses, brackets, and braces, we often write $f^{-1}(a,b) = f^{-1}((a,b))$, $f^{-1}[a,\infty) = f^{-1}([a,\infty))$, and so forth. We also use shorthands such as

$$\begin{aligned}
\{f \ge a\} &= \{x \in X : f(x) \ge a\} &&= f^{-1}[a,\infty],\\
\{f = a\} &= \{x \in X : f(x) = a\} &&= f^{-1}\{a\},\\
\{a < f < b\} &= \{x \in X : a < f(x) < b\} &&= f^{-1}(a,b),\\
\{f \ge g\} &= \{x \in X : f(x) \ge g(x)\},\\
\{f = g\} &= \{x \in X : f(x) = g(x)\}.
\end{aligned}$$

If $f\colon S \to [-\infty,\infty]$ is an extended real-valued function on a set $S \subseteq \mathbb{R}$, then f is *monotone increasing* on S if

$$x,y \in S,\ x \le y \implies f(x) \le f(y).$$

We say that f is *strictly increasing* on S if

$$x,y \in S,\ x < y \implies f(x) < f(y).$$

Monotone decreasing and *strictly decreasing* functions are defined in an entirely similar manner.

Notation for Extended Real-Valued and Complex-Valued Functions

A function of the form $f\colon X \to \mathbb{C}$ is said to be *complex valued*. We have the inclusions $\mathbb{R} \subseteq [-\infty,\infty]$ and $\mathbb{R} \subseteq \mathbb{C}$, so every real-valued function is both an extended real-valued and a complex-valued function. However, neither $[-\infty,\infty]$ nor $\mathbb{C}$ is a subset of the other, so an extended real-valued function need not be a complex-valued function, and a complex-valued function need not be an extended real-valued function. Hence there are usually two cases to consider:

- extended real-valued functions of the form $f\colon X \to [-\infty,\infty]$, and
- complex-valued functions of the form $f\colon X \to \mathbb{C}$.

To avoid excessive duplication, we introduce a notation that will allow us to consider both cases together.

Notation (Scalars and the Symbol $\mathbb{F}$). We let the symbol $\mathbb{F}$ denote a choice of either the extended real line $[-\infty, \infty]$ or the complex plane $\mathbb{C}$. Associated with this choice, we make the following declarations.

- If $\mathbb{F} = [-\infty, \infty]$, then the word *scalar* means a *finite real number* $c \in \mathbb{R}$.
- If $\mathbb{F} = \mathbb{C}$, then the word *scalar* means a *complex number* $c \in \mathbb{C}$. $\diamondsuit$

Suprema and Infima

A set of real numbers S is *bounded above* if there exists a real number M such that $x \le M$ for every $x \in S$. Any such number M is called an *upper bound* for S. The definition of *bounded below* is similar, and we say that S is *bounded* if it is bounded both above and below.

A number $x \in \mathbb{R}$ is the *supremum*, or *least upper bound*, of S if

- x is an upper bound for S, and
- if y is any upper bound for S, then $x \le y$.

We denote the supremum of S, if one exists, by $x = \sup(S)$. The *infimum*, or greatest lower bound, of S is defined in an entirely analogous manner, and is denoted by $\inf(S)$.

It is not obvious that every set that is bounded above has a supremum. We take the existence of suprema as the following axiom.

Axiom (Supremum Property of $\mathbb{R}$). Let S be a nonempty subset of $\mathbb{R}$. If S is bounded above, then there exists a real number $x = \sup(S)$ that is the supremum of S. $\diamondsuit$

We extend the definition of supremum to sets that are not bounded above by declaring that $\sup(S) = \infty$ if S is not bounded above. We also declare that $\sup(\varnothing) = -\infty$. With these conventions, every set $S \subseteq \mathbb{R}$ has a supremum in the extended real sense.

If $S = (x_n)_{n \in \mathbb{N}}$ is countable, then we often write $\sup_n x_n$ or $\sup x_n$ to denote the supremum instead of $\sup(S)$, and similarly we may write $\inf_n x_n$ or $\inf x_n$ instead of $\inf(S)$.

If $(x_n)_{n \in \mathbb{N}}$ and $(y_n)_{n \in \mathbb{N}}$ are two sequences of real numbers, then

$$\inf_n x_n + \inf_n y_n \;\le\; \inf_n\,(x_n + y_n) \;\le\; \sup_n\,(x_n + y_n) \;\le\; \sup_n x_n + \sup_n y_n.$$

Any or all of the inequalities on the preceding line can be strict. If $c > 0$ then

$$\sup_n c x_n \;=\; c \sup_n x_n \qquad \text{and} \qquad \sup_n\,(-c x_n) \;=\; -c \inf_n x_n.$$

Convergent and Cauchy Sequences of Scalars

Given a sequence of real or complex numbers $(x_n)_{n\in\mathbb{N}}$, we say that $(x_n)_{n\in\mathbb{N}}$ *converges* if there exists a scalar x such that for every $\varepsilon > 0$ there is an $N > 0$ such that

$$n > N \quad \Longrightarrow \quad |x - x_n| < \varepsilon.$$

In this case we say that x_n *converges to* x *as* $n \to \infty$ and write

$$x_n \to x \qquad \text{or} \qquad \lim_{n\to\infty} x_n \;=\; x \qquad \text{or} \qquad \lim x_n \;=\; x.$$

We say that $(x_n)_{n\in\mathbb{N}}$ is a *Cauchy sequence* if for every $\varepsilon > 0$ there exists an integer $N > 0$ such that

$$m, n \;>\; N \quad \Longrightarrow \quad |x_m - x_n| < \varepsilon. \qquad \diamondsuit$$

The following equivalence holds for any sequence of scalars $(x_n)_{n\in\mathbb{N}}$:

$$(x_n)_{n\in\mathbb{N}} \text{ is convergent} \quad \Longleftrightarrow \quad (x_n)_{n\in\mathbb{N}} \text{ is Cauchy.}$$

Convergence in the Extended Real Sense

Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of *real* numbers. We say that $(x_n)_{n\in\mathbb{N}}$ *diverges to* ∞ as $n \to \infty$ if given any $R > 0$ there is an $N > 0$ such that $x_n > R$ for all $n > N$. In this case we write

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

We define *divergence to* $-\infty$ similarly.

We say that $(x_n)_{n\in\mathbb{N}}$ *converges in the extended real sense* if

- x_n converges to a real number x as $n \to \infty$, *or*
- x_n diverges to ∞ as $n \to \infty$, *or*
- x_n diverges to $-\infty$ as $n \to \infty$.

Every monotone increasing sequence of real numbers $(x_n)_{n\in\mathbb{N}}$ converges in the extended real sense, and in this case $\lim x_n = \sup x_n$. Similarly, a monotone decreasing sequence of real numbers converges in the extended real sense and its limit equals its infimum.

Limsup and Liminf

The *limit superior*, or *limsup*, of a sequence of real numbers $(x_n)_{n\in\mathbb{N}}$ is

$$\limsup_{n\to\infty} x_n \;=\; \inf_{n\in\mathbb{N}} \sup_{m\ge n} x_m \;=\; \lim_{n\to\infty} \sup_{m\ge n} x_m.$$

Likewise, the *limit inferior*, or *liminf*, of $(x_n)_{n\in\mathbb{N}}$ is

$$\liminf_{n\to\infty} x_n \;=\; \sup_{n\in\mathbb{N}} \inf_{m\ge n} x_m \;=\; \lim_{n\to\infty} \inf_{m\ge n} x_m.$$

The liminf and limsup of every sequence of real numbers exists in the extended real sense. Further,

$$\begin{array}{c}(x_n)_{n\in\mathbb{N}} \text{ converges in} \\ \text{the extended real sense}\end{array} \quad\Longleftrightarrow\quad \liminf_{n\to\infty} x_n \;=\; \limsup_{n\to\infty} x_n,$$

and in this case, $\lim x_n = \liminf x_n = \limsup x_n$.

If $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are two sequences of real numbers, then

$$\begin{aligned}\liminf_{n\to\infty} x_n \;+\; \liminf_{n\to\infty} y_n \;&\le\; \liminf_{n\to\infty}\,(x_n+y_n)\\ &\le\; \limsup_{n\to\infty} x_n \;+\; \liminf_{n\to\infty} y_n\\ &\le\; \limsup_{n\to\infty}\,(x_n+y_n)\\ &\le\; \limsup_{n\to\infty} x_n \;+\; \limsup_{n\to\infty} y_n,\end{aligned}$$

as long as none of these sums takes the indeterminate forms $\infty-\infty$ or $-\infty+\infty$. Strict inequality can hold on any line above. If the sequence $\{x_n\}_{n\in\mathbb{N}}$ converges, then

$$\liminf_{n\to\infty}\,(x_n+y_n) \;=\; \lim_{n\to\infty} x_n \;+\; \liminf_{n\to\infty} y_n$$

and a similar equality holds with liminf replaced by limsup.

Given any sequence of real numbers $(x_n)_{n\in\mathbb{N}}$, there exist subsequences $(x_{n_k})_{k\in\mathbb{N}}$ and $(x_{m_j})_{j\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} x_{n_k} \;=\; \limsup_{n\to\infty} x_n \qquad\text{and}\qquad \lim_{j\to\infty} x_{m_j} \;=\; \liminf_{n\to\infty} x_n.$$

In fact, if $(x_n)_{n\in\mathbb{N}}$ is bounded above then $\limsup x_n$ is the largest accumulation point of $(x_n)_{n\in\mathbb{N}}$, and likewise if $(x_n)_{n\in\mathbb{N}}$ is bounded below then $\liminf x_n$ is the smallest accumulation point.

On occasion we deal with real-parameter versions of liminf and limsup. Given a real-valued function f whose domain includes an interval centered at a point $x\in\mathbb{R}$, we define

$$\limsup_{t\to x} f(t) \;=\; \inf_{\delta>0} \sup_{|t-x|<\delta} f(t) \;=\; \lim_{\delta\to 0} \sup_{|t-x|<\delta} f(t),$$

and $\liminf_{t\to x} f(t)$ is defined analogously. The properties of these real-parameter versions of liminf and limsup are similar to those of the sequence versions.

Infinite Series

If $(c_n)_{n\in\mathbb{N}}$ is a sequence of real or complex numbers then we say that the series $\sum_{n=1}^\infty c_n$ *converges* if there is a scalar s such that the *partial sums* $s_N = \sum_{n=1}^N c_n$ converge to s as $N\to\infty$. In this case $\sum_{n=1}^\infty c_n$ is assigned the value s:

$$\sum_{n=1}^\infty c_n \;=\; \lim_{N\to\infty} s_N \;=\; \lim_{N\to\infty}\sum_{n=1}^N c_n \;=\; s.$$

If every c_n is a *real number*, then we say that the series $\sum_{n=1}^\infty c_n$ *converges in the extended real sense* if

- s_N converges to a real number s as $N\to\infty$, or
- s_N diverges to ∞ as $N\to\infty$, or
- s_N diverges to $-\infty$ as $N\to\infty$.

If every c_n is a *nonnegative real number*, then the series $\sum_{n=1}^\infty c_n$ will converge in the extended real sense to a nonnegative extended real number.

If $c_n \ge 0$ for every n then we write

$$\sum_{n=1}^\infty c_n \;<\; \infty$$

to mean that *the series $\sum_{n=1}^\infty c_n$ converges*, i.e., it converges to a nonnegative, finite real number. Similarly, if $c_n \ge 0$ for every n then $\sum c_n = \infty$ means that the series diverges to infinity.

Pointwise Convergence of Functions

If X is a set and $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of extended real-valued or complex-valued functions whose domain is X, then we say that f_n *converges pointwise to a function f* if

$$f(x) \;=\; \lim_{n\to\infty} f_n(x), \qquad \text{all } x\in X.$$

In this case we write $f_n(x)\to f(x)$ *for every $x\in X$* or $f_n\to f$ *pointwise.*

If $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of extended real-valued functions whose domain is a set X, then we say that $\{f_n\}_{n\in\mathbb{N}}$ is a *monotone increasing sequence* if

$\{f_n(x)\}_{n\in\mathbb{N}}$ is monotone increasing for each x, i.e.,

$$f_1(x) \le f_2(x) \le \cdots \qquad \text{for all } x \in X.$$

In this case $f(x) = \lim_{n\to\infty} f_n(x)$ exists for each $x \in X$ in the extended real sense, and we say that f_n *increases pointwise to* f. We denote this by writing

$$f_n \nearrow f \quad \text{on } X.$$

Continuity and Uniform Continuity

Let E be a subset of $\mathbb{R}^d$. We say that a function $f\colon E \to \mathbb{C}$ is *continuous on* E if given any points $x_n, x \in E$ such that $x_n \to x$ we have $f(x_n) \to f(x)$.

Derivatives and Everywhere Differentiability

Let f be a complex-valued function whose domain includes an open interval centered at a point $x \in \mathbb{R}$. We say that f is *differentiable at* x if the limit

$$f'(x) = \lim_{y\to x} \frac{f(y)-f(x)}{y-x}$$

exists (as a scalar, not as $\pm\infty$).

Let $[a,b]$ be a closed interval in the real line. A function f is *everywhere differentiable* or *differentiable everywhere* on $[a,b]$ if it is differentiable at each point in the interior (a,b) and if the appropriate one-sided derivatives exist at the endpoints a and b. That is, f is everywhere differentiable on $[a,b]$ if

$$f'(x) = \lim_{y\to x,\, y\in[a,b]} \frac{f(y)-f(x)}{y-x} \quad \text{exists for each } x \in [a,b].$$

We use similar terminology if f is defined on other types of intervals in $\mathbb{R}$. For example, $x^{3/2}$ is differentiable everywhere on $[0,1]$ and $x^{1/2}$ is differentiable everywhere on $(0,1]$, but $x^{1/2}$ is not differentiable everywhere on $[0,1]$.

Chapter 1
Abstract Measure Theory

Lebesgue measure is one of the premier examples of a measure on $\mathbb{R}^d$, but it is not the only measure and certainly not the only important measure on $\mathbb{R}^d$. Further, $\mathbb{R}^d$ is not the only domain on which we encounter measures. This chapter develops the theory of measures from an abstract viewpoint. The fact that we have already examined Lebesgue measure in detail will simplify this task considerably, as there are many aspects of abstract measure theory that are precisely analogous to results for Lebesgue measure. However, other portions of the theory require some extra care. In return for developing this abstract theory of measure, we will be able to construct a powerful and useful theory of integration with respect to measures.

For proofs of statements made about Lebesgue measure, see [Heil18], which we will refer to as *Volume 1*.

This chapter draws heavily from Folland's text [Fol99], and so I encourage you to consult that text as a comparison.

1.1 Sigma Algebras

There are three components to any measure. First there is the set X whose subsets we wish to measure. Unfortunately, we often cannot construct a measure that is suitably well-behaved for our application on all of the subsets of X, and hence the second component of a measure is the collection Σ of subsets of X that we will actually be allowed to measure. Finally, there is the measure itself, which is a mapping $\mu\colon \Sigma \to [0,\infty]$.

We cannot choose Σ at random; it must satisfy the properties of a σ-algebra. We state the definition here, and also introduce a terminology for referring to a set together with a σ-algebra.

Definition 1.1.1 (Sigma Algebra). Let X be a set.

(a) A *σ-algebra* or *σ-field* on X is a nonempty collection Σ of subsets of X that is closed under complements and countable unions.

(b) If Σ is a σ-algebra on X, then we call (X, Σ) a *measurable space*, and we refer to the elements of Σ as the *measurable subsets of X*. $\diamondsuit$

By definition, *countable* means either finite or countably infinite, so a σ-algebra is closed under both finite and countably infinite unions. Since a σ-algebra is also closed under complements, it follows that it is closed under countable intersections. Additionally, a σ-algebra is also closed under *relative complements*, for if $A, B \in \Sigma$ then $A \setminus B = A \cap B^{\mathrm{C}} \in \Sigma$.

If Σ is a σ-algebra then it must contain at least one subset E of X. Therefore $E^{\mathrm{C}} = X \setminus E$ belongs to Σ, and hence both

$$X = E \cup E^{\mathrm{C}} \qquad \text{and} \qquad \emptyset = X \setminus X$$

are elements of Σ.

Example 1.1.2. (a) Trivial examples of σ-algebras on a set X are $\Sigma = \{\emptyset, X\}$ and the power set $\mathcal{P}(X) = \{E : E \subseteq X\}$.

(b) We proved in Volume 1 that the class $\mathcal{L} = \mathcal{L}_{\mathbb{R}^d}$ of all Lebesgue measurable subsets of $\mathbb{R}^d$ is a σ-algebra on $\mathbb{R}^d$. $\diamondsuit$

We often encounter situations where we have a particular family $\mathcal{E}$ of subsets of X that we want to measure, but $\mathcal{E}$ is not a σ-algebra. In this case there will usually be many larger collections that are σ-algebras and include all of the sets from $\mathcal{E}$. For example, the power set $\mathcal{P}(X)$ is a σ-algebra that contains $\mathcal{E}$. However, we often seek the "smallest possible" σ-algebra that contains $\mathcal{E}$. The next exercise constructs this smallest σ algebra.

Exercise 1.1.3. Let $\mathcal{E}$ be a collection of subsets of a set X. Show that

$$\Sigma(\mathcal{E}) = \bigcap \bigl\{\Sigma : \Sigma \text{ is a } \sigma\text{-algebra and } \mathcal{E} \subseteq \Sigma\bigr\} \tag{1.1}$$

is a σ-algebra on X. We call $\Sigma(\mathcal{E})$ the *σ-algebra generated by $\mathcal{E}$*. $\diamondsuit$

Note that if Σ_1, Σ_2 are σ-algebras, then $\Sigma_1 \cap \Sigma_2$ is not formed by intersecting the elements of Σ_1 with those of Σ_2. That is, $\Sigma_1 \cap \Sigma_2$ does *not* mean $\{A \cap B : A \in \Sigma_1,\ B \in \Sigma_2\}$. Rather, $\Sigma_1 \cap \Sigma_2$ is the collection of all sets that are *common* to both Σ_1 and Σ_2:

$$\Sigma_1 \cap \Sigma_2 = \bigl\{A : A \in \Sigma_1 \text{ and } A \in \Sigma_2\bigr\}. \tag{1.2}$$

Similarly, the collection $\Sigma(\mathcal{E})$ defined in equation (1.1) consists of all those sets A that belong to every σ-algebra Σ that satisfies $\mathcal{E} \subseteq \Sigma$.

Example 1.1.4. Let $\mathcal{F}$ be the collection of all F_σ subsets of $\mathbb{R}^d$, and let $\mathcal{G}$ be the collection of all G_δ subsets. These are not σ-algebras on $\mathbb{R}^d$, but we can use them to illustrate equation (1.2). The intersection $\mathcal{F} \cap \mathcal{G}$ does not consist of all possible intersections of F_σ sets with G_δ sets, but rather is the collection of sets that are common to both $\mathcal{F}$ and $\mathcal{G}$:

$$\mathcal{F} \cap \mathcal{G} = \{S \subseteq \mathbb{R}^d : S \text{ is both an } F_\sigma \text{ and a } G_\delta \text{ set}\}. \qquad \diamondsuit$$

The following exercise explains why we call $\Sigma(\mathcal{E})$ the *smallest σ-algebra that contains $\mathcal{E}$.*

Exercise 1.1.5. Let $\mathcal{E}$ be a collection of subsets of a set X. Show that the σ-algebra $\Sigma(\mathcal{E})$ contains $\mathcal{E}$, and if Σ is any other σ-algebra that contains $\mathcal{E}$ then $\Sigma(\mathcal{E}) \subseteq \Sigma$. $\diamondsuit$

Here is an example of a σ-algebra and a generating family for that σ-algebra.

Exercise 1.1.6. Given a set X, let Σ consist of all sets $E \subseteq X$ such that at least one of E or $X \setminus E$ is countable.

(a) Show that Σ is a σ-algebra on X.

(b) Let $\mathcal{S} = \bigl\{\{x\} : x \in X\bigr\}$ be the set of all singletons of elements of X. Show that $\Sigma = \Sigma(\mathcal{S})$, i.e., Σ is the σ-algebra generated by $\mathcal{S}$. $\diamondsuit$

When working with $\mathbb{R}^d$, we inevitably must deal with the topology of $\mathbb{R}^d$. The Lebesgue σ-algebra $\mathcal{L}_{\mathbb{R}^d}$ contains all of the open subsets of $\mathbb{R}^d$, but is it the *smallest* σ-algebra that contains the open sets? Problem 1.1.19 shows that the answer to this is no—the σ-algebra generated by the open subsets of $\mathbb{R}^d$ is a *proper* subset of $\mathcal{L}_{\mathbb{R}^d}$. We have a special name for this σ-algebra.

Definition 1.1.7 (Borel σ-algebra on $\mathbb{R}^d$). The *Borel σ-algebra $\mathcal{B}_{\mathbb{R}^d}$ on* $\mathbb{R}^d$ is the smallest σ-algebra that contains all of the open subsets of $\mathbb{R}^d$. That is, if we set $\mathcal{U} = \{U \subseteq \mathbb{R}^d : U \text{ is open}\}$, then

$$\mathcal{B}_{\mathbb{R}^d} = \Sigma(\mathcal{U}).$$

The elements of $\mathcal{B}_{\mathbb{R}^d}$ are called the *Borel subsets of* $\mathbb{R}^d$. $\diamondsuit$

In particular, $\mathcal{B}_{\mathbb{R}^d}$ includes all of the open and closed subsets of $\mathbb{R}^d$, as well as the G_δ and F_σ subsets of $\mathbb{R}^d$. However, not every Lebesgue measurable subset of $\mathbb{R}^d$ is a Borel set, and not every Borel set is a G_δ or an F_σ set. On the other hand, if E is a Lebesgue measurable subset of $\mathbb{R}^d$ then there exists a G_δ set H that contains E and satisfies $|H \setminus E| = 0$. Hence we can obtain the Lebesgue σ-algebra by "adjoining" sets of measure zero to the Borel σ-algebra: A set E belongs to $\mathcal{L}_{\mathbb{R}^d}$ if and only if can be written as $E = H \cup Z$ where $H \in \mathcal{B}_{\mathbb{R}^d}$ and $|Z| = 0$.

The Borel subsets of $\mathbb{R}^d$ are generated from the open subsets of $\mathbb{R}^d$. There is nothing special about $\mathbb{R}^d$ in this regard—whenever we have a space that has a topology, we can define Borel sets in a similar manner. This is stated precisely in the following definition.

Definition 1.1.8 (Borel σ-algebra on X). Let X be a topological space. Then the *Borel σ-algebra $\mathcal{B}_X$ on X* is the smallest σ-algebra that contains all of the open subsets of X. The elements of $\mathcal{B}_X$ are called the *Borel subsets of X*. $\diamondsuit$

We end this section by pointing out a simple but useful fact about countable unions of sets.

Exercise 1.1.9 (The Disjointization Trick). If Σ is a σ-algebra of subsets of X and $E_1, E_2, \ldots \in \Sigma$, then the sets F_k defined by

$$F_1 \;=\; E_1, \quad F_2 \;=\; E_2 \backslash E_1, \quad F_3 \;=\; E_3 \backslash (E_1 \cup E_2), \quad \ldots \tag{1.3}$$

are disjoint, belong to Σ, and satisfy

$$\bigcup_k F_k \;=\; \bigcup_k E_k. \qquad \diamondsuit$$

This disjointization trick can be used to give some equivalent characterizations of σ-algebras (see Problem 1.1.13).

Problems

1.1.10. Suppose that $\mathcal{E},\, \mathcal{F} \subseteq \mathcal{P}(X)$. Show that if $\mathcal{E} \subseteq \Sigma(\mathcal{F})$ then $\Sigma(\mathcal{E}) \subseteq \Sigma(\mathcal{F})$.

1.1.11. Let (X, Σ) be a measure space. Given $Y \subseteq X$, set $\Sigma_Y = \{E \cap Y : E \in \Sigma\}$. Show that Σ_Y is a σ-algebra on Y.

1.1.12. Show that $\Sigma = \{E \subseteq \mathbb{R}^d : |E| = 0 \text{ or } |\mathbb{R}^d \backslash E| = 0\}$ is a σ-algebra on $\mathbb{R}^d$.

1.1.13. An *algebra* (or *field*) is a nonempty collection $\mathcal{A}$ of subsets of X that is closed under complements and *finite* unions.

(a) Show that an algebra $\mathcal{A}$ is a σ-algebra if and only if $\mathcal{A}$ is closed under countable disjoint unions.

(b) Show that an algebra $\mathcal{A}$ is a σ-algebra if and only if $\mathcal{A}$ is closed under countable increasing unions.

1.1.14. Let $\mathcal{B}_{\mathbb{R}}$ be the Borel σ-algebra on $\mathbb{R}$. Show that each of the following collections generates $\mathcal{B}_{\mathbb{R}}$.

(a) $\mathcal{E}_1 = \{(a, b) : a < b\}$.

(b) $\mathcal{E}_2 = \{[a, b] : a < b\}$.

(c) $\mathcal{E}_3 = \{[a, b) : a < b\}$.

(d) $\mathcal{E}_4 = \{(a, \infty) : a \in \mathbb{R}\}$.

(e) $\mathcal{E}_5 = \{[a,\infty) : a \in \mathbb{R}\}$.

(f) $\mathcal{E}_6 = \{(-\infty,a) : a \in \mathbb{R}\}$.

(g) $\mathcal{E}_7 = \{(-\infty,a] : a \in \mathbb{R}\}$.

(h) $\mathcal{E}_8 = \{(r,\infty) : r \in \mathbb{Q}\}$.

(i) $\mathcal{E}_9 = \{(-\infty,r) : r \in \mathbb{Q}\}$.

1.1.15. Let $\mathcal{B}_{\mathbb{R}^2}$ be the Borel σ-algebra on $\mathbb{R}^2$. Show that each of the following collections generates $\mathcal{B}_{\mathbb{R}^2}$.

(a) $\mathcal{E}_1 = \{(a,b) \times (c,d) : a < b, c < d\}$.

(b) $\mathcal{E}_2 = \{(a,b) \times \mathbb{R} : a < b\} \cup \{\mathbb{R} \times (c,d) : c < d\}$.

1.1.16.* Show that any σ-algebra Σ that contains infinitely many subsets of X must be uncountable.

1.1.17. Given a subset A of a set X, show that

$$\Sigma \;=\; \{S : S \subseteq A\} \;\cup\; \{X \setminus S : S \subseteq A\}$$

is a σ-algebra on X.

1.1.18. Suppose that $\Sigma_1 \subseteq \Sigma_2 \subseteq \cdots$ are nested increasing σ-algebras on X. Must $\Sigma = \cup \Sigma_k$ be a σ-algebra?

1.1.19. Let C be the Cantor set, let φ the Cantor–Lebesgue function, and set $g(x) = \varphi(x) + x$ for $x \in [0,1]$.

(a) Prove that $g\colon [0,1] \to [0,2]$ and its inverse $h = g^{-1}\colon [0,2] \to [0,1]$ are each continuous, strictly increasing bijections.

(b) Show that $g(C)$ is a closed subset of $[0,2]$, and $|g(C)| = 1$.

(c) Since $g(C)$ has positive measure, it contains a nonmeasurable set N. Show that $A = h(N)$ is a Lebesgue measurable subset of $[0,1]$. (Note that $N = h^{-1}(A)$ is not measurable, so this shows that the inverse image of a Lebesgue measurable set under a continuous function need not be Lebesgue measurable.)

(d) Show that, although A is Lebesgue measurable, it is not a Borel set.

(e) Set $f = \chi_A$. Prove that $f \circ h$ is not a Lebesgue measurable function, even though f is Lebesgue measurable and h is continuous.

1.2 Measures

Now we turn to measures themselves. Here is the abstract definition of a measure.

Definition 1.2.1 (Measure). Let (X,Σ) be a measurable space. A function $\mu\colon \Sigma \to [0,\infty]$ is a *measure* on (X,Σ) if the following two statements hold:

(a) $\mu(\emptyset) = 0$, and

(a) μ is *countably additive*, i.e.,

$$E_1, E_2, \dots \in \Sigma \text{ are disjoint} \implies \mu\Bigl(\bigcup_k E_k\Bigr) = \sum_k \mu(E_k). \tag{1.4}$$

In this case we say that (X, Σ, μ) is a *measure space*. We refer to the elements of Σ as the *μ-measurable subsets* of X. If the measure μ is clear from context, then we may simply call them the *measurable subsets* of X. The number $\mu(E)$ is the *μ-measure* or simply the *measure* of $E \in \Sigma$. ♢

Remark 1.2.2. (a) We use the phrases "(X, Σ, μ) is a measure space" and "μ is a measure on (X, Σ)" interchangeably. If the σ-algebra Σ is understood then we often simply write "μ is a measure on X."

(b) To avoid multiplicities of parentheses and braces, we usually write $\mu\{x\}$ instead of $\mu(\{x\})$. Similarly, if μ is a measure on $\mathbb{R}$ then we usually write $\mu[a, b)$ instead of $\mu([a, b))$, and so forth.

(c) Since $\mu(E_k) \ge 0$ for every k, the series $\sum \mu(E_k)$ that appears in equation (1.4) always exists in the extended real sense, i.e., it either converges to a finite nonnegative number, or it is ∞.

(d) Countable additivity implies finite additivity, but Problem 1.3.14 shows that finite additivity need not imply countable additivity. ♢

We introduce some terminology for measures with special properties.

Definition 1.2.3. Let (X, Σ, μ) be a measure space.

(a) If $\mu(X) < \infty$ then we say that μ is a *bounded measure* or a *finite measure*, and in this case we call (X, Σ, μ) a *finite measure space*.

(b) A measure that is not bounded is called an *unbounded measure*. In this case we say that (X, Σ, μ) is an *infinite measure space*.

(c) If there exist countably many sets $E_1, E_2, \dots \in \Sigma$ such that $\mu(E_k) < \infty$ for every k and $X = \cup E_k$, then μ is a *σ-finite measure*.

(d) If every set $E \in \Sigma$ with $\mu(E) = \infty$ has a subset $F \subseteq E$ such that $F \in \Sigma$ and $0 < \mu(F) < \infty$, then μ is a *semifinite measure*. ♢

Suppose that μ is a measure on a measurable space (X, Σ), and Σ' is a σ-algebra contained in Σ. (We could call Σ' a sub-σ-algebra, but that terminology seems rather awkward, so we will usually avoid it.) In this case, $\mu|_{\Sigma'}$ (μ restricted to Σ') is a measure on (X, Σ'). Technically, μ and $\mu|_{\Sigma'}$ are two different measures, but often we refer to the restriction of μ as "μ on Σ'," or even just as μ. Lebesgue measure is a typical illustration of this situation, as we describe next.

Example 1.2.4 (Lebesgue Measure). We proved in Volume 1 that Lebesgue measure is a measure on the measurable space $(\mathbb{R}^d, \mathcal{L}_{\mathbb{R}^d})$. Many texts assign a symbol such as m or λ to represent Lebesgue measure, in which case we would denote the Lebesgue measure of $E \in \mathcal{L}_{\mathbb{R}^d}$ by $m(E)$ or $\lambda(E)$. However, for our purposes it is usually more convenient to simply write $|E|$ for the Lebesgue measure of E, and to speak of "Lebesgue measure" without assigning a specific symbol to represent this measure. In those cases where it is convenient to have a symbol for Lebesgue measure, we write "Lebesgue measure dx."

Lebesgue measure is a σ-finite measure on $\mathbb{R}^d$ that is not a bounded measure (because $|\mathbb{R}^d| = \infty$). When we deal with Lebesgue measure in this volume, we will usually want to take the σ-algebra to be the Lebesgue σ-algebra. Therefore, when we speak of Lebesgue measure without qualification we assume that the σ-algebra is $\mathcal{L}_{\mathbb{R}^d}$. However, we sometimes need to restrict to a smaller σ-algebra. After the Lebesgue σ-algebra $\mathcal{L}_{\mathbb{R}^d}$, the σ-algebra on $\mathbb{R}^d$ that we encounter most often is the Borel σ-algebra $\mathcal{B}_{\mathbb{R}^d}$. We refer to the restriction of Lebesgue measure to the Borel subsets of $\mathbb{R}^d$ as *Lebesgue measure on $\mathcal{B}_{\mathbb{R}^d}$*. $\diamondsuit$

Here are some new examples of measures.

Exercise 1.2.5 (The Delta Measure). Let X be a set, and fix an element $a \in X$. For each set $E \subseteq X$ define

$$\delta_a(E) \;=\; \begin{cases} 1, & a \in E, \\ 0, & a \notin E. \end{cases}$$

Show that δ_a is a bounded measure with respect to the σ-algebra $\mathcal{P}(X)$ (and hence also with respect to every σ-algebra on X). $\diamondsuit$

The measure δ_a has many names, including the *δ or delta measure at a*, the *Dirac measure at a*, and the *point mass at a*. If X is a vector space and $a = 0$ then we often use the shorthand $\delta = \delta_0$ for the point mass at the origin, i.e.,

$$\delta(E) \;=\; \begin{cases} 1, & 0 \in E, \\ 0, & 0 \notin E. \end{cases}$$

If we compare Lebesgue measure on $\mathbb{R}^d$ with a delta measure δ_a where $a \in \mathbb{R}^d$, we see many differences, such as the following.

- δ_a is defined on every subset of $\mathbb{R}^d$ (the σ-algebra is $\mathcal{P}(\mathbb{R}^d)$).
- δ_a is a bounded measure: $\delta_a(\mathbb{R}^d) = 1$.
- Singletons can have nonzero measure: $\delta_a\{a\} = 1$.
- Sets with infinite Lebesgue measure can have zero measure with respect to δ_a: $\delta_a(\mathbb{R}^d \backslash \{a\}) = 0$.
- δ_a is not translation-invariant: $\delta_a(E)$ need not equal $\delta_a(E + h)$.

Next we define a measure that is quite different from either Lebesgue measure or a delta measure. In this exercise the notation $\#E$ means the number of elements in the finite set E.

Exercise 1.2.6 (Counting Measure). Let X be a set. For each set $E \subseteq X$ define

$$\mu(E) \;=\; \begin{cases} \#E, & E \text{ finite}, \\ \infty, & E \text{ infinite}. \end{cases}$$

Show that μ is an unbounded measure with respect to the σ-algebra $\mathcal{P}(X)$ (and hence also with respect to every σ-algebra on X). This measure is called *counting measure* on X. $\diamondsuit$

Focusing on counting measure on $\mathbb{R}^d$, here are some ways in which counting measure is similar to Lebesgue measure or a delta measure, but also more ways in which it is different.

- μ is defined on every subset of $\mathbb{R}^d$ (the σ-algebra is $\mathcal{P}(\mathbb{R}^d)$).
- μ is unbounded and is not σ-finite, although it is semifinite.
- Only finite sets have finite measure with respect to μ.
- μ is translation-invariant.

Since the basic definition of σ-algebras revolves around countable unions, it should not be surprising that it is often unpleasant to work with measures that are not σ-finite. Fortunately, most of the measures that we encounter in practice are σ-finite.

Remark 1.2.7. Although counting measure on $\mathbb{R}^d$ is not σ-finite, counting measure on a countable set such as $\mathbb{N}$ or $\mathbb{Z}^d$ is σ-finite. In fact, this is an important measure because we will see later that integration with respect to counting measure is related to infinite series. For example, we will eventually see that if μ is counting measure on the natural numbers $\mathbb{N}$ then the integral of a function $f \colon \mathbb{N} \to [0,\infty]$ with respect to μ is

$$\int_{\mathbb{N}} f \, d\mu \;=\; \sum_{n=1}^{\infty} f(n).$$

Thus, results about series are just special cases of results about integration with respect to a measure. $\diamondsuit$

As we have seen, there can be many different measures on a given space. In Volume 1 we focused on Lebesgue measure on $\mathbb{R}^d$. There are many other measures on $\mathbb{R}^d$, but we adopt the following convention that Lebesgue measure is "default" measure on $\mathbb{R}^d$.

Notation 1.2.8 (Default Measure on $\mathbb{R}^d$). Unless we specifically state otherwise, whenever we deal with $\mathbb{R}^d$ we will implicitly assume that we have chosen to work with Lebesgue measure. Further, unless specifically stated otherwise, we will assume that the σ-algebra associated with Lebesgue measure is the Lebesgue σ-algebra $\mathcal{L}_{\mathbb{R}^d}$. For example, using this convention, the phrase "let E be a measurable subset of $\mathbb{R}^d$" is interpreted to mean that E is a subset of $\mathbb{R}^d$ that is measurable with respect to Lebesgue measure (i.e., E belongs to $\mathcal{L}_{\mathbb{R}^d}$). $\diamondsuit$

Here is one way that we can derive new measures on $\mathbb{R}^d$ from Lebesgue measure.

Exercise 1.2.9. Assume that $f\colon \mathbb{R}^d \to [0,\infty]$ is a Lebesgue measurable function, and define $\mu_f(E) = \int_E f(x)\,dx$ for $E \in \mathcal{L}_{\mathbb{R}^d}$.

(a) Prove that μ_f is a measure on $(\mathbb{R}^d, \mathcal{L}_{\mathbb{R}^d})$.

(b) Show that μ_f is bounded if and only if $f \in L^1(\mathbb{R}^d)$.

(c) For which f is μ_f translation-invariant?

(d) Implicitly restricting δ to the Lebesgue σ-algebra $\mathcal{L}_{\mathbb{R}^d}$, is there any function f such that $\mu_f = \delta$? Is there any f such that μ_f is counting measure on $\mathbb{R}^d$? $\diamondsuit$

Problems

1.2.10. (a) Prove that a nonnegative finite linear combination of measures is a measure, i.e., if $\mu_1, \dots, \mu_N$ are measures on (X, Σ) and $c_1, \dots, c_N \ge 0$, then $\mu = \sum_{n=1}^N c_k \mu_k$, $E \in \Sigma$, defines a measure on (X, Σ).

(b) For each $k \in \mathbb{N}$ let μ_k be a measure on (X, Σ). What conditions (if any) on the extended real numbers c_k are needed so that $\mu(E) = \sum_{n=1}^\infty c_k \mu_k(E)$ defines a measure on (X, Σ)?

1.2.11. Let (X, Σ, μ) be a measure space, and let $Y \in \Sigma$ be fixed. Show that $\mu_Y\colon \Sigma \to [0,\infty]$ defined by $\mu_Y(E) = \mu(E \cap Y)$ is a measure on (X, Σ).

1.2.12. Let μ be a measure on $\big(\mathbb{N}, \mathcal{P}(\mathbb{N})\big)$. Show that μ is completely determined by the values $(\mu\{k\})_{k\in\mathbb{N}}$. In other words, show that $\mu \mapsto (\mu\{k\})_{k\in\mathbb{N}}$ is an injective map of the measures on $\big(\mathbb{N}, \mathcal{P}(\mathbb{N})\big)$ into the space of sequences of extended real numbers. Identify the range of this map, and identify the sequences that correspond to bounded measures on $\mathbb{N}$.

1.2.13. (a) Given an infinite set X, for each $E \subseteq X$ define $\mu(E) = 0$ if E is finite, and $\mu(E) = \infty$ if E is infinite. Show that μ is finitely additive but is not countably additive.

(b) What if we define $\mu(\emptyset) = 0$ and $\mu(E) = \infty$ for every nonempty $E \subseteq X$?

1.2.14. Let X be an uncountable set. Let Σ consist of all subsets A of X such that either A or $X \setminus A$ is at most countable. Given $E \in \Sigma$, define $\mu(E) = 0$ if E is countable, and $\mu(E) = 1$ otherwise. Show that Σ is a σ-algebra and μ is a measure on (X, Σ).

1.3 Basic Properties of Measures

This section presents some properties of abstract measures that are analogous to properties of Lebesgue measure that we encountered in Volume 1. As most of the proofs of these properties are almost identical to those for Lebesgue measure, we state these results as exercises.

Exercise 1.3.1. Given a measure space (X, Σ, μ), prove the following statements.

(a) Monotonicity: If A, $B \in \Sigma$ and $A \subseteq B$ then $\mu(A) \le \mu(B)$.

(b) If A, $B \in \Sigma$, $A \subseteq B$, and $\mu(A) < \infty$ then $\mu(B \setminus A) = \mu(B) - \mu(A)$. ♢

Remark 1.3.2. In Chapter 3 we will study *signed measures*, which satisfy countable additivity but allow the measure to take values in the range $-\infty \le \mu(E) \le \infty$. An important difference between *measures* and *signed measures* is that monotonicity need not hold for signed measures! If μ is a signed measure and $A \subseteq B$ then, since $\mu(B \setminus A)$ might be negative, we cannot infer from $\mu(A) + \mu(B \setminus A) = \mu(B)$ that $\mu(A) \le \mu(B)$. ♢

Since *measures* are monotonic, we immediately obtain the following corollary of Exercise 1.3.1.

Corollary 1.3.3. *Let (X, Σ, μ) be a measure space, and suppose that $E \in \Sigma$ satisfies $\mu(E) = 0$. If $A \subseteq E$ and $A \in \Sigma$, then $\mu(A) = 0$.* □

Thus, all of the *measurable* subsets of a zero measure set have zero measure. In general, however, a set with zero measure may contain subsets that are not measurable. We give the following name to measures that have the property that *every* subset of a measurable set with zero measure are measurable.

Definition 1.3.4 (Complete Measure). Let (X, Σ, μ) be a measure space. we say that μ is *complete* if

$$E \in \Sigma,\ \mu(E) = 0 \quad \Longrightarrow \quad A \in \Sigma \text{ for all } A \subseteq E. \qquad \diamondsuit$$

The reader should be aware that the terms "complete" and "algebra" are heavily overused in mathematics and appear in many unrelated definitions and contexts.

Example 1.3.5. If Z is a Lebesgue measurable subset of $\mathbb{R}^d$ that has measure zero, then every subset of Z is Lebesgue measurable. Thus Lebesgue measure is complete, although we should emphasize that we are (as usual) implicitly taking the σ-algebra to be $\mathcal{L}_{\mathbb{R}^d}$. If we change the σ-algebra, then Lebesgue measure restricted to this new σ-algebra may not be complete. For example, consider Lebesgue measure restricted to the Borel σ-algebra $\mathcal{B}_{\mathbb{R}^d}$. If $Z \in \mathcal{B}_{\mathbb{R}^d}$ is a Borel set and $|Z| = 0$, then it is possible that Z may contain a subset A that is not a Borel set. This set A is not measurable *with respect to* $\mathcal{B}_{\mathbb{R}^d}$, so Lebesgue measure is not a complete measure on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. $\diamondsuit$

We will consider completeness of measures in more detail in Section 1.4. For now, we give a name to those sets E that have zero measure.

Notation 1.3.6 (Null Sets). Given a measure space (X, Σ, μ), a set $E \in \Sigma$ such that $\mu(E) = 0$ is called a *μ-null set*, a set with *μ-measure zero*, or simply a *zero measure set*.

A property that holds for all $x \in X$ except possibly for x in a μ-null set E is said to hold *μ-almost everywhere* (abbreviated μ-a.e.). $\diamondsuit$

We often omit writing the symbol μ if it is clear from context, e.g., we may simply say that E is a *null set* instead writing *μ-null set*, or say that a property holds *almost everywhere* instead of *μ-almost everywhere*.

Remark 1.3.7. Null sets can be "very large" in senses other than their measure. For example, for the δ-measure on $\mathbb{R}^d$ we have $\delta(\mathbb{R}^d \setminus \{0\}) = 0$. Thus $\mathbb{R}^d \setminus \{0\}$ is a δ-null set, and consequently $\chi_{\mathbb{R}^d \setminus \{0\}} = 0$ δ-a.e.

Abstract measures satisfy continuity from above and below in the following sense.

Exercise 1.3.8. Let (X, Σ, μ) be a measure space. Given measurable sets $E_1, E_2, \dots$ in X, prove that the following statements hold.

(a) Continuity from below: If $E_1 \subseteq E_2 \subseteq \cdots$, then

$$\mu\biggl(\bigcup_{k=1}^{\infty} E_k\biggr) \;=\; \lim_{k \to \infty} \mu(E_k).$$

(b) Continuity from above: If $E_1 \supseteq E_2 \supseteq \cdots$ and $\mu(E_k) < \infty$ for some k, then

$$\mu\biggl(\bigcap_{k=1}^{\infty} E_k\biggr) \;=\; \lim_{k \to \infty} \mu(E_k). \qquad \diamondsuit$$

The final property that we will discuss in this section is countable subadditivity. Here we require a different approach than we took when we considered Lebesgue measure. When we constructed Lebesgue measure, we first constructed exterior Lebesgue measure, which is subadditive, and then restricted to the Lebesgue measurable sets to obtain Lebesgue measure. Hence

Lebesgue measure simply inherits subadditivity from exterior Lebesgue measure. With Lebesgue measure, the difficulty was in proving that countable additivity holds on the Lebesgue measurable sets. In contrast, an abstract measure μ is countably additive by definition, and we must *deduce* countable subadditivity from that hypothesis.

Theorem 1.3.9 (Countable Subadditivity). *Let (X, Σ, μ) be a measure space. If $E_1, E_2, \dots \in \Sigma$, then*

$$\mu\biggl(\bigcup_k E_k\biggr) \;\le\; \sum_k \mu(E_k).$$

Proof. Using the disjointization trick (Exercise 1.1.9), we can write $\cup E_k = \cup F_k$ where the sets F_k are measurable and disjoint, and $F_k \subseteq E_k$ for each k. Combining countable additivity and monotonicity, it follows that

$$\mu\biggl(\bigcup_k E_k\biggr) \;=\; \mu\biggl(\bigcup_k F_k\biggr) \;=\; \sum_k \mu(F_k) \;\le\; \sum_k \mu(E_k). \quad \square$$

Problems

1.3.10. Show that every σ-finite measure is semifinite.

1.3.11. Suppose that (X, Σ, μ) is a measure space and $\{x\} \in \Sigma$ for every $x \in X$. Show that if μ is a finite measure, then $E = \{x \in X : \mu\{x\} > 0\}$ is countable.

1.3.12. Let (X, Σ, μ) be a measure space and fix $E_k \in \Sigma$ for $k \in \mathbb{N}$. Define

$$\limsup E_k \;=\; \bigcap_{j=1}^{\infty}\biggl(\bigcup_{k=j}^{\infty} E_k\biggr), \qquad \liminf E_k \;=\; \bigcup_{j=1}^{\infty}\biggl(\bigcap_{k=j}^{\infty} E_k\biggr).$$

(a) Show that

$$\mu\Bigl(\liminf E_k\Bigr) \;\le\; \liminf \mu(E_k).$$

(b) Prove that if $\mu(\cup E_k) < \infty$, then

$$\mu\Bigl(\limsup E_k\Bigr) \;\ge\; \limsup \mu(E_k).$$

1.3.13. Let (X, Σ, μ) be a measure space. Show that if μ is semifinite, $E \in \Sigma$, and $\mu(E) = \infty$, then for any $C > 0$ there exists some measurable set $A \subseteq E$ that satisfies $C < \mu(A) < \infty$.

1.3.14. Suppose that μ is a *finitely* additive function on a σ-algebra Σ of subsets of X, i.e., $\mu(\emptyset) = 0$ and $\mu\bigl(\cup_{k=1}^N E_k\bigr) = \sum_{k=1}^N \mu(E_k)$ for any finite collection of disjoint sets $E_1, \dots, E_N \in \Sigma$. Prove the following statements.

(a) μ is a measure if and only if it satisfies continuity from below.

(b) If $\mu(X) < \infty$, then μ is a measure if and only if it satisfies continuity from above.

1.4 The Completion of a Measure

By definition, a set $E \subseteq X$ is a null set for a measure μ on X if $E \in \Sigma$ and $\mu(E) = 0$. In general, an arbitrary subset A of E need not be measurable, but if A happens to be measurable then monotonicity implies that $\mu(A) = 0$. A *complete measure* is one such that every subset A of every null set E is measurable (Definition 1.3.4).

Complete measures are often more convenient to work with than incomplete measures. Fortunately, if we have a incomplete measure μ in hand, there is a way to extend μ to a larger σ-algebra $\overline{\Sigma}$ in such a way that the extended measure is complete. This new extended measure $\overline{\mu}$ is called the *completion* of μ, and its construction is given in the next exercise.

Exercise 1.4.1. Let (X, Σ, μ) be a measure space, and let $\mathcal{N}$ be the collection of all μ-null sets in X:

$$\mathcal{N} \;=\; \bigl\{N \in \Sigma : \mu(N) = 0\bigr\}.$$

Define

$$\overline{\Sigma} \;=\; \{E \cup Z : E \in \Sigma,\, Z \subseteq N \in \mathcal{N}\},$$

and prove the following statements.

(a) $\overline{\Sigma}$ is a σ-algebra on X.

(b) For each set $E \cup Z \in \overline{\Sigma}$, define

$$\overline{\mu}(E \cup Z) \;=\; \mu(E).$$

Then $\overline{\mu}$ is a well-defined function on $\overline{\Sigma}$.

(c) $\overline{\mu}$ is a measure on $(X, \overline{\Sigma})$.

(d) $\overline{\mu}$ is the *unique* measure on $(X, \overline{\Sigma})$ that coincides with μ on Σ.

(e) $\overline{\mu}$ is complete. $\diamondsuit$

Example 1.4.2. Let $\mathcal{B}_{\mathbb{R}^d}$ be the Borel σ-algebra on $\mathbb{R}^d$, and let μ be Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. Since every open subset of $\mathbb{R}^d$ is Lebesgue measurable, $\mathcal{B}_{\mathbb{R}^d}$ is contained in the σ-algebra $\mathcal{L}_{\mathbb{R}^d}$ of Lebesgue measurable subsets of $\mathbb{R}^d$. Using results from Volume 1, the σ-algebra $\overline{\mathcal{B}_{\mathbb{R}^d}}$ constructed in Exercise 1.4.1 is precisely $\mathcal{L}_{\mathbb{R}^d}$, and $\overline{\mu}$ is Lebesgue measure $|\cdot|$ on $(\mathbb{R}^d, \mathcal{L}_{\mathbb{R}^d})$. $\diamondsuit$

Example 1.4.3. Consider the δ-measure as a measure on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. In this case
$$\overline{\mathcal{B}_{\mathbb{R}^d}} = \mathcal{P}(\mathbb{R}^d),$$
and $\overline{\delta} = \delta$ on $(\mathbb{R}^d, \mathcal{P}(\mathbb{R}^d))$. $\diamondsuit$

1.5 Outer Measures

In Volume 1 we created exterior Lebesgue measure by beginning with a class of sets whose measures were known (boxes), followed by an extension to arbitrary subsets of $\mathbb{R}^d$ via countable coverings by boxes. We could not do this in a "good" way for all sets, so after constructing exterior Lebesgue measure on the measure space $(\mathbb{R}^d, \mathcal{P}(\mathbb{R}^d))$, we restricted to the measure space $(\mathbb{R}^d, \mathcal{L}_{\mathbb{R}^d})$ to obtain Lebesgue measure, which is countably additive on this smaller measure space.

In contrast, in this chapter we began with an abstract measure that, by definition, satisfies countable additivity on an associated σ-algebra. This leaves us with a fundamental question: How can we construct such abstract measures? In this this section and the next we will show that we can create measures through a process similar to the one that we used to construct Lebesgue measure from exterior Lebesgue measure. Specifically, given a class of subsets of X that we know how we want to measure, we will construct an exterior or outer measure μ^* that is defined on all subsets of X but is not a true measure, and then construct an actual measure μ by restricting μ^* to an appropriate σ-algebra Σ of "measurable sets."

It will be easier if we break this process into two parts:

(i) the construction of an outer measure μ^*, and

(ii) the construction of a measure μ from the outer measure μ^*.

We will address the second item in this section, showing how to obtain a measure μ from an outer measure μ^*. In the next section we will consider the issue of constructing an outer measure μ^* from scratch.

Our first task is to precisely define outer measures. Considering the example of exterior Lebesgue measure, it seems that the most important requirements are that an outer measure μ^* should be defined on all subsets of X and that it should satisfy countable subadditivity. However, there is another important but hidden property, which is *monotonicity*. Although *countable additivity* implies monotonicity, *countable subadditivity* does not. Therefore we need to include monotonicity as part of the definition of an outer measure.

Definition 1.5.1 (Outer Measure). Let X be a nonempty set. An *outer measure* or *exterior measure* on X is a function $\mu^*\colon \mathcal{P}(X) \to [0,\infty]$ that satisfies the following conditions.

(a) $\mu^*(\emptyset) = 0$.

(b) *Monotonicity*: If $A \subseteq B$, then $\mu^*(A) \le \mu^*(B)$.

(c) *Countable subadditivity*: If $E_1, E_2, \ldots \subseteq X$, then

$$\mu^*\Bigl(\bigcup_k E_k\Bigr) \;\le\; \sum_k \mu^*(E_k). \qquad \diamondsuit$$

Given an outer measure μ^*, our goal is to create a σ-algebra Σ on X such that μ^* restricted to Σ is countably additive. The elements of Σ will be our "good sets," the sets that are measurable with respect to μ^*. But how do we define measurability for an arbitrary outer measure? We might not be given a topology on X, so we cannot define measurability in terms of surrounding open sets, as we did for Lebesgue measure. On the other hand, the formulation of Lebesgue measurability given by Carathéodory's Criterion does not involve topology, and as such it is the appropriate motivation for the following definition.

Definition 1.5.2 (Measurable Set). Let μ^* be an outer measure on a set X. Then a set $E \subseteq X$ is μ^*-*measurable*, or simply *measurable* for short, if

$$\forall\, A \subseteq X, \quad \mu^*(A) \;=\; \mu^*(A \cap E) + \mu^*(A \backslash E). \qquad \diamondsuit$$

Remark 1.5.3. (a) It is often helpful to recall that $\mu^*(A \backslash E) = \mu^*(A \cap E^{\mathrm{C}})$, so we can equivalently write the condition for measurability as

$$\forall\, A \subseteq X, \quad \mu^*(A) \;=\; \mu^*(A \cap E) + \mu^*(A \cap E^{\mathrm{C}}).$$

(b) The empty set is μ^*-measurable by virtue of the fact that $\mu^*(\emptyset) = 0$.

(c) By subadditivity, we always have the inequality

$$\mu^*(A) \;\le\; \mu^*(A \cap E) + \mu^*(A \backslash E).$$

Hence, to establish that a set E is measurable, we just have to prove that the opposite inequality holds for every set $A \subseteq X$. $\diamondsuit$

The next exercise, which follows by combining subadditivity with monotonicity, asks for a proof that every subset of X that has outer measure zero is measurable.

Exercise 1.5.4. Let μ^* be an outer measure on X. Show that if $E \subseteq X$ and $\mu^*(E) = 0$ then E is μ^*-measurable. $\diamondsuit$

Now we will prove that the collection of μ-measurable sets forms a σ-algebra on X, and μ^* restricted to this σ-algebra is a complete measure on X.

Theorem 1.5.5 (Carathéodory's Theorem). *If μ^* is an outer measure on a set X, then the following statements hold.*

(a) $\Sigma = \bigl\{E \subseteq X : E$ *is* μ^**-measurable*$\bigr\}$ *is a* σ*-algebra on* X.

(b) $\mu = \mu^*|_\Sigma$ *is a measure on* (X, Σ).

(c) μ *is a complete measure. In fact,* every *set* $E \subseteq X$ *that satisfies* $\mu^*(E) = 0$ *is* μ^**-measurable.*

Proof. (a) We break the proof into a series of steps.

Step 1. Σ is not empty since the empty set is μ^*-measurable.

Step 2. To show that Σ is closed under complements, fix any set $E \in \Sigma$ and let A be an arbitrary subset of X. Using the fact that E is μ^*-measurable, we compute that

$$\begin{aligned}
\mu^*(A) &= \mu^*(A \cap E) + \mu^*(A \setminus E) \\
&= \mu^*(A \cap E) + \mu^*(A \cap E^{\mathrm{C}}) \\
&= \mu^*(A \cap E^{\mathrm{C}}) + \mu^*\bigl(A \cap (E^{\mathrm{C}})^{\mathrm{C}}\bigr) \\
&= \mu^*(A \cap E^{\mathrm{C}}) + \mu^*(A \setminus E^{\mathrm{C}}).
\end{aligned}$$

Hence E^{C} is μ^*-measurable, so $E^{\mathrm{C}} \in \Sigma$.

Step 3. Ultimately we want to show that Σ is closed under countable unions, but to begin we will show that it is closed under finite unions. By induction, to prove this it suffices to show that if $E, F \in \Sigma$ then $E \cup F \in \Sigma$. Choose any set $A \subseteq X$. By subadditivity,

$$\mu^*(A) \;\le\; \mu^*\bigl(A \cap (E \cup F)\bigr) + \mu^*\bigl(A \cap (E \cup F)^{\mathrm{C}}\bigr).$$

To prove the opposite inequality, note that since F is μ^*-measurable and $A \cap E$ is a subset of X, we have

$$\mu^*(A \cap E \cap F) + \mu^*(A \cap E \cap F^{\mathrm{C}}) \;=\; \mu^*(A \cap E).$$

Similarly,

$$\mu^*(A \cap E^{\mathrm{C}} \cap F) + \mu^*(A \cap E^{\mathrm{C}} \cap F^{\mathrm{C}}) \;=\; \mu^*(A \cap E^{\mathrm{C}}).$$

Applying these equalities and using the fact that μ^* is finitely subadditive, we see that

$$\begin{aligned}
&\mu^*\bigl(A \cap (E \cup F)\bigr) \;+\; \mu^*\bigl(A \cap (E \cup F)^{\mathrm{C}}\bigr). \\
&= \mu^*\bigl((A \cap E \cap F) \cup (A \cap E \cap F^{\mathrm{C}}) \cup (A \cap E^{\mathrm{C}} \cap F)\bigr) \;+\; \mu^*(A \cap (E \cup F)^{\mathrm{C}}) \\
&\le \mu^*(A \cap E \cap F) + \mu^*(A \cap E \cap F^{\mathrm{C}}) + \mu^*(A \cap E^{\mathrm{C}} \cap F) + \mu^*(A \cap E^{\mathrm{C}} \cap F^{\mathrm{C}})
\end{aligned}$$

$$= \mu^*(A \cap E) \;+\; \mu^*(A \cap E^{\mathrm{C}})$$
$$= \mu^*(A),$$

where at the final equality we have used the fact that E is μ^*-measurable. Therefore $E \cup F \in \Sigma$.

Step 4. Next we will show that μ^* is finitely additive on Σ. Suppose that $E,\, F \in \Sigma$ are *disjoint*. Since E is measurable, we have

$$\mu^*(E \cup F) \;=\; \mu^*\big((E \cup F) \cap E\big) + \mu^*\big((E \cup F) \cap E^{\mathrm{C}}\big) \;=\; \mu^*(E) + \mu^*(F).$$

By induction, it follows that μ^* is finitely additive when we choose sets from Σ. However, μ^* need not be finitely additive on arbitrary subsets of X!

Step 5. Now we will show that Σ is closed under countable unions. By Problem 1.1.13, it suffices to show that Σ is closed under countable *disjoint* unions. So, assume that $E_1, E_2, \ldots \in \Sigma$ are disjoint sets and define

$$F \;=\; \bigcup_{k=1}^{\infty} E_k \qquad \text{and} \qquad F_n \;=\; \bigcup_{k=1}^{n} E_k, \quad n \in \mathbb{N}.$$

Note that $F_n \in \Sigma$ since Σ is closed under finite unions.

Choose any set $A \subseteq X$. We claim that

$$\mu^*(A \cap F_n) \;=\; \sum_{k=1}^{n} \mu^*(A \cap E_k). \tag{1.5}$$

Now, if we knew that μ^* was finitely additive on all subsets of X then equation (1.5) would be immediate. However, we only know that μ^* is finitely additive on the μ^*-measurable sets. Since A is an arbitrary set, this does not help us. Instead, we prove equation (1.5) by induction.

Since $F_1 = E_1$, equation (1.5) is trivial when $n = 1$. Suppose that equation (1.5) holds for some integer $n \geq 1$. Then, since E_{n+1} is μ^*-measurable,

$$\begin{aligned}
\mu^*(A \cap F_{n+1}) \;&=\; \mu^*\Big(A \cap \bigcup_{k=1}^{n+1} E_k\Big) \\
&=\; \mu^*\Big(A \cap \bigcup_{k=1}^{n+1} E_k \,\cap\, E_{n+1}\Big) + \mu^*\Big(A \cap \bigcup_{k=1}^{n+1} E_k \,\cap\, E_{n+1}^{\mathrm{C}}\Big) \\
&=\; \mu^*(A \cap E_{n+1}) + \mu^*(A \cap \bigcup_{k=1}^{n} E_n) \qquad \text{(by disjointness)} \\
&=\; \mu^*(A \cap E_{n+1}) + \sum_{k=1}^{n} \mu^*(A \cap E_k).
\end{aligned}$$

It follows that equation (1.5) holds for all n by induction.

Next we compute that

$$\begin{aligned}
&\sum_{k=1}^{n} \mu^*(A \cap E_k) \;+\; \mu^*(A \cap F^{\mathrm{C}}) \\
&\quad\le\; \sum_{k=1}^{n} \mu^*(A \cap E_k) \;+\; \mu^*(A \cap F_n^{\mathrm{C}}) && (\text{since } F^{\mathrm{C}} \subseteq F_n^{\mathrm{C}}) \\
&\quad=\; \mu^*(A \cap F_n) + \mu^*(A \cap F_n^{\mathrm{C}}) && (\text{by equation (1.5)}) \\
&\quad=\; \mu^*(A) && (\text{since } F_n \in \Sigma). \qquad (1.6)
\end{aligned}$$

Applying subadditivity and taking the limit as $n \to \infty$,

$$\begin{aligned}
\mu^*(A) &\le\; \mu^*(A \cap F) + \mu^*(A \cap F^{\mathrm{C}}) && (\text{subadditivity}) \\
&\le\; \sum_{k=1}^{\infty} \mu^*(A \cap E_k) \;+\; \mu^*(A \cap F^{\mathrm{C}}) && (\text{subadditivity}) \\
&\le\; \mu^*(A) && (\text{by equation (1.6)}).
\end{aligned}$$

Therefore equality holds in the preceding lines:

$$\begin{aligned}
\mu^*(A) &=\; \sum_{k=1}^{\infty} \mu^*(A \cap E_k) \;+\; \mu^*(A \cap F^{\mathrm{C}}) \qquad (1.7) \\
&=\; \mu^*(A \cap F) + \mu^*(A \cap F^{\mathrm{C}}),
\end{aligned}$$

so $F \in \Sigma$.

(b) To show that μ^* *restricted to* Σ is countably additive, let $E_1, E_2, \dots$ be disjoint sets in Σ, and set $F = \cup E_k$. Then since F is μ^*-measurable, by applying equation (1.7) with $A = F$ we see that

$$\mu^*(F) \;=\; \sum_{k=1}^{\infty} \mu^*(F \cap E_k) \;+\; \mu^*(F \cap F^{\mathrm{C}}) \;=\; \sum_{k=1}^{\infty} \mu^*(E_k).$$

Hence μ^* is countably additive on Σ, and therefore $\mu = \mu^*|_\Sigma$ is a measure.

(c) This follows from Exercise 1.5.4. $\square$

Problems

1.5.6. Let μ^* be an outer measure on X, and let $A, B \subseteq X$ be μ^*-measurable. Show that $\mu^*(A \cup B) + \mu^*(A \cap B) \;=\; \mu^*(A) + \mu^*(B)$.

1.5.7. Let μ^* be an outer measure on X. Show that if A, B are disjoint subsets of X and A is μ^*-measurable, then $\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$.

1.5.8. Given an uncountable set X, define $\mu^*(E) = 0$ if $E \subseteq X$ is countable, and $\mu^*(E) = 1$ if $E \subseteq X$ is uncountable, Show that μ^* is an outer measure on X, and identify the μ^*-measurable subsets of X.

1.5.9. Let μ^* be an outer measure on a set X, and let $\{A_j\}_{j=1}^\infty$ be a sequence of disjoint μ^*-measurable subsets of X.

(a) Prove that for any $E \subseteq X$ (measurable or not) and any integer $n > 0$,

$$\mu^*\Big(E \cap \Big(\bigcup_{j=1}^n A_j\Big)\Big) \;=\; \sum_{j=1}^n \mu^*(E \cap A_j).$$

(b) Prove that for any $E \subseteq X$ (measurable or not),

$$\mu^*\Big(E \cap \Big(\bigcup_{j=1}^\infty A_j\Big)\Big) \;=\; \sum_{j=1}^\infty \mu^*(E \cap A_j).$$

1.6 The Construction of an Outer Measure

Now we want to show now that if we are given a particular class of "elementary sets" $\mathcal{E}$ whose measures are specified, then we can extend this to an outer measure on X. We do this by employing the same technique that we used to create exterior Lebesgue measure, i.e., we cover arbitrary sets by countable unions of elementary sets in all possible ways.

Theorem 1.6.1. *Let $\mathcal{E} \subseteq \mathcal{P}(X)$ be a fixed collection of sets such that*

(a) $\emptyset \in \mathcal{E}$, *and*

(b) *there exist countably many sets $E_k \in \mathcal{E}$ such that $\cup E_k = X$.*

Suppose that $\rho\colon \mathcal{E} \to [0,\infty]$ satisfies $\rho(\emptyset) = 0$. For each $A \subseteq X$, define

$$\mu^*(A) \;=\; \inf\Big\{\sum_k \rho(E_k)\Big\}, \tag{1.8}$$

where the infimum is taken over all finite or countable covers of A by sets $E_k \in \mathcal{E}$. Then μ^* is an outer measure on X.

Proof. The given hypotheses ensure that every subset of X has at least one covering by elements of $\mathcal{E}$. Hence the infimum in equation (1.8) is not taken over the empty set, and therefore defines a value in $[0,\infty]$ for each $A \subseteq X$.

Since $\{\emptyset\}$ is one covering of $\emptyset$ by elements of $\mathcal{E}$, we have

$$0 \le \mu^*(\emptyset) \le \rho(\emptyset) = 0,$$

and monotonicity follows from the fact that if $A \subseteq B$ then every covering of B by sets $E_k \in \mathcal{E}$ is also a covering of A. Finally, the proof that μ^* is countably subadditive is just like the proof that exterior Lebesgue measure is countably subadditive, so we assign this as an exercise. □

We refer to the elements of the collection $\mathcal{E}$ in Theorem 1.6.1 as *elementary sets*. By Theorem 1.5.5, since the function μ^* constructed in Theorem 1.6.1 is an outer measure, we know that there is an associated σ-algebra Σ of μ^*-measurable sets, and we also know that $\mu = \mu^*|_\Sigma$ is a complete measure on (X, Σ). However, there are still two important questions that we have not addressed.

- Are the elementary sets measurable, i.e., do we have $\mathcal{E} \subseteq \Sigma$?
- Is $\mu^*(E) = \rho(E)$ for $E \in \mathcal{E}$?

Unfortunately, the following example shows that the answers to these questions are *no* in general.

Example 1.6.2. Let X be a set that contains at least two elements. Let E be a nonempty proper subset of X, and define $\mathcal{E} = \{\emptyset, E, E^{\mathrm{C}}, X\}$.

(a) If we define

$$\rho(\emptyset) = 0, \qquad \rho(E) = \frac{1}{4}, \qquad \rho(E^{\mathrm{C}}) = \frac{1}{4}, \qquad \rho(X) = 1,$$

then $\mu^*(X) = \frac{1}{2} \neq \rho(X)$, so the outer measure μ^* does not agree with ρ on the elementary sets.

(b) If we define

$$\rho(\emptyset) = 0, \qquad \rho(E) = 1 \qquad \rho(E^{\mathrm{C}}) = 1 \qquad \rho(X) = 1,$$

then

$$\mu^*(X) = 1 \neq 2 = \mu^*(X \cap E) + \mu^*(X \cap E^{\mathrm{C}}),$$

so E is not μ^*-measurable, even though it is an elementary set. ♢

Thus, we need to impose some extra conditions on the function ρ and the class $\mathcal{E}$ of elementary sets.

Definition 1.6.3 (Premeasure). Given a set X, let $\mathcal{A} \subseteq \mathcal{P}(X)$ be an *algebra*, i.e., $\mathcal{A}$ is nonempty and is closed under complements and *finite* unions. A *premeasure* on $\mathcal{A}$ is a function $\mu_0 \colon \mathcal{A} \to [0, \infty]$ satisfying

(a) $\mu_0(\emptyset) = 0$, and

(b) if $E_1, E_2, \dots \in \mathcal{A}$ are disjoint and $\bigcup_{k=1}^{\infty} E_k \in \mathcal{A}$, then

$$\mu_0\Bigl(\bigcup_{k=1}^{\infty} E_k\Bigr) \;=\; \sum_{k=1}^{\infty} \mu_0(E_k). \qquad \diamondsuit$$

Note that in requirement (b) we are not assuming that $\mathcal{A}$ is closed under countable unions. We only require that *if* the union of the disjoint sets A_k belongs to $\mathcal{A}$ *then* μ_0 will be countably additive on those sets. On the other hand, the next lemma shows that every premeasure is *finitely additive* and *monotonic*.

Lemma 1.6.4. *A premeasure μ_0 on an algebra $\mathcal{A}$ is monotonic and finitely additive on $\mathcal{A}$.* $\diamondsuit$

Proof. To show finite additivity, choose any disjoint sets $A_1, \dots, A_N \in \mathcal{A}$. Define $A_k = \emptyset$ for $k > 0$. Then, since $\mathcal{A}$ is closed under *finite unions*, we have

$$\bigcup_{k=1}^{\infty} A_k \;=\; \bigcup_{k=1}^{N} A_k \;\in\; \mathcal{A}.$$

Therefore, by definition of a premeasure,

$$\mu_0\Bigl(\bigcup_{k=1}^{N} A_k\Bigr) \;=\; \mu_0\Bigl(\bigcup_{k=1}^{\infty} A_k\Bigr) \;=\; \sum_{k=1}^{\infty} \mu_0(A_k) \;=\; \sum_{k=1}^{N} \mu_0(A_k).$$

Hence μ^* is finitely additive *on the algebra* $\mathcal{A}$.

Now suppose that $A, B \in \mathcal{A}$ and $A \subseteq B$. Then, by finite additivity,

$$\mu_0(B) \;=\; \mu_0(A) + \mu_0(B \backslash A) \;\geq\; \mu_0(A).$$

Therefore μ_0 is monotonic *on the algebra* $\mathcal{A}$. $\square$

Note that the collection of all boxes in $\mathbb{R}^d$ does not form an algebra, since it is not closed under either complements or finite unions. Thus it is not *entirely* obvious how the construction of Lebesgue measure relates to premeasures. We will consider this issue in detail in Section 1.7.

Given a premeasure μ_0, we associate the outer measure μ^* defined by

$$\mu^*(E) \;=\; \inf\Biggl\{\sum_{k=1}^{\infty} \mu_0(E_k) : E_k \in \mathcal{A},\, E \subseteq \bigcup_k E_k\Biggr\}.$$

Note that $\mathcal{A}$ is nonempty, so there exists some set $A \in \mathcal{A}$, and therefore $A^{\mathrm{C}} \in \mathcal{A}$ as well. Hence $X = A \cup A^{\mathrm{C}}$. This shows that the hypotheses of Theorem 1.6.1 are satisfied, so μ^* is indeed an outer measure. Let Σ denote the corresponding σ-algebra of μ^*-measurable sets. Carathéodory's Theorem implies that $\mu = \mu^*|_{\Sigma}$ is a complete measure. Our next goal is to show that this measure is "well-behaved."

Theorem 1.6.5. *Given a premeasure μ_0 on an algebra $\mathcal{A}$, let μ^* be the associated outer measure. Then the following statements hold.*

(a) $\mu^*|_{\mathcal{A}} = \mu_0$, *i.e.,* $\mu^*(E) = \mu_0(E)$ *for every* $E \in \mathcal{A}$.

(b) $\mathcal{A} \subseteq \Sigma$, *i.e., every set in* $\mathcal{A}$ *is* μ^**-measurable. Consequently,* $\mu(E) = \mu_0(E)$ *for every set* $E \in \mathcal{A}$.

(c) *If* ν *is any measure on* Σ *such that* $\nu|_{\mathcal{A}} = \mu_0$, *then* $\nu(E) \le \mu(E)$ *for all* $E \in \Sigma$, *with equality holding if* $\mu(E) < \infty$. *Furthermore, if* μ_0 *is* σ*-finite, then* $\nu = \mu$.

Proof. (a) Suppose that $E \in \mathcal{A}$. Then $\{E\}$ is a covering of E by a single set from $\mathcal{A}$, so $\mu^*(E) \le \mu_0(E)$. For the converse inequality, let $\{E_k\}$ be any countable collection of sets from $\mathcal{A}$ that covers E. Disjointize these sets by defining

$$F_1 \;=\; E \cap E_1 \qquad \text{and} \qquad F_n \;=\; E \cap \biggl(E_n \setminus \bigcup_{k=1}^{n-1} E_k\biggr), \quad n > 1.$$

Then $F_1, F_2, \ldots \in \mathcal{A}$ and $\cup F_n = E \in \mathcal{A}$, so by the definition of premeasure and the fact that μ_0 is monotonic we have

$$\mu_0(E) \;=\; \sum_{n=1}^{\infty} \mu_0(F_n) \;\le\; \sum_{n=1}^{\infty} \mu_0(E_n).$$

This is true for every covering of E, so $\mu_0(E) \le \mu^*(E)$ and therefore μ^* agrees with μ_0 on $\mathcal{A}$.

(b) Suppose that $E \in \mathcal{A}$ and $A \subseteq X$. If we fix $\varepsilon > 0$, then there exists a countable covering $\{E_k\}$ of A by sets $E_k \in \mathcal{A}$ such that

$$\sum_{k=1}^{\infty} \mu_0(E_k) \;\le\; \mu^*(A) + \varepsilon.$$

Hence,

$$\begin{aligned}
\mu^*(A) \;&\le\; \mu^*(A \cap E) + \mu^*(A \cap E^{\mathrm{C}}) && \text{(subadditivity)} \\
&\le\; \mu^*\biggl(\Bigl(\bigcup_k E_k\Bigr) \cap E\biggr) + \mu^*\biggl(\Bigl(\bigcup_k E_k\Bigr) \cap E^{\mathrm{C}}\biggr) && \text{(monotonicity)} \\
&=\; \mu^*\Bigl(\bigcup_k (E_k \cap E)\Bigr) + \mu^*\Bigl(\bigcup_k (E_k \cap E^{\mathrm{C}})\Bigr) \\
&\le\; \sum_{k=1}^{\infty} \mu^*(E_k \cap E) + \sum_{k=1}^{\infty} \mu^*(E_k \cap E^{\mathrm{C}}) && \text{(subadditivity)} \\
&=\; \sum_{k=1}^{\infty} \Bigl(\mu_0(E_k \cap E) + \mu_0(E_k \cap E^{\mathrm{C}})\Bigr) && \text{(part (a))}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{\infty} \mu_0(E_k) && \text{(countable additivity on } \mathcal{A}) \\
&\leq \mu^*(A) + \varepsilon.
\end{aligned}
$$

Since this is true for every ε, we conclude that

$$\mu^*(A) \;=\; \mu^*(A \cap E) + \mu^*(A \cap E^{\mathrm{C}}),$$

and therefore E is μ^*-measurable.

(c) Suppose that ν is any measure on Σ that extends μ_0, and fix $A \in \Sigma$. If $\{E_k\}$ is a countable cover of A by sets $E_k \in \mathcal{A}$, then

$$\nu(A) \;\leq\; \sum_{k=1}^{\infty} \nu(E_k) \;=\; \sum_{k=1}^{\infty} \mu_0(E_k).$$

Since this is true for every covering, we conclude that $\nu(A) \leq \mu^*(A) = \mu(A)$.

Suppose in addition that $\mu(A) < \infty$. Then given $\varepsilon > 0$ we can find sets $E_k \in \mathcal{A}$ such that $\cup E_k \supseteq A$ and

$$\sum_k \mu_0(E_k) \;\leq\; \mu^*(A) + \varepsilon.$$

Set $E = \cup E_k$. Then

$$
\begin{aligned}
\mu(E) &\leq \sum_k \mu(E_k) && \text{(subadditivity)} \\
&= \sum_k \mu_0(E_k) && \text{(part (a))} \\
&\leq \mu^*(A) + \varepsilon \\
&= \mu(A) + \varepsilon && \text{(since } A \in \Sigma).
\end{aligned}
$$

Since all quantities are finite, we can rearrange and use additivity to conclude that

$$\mu(E \backslash A) \;=\; \mu(E) - \mu(A) \;\leq\; \varepsilon.$$

Now, $\mathcal{A}$ is closed under finite unions, so $\bigcup_{k=1}^{N} E_k \in \mathcal{A}$ for every N. By continuity from below and the fact that μ and ν both extend μ_0, it follows that

$$\nu(E) \;=\; \lim_{N\to\infty} \nu\Bigl(\bigcup_{k=1}^{N} E_k\Bigr) \;=\; \lim_{N\to\infty} \mu_0\Bigl(\bigcup_{k=1}^{N} E_k\Bigr) \;=\; \lim_{N\to\infty} \mu\Bigl(\bigcup_{k=1}^{N} E_k\Bigr) \;=\; \mu(E).$$

Hence

$$\mu(A) \;\leq\; \mu(E) \;=\; \nu(E) \;=\; \nu(A) + \nu(E \backslash A) \;\leq\; \nu(A) + \mu(E \backslash A) \;\leq\; \nu(A) + \varepsilon.$$

Since ε is arbitrary, $\mu(A) = \nu(A)$.

Finally, suppose that μ_0 is σ-finite, i.e., we can write $X = \cup A_k$ where $\mu_0(A_k) < \infty$ for each k. By applying the disjointization trick, we can assume that the sets A_k are disjoint. Then since each A_k has finite measure, we have for any $E \in \Sigma$ that

$$\nu(E) = \sum_{k=1}^{\infty} \nu(E \cap A_k) = \sum_{k=1}^{\infty} \mu(E \cap A_k) = \mu(E). \qquad \diamondsuit$$

Problems

1.6.6. Show that if μ^* is the outer measure induced from a premeasure μ_0, then every set $E \subseteq X$ that satisfies $\mu^*(E) = 0$ is μ^*-measurable. Conclude that $\mu = \mu|_\Sigma$ is a complete measure on (X, Σ).

1.6.7. Let μ_0 be a premeasure on an algebra $\mathcal{A}$ of subsets of X. Suppose μ_0 is bounded ($\mu_0(X) < \infty$) and there is a set $A \subseteq X$ such that $\mu^*(A) = \mu_0(X)$. Show that $\mu^*(E) = \mu^*(E \cap A)$ for every μ^*-measurable set E.

1.6.8. Let $\mathcal{A}$ be an algebra on a set X. Let $\mathcal{A}_\sigma$ be the collection of countable unions of sets from $\mathcal{A}$, and let $\mathcal{A}_{\sigma\delta}$ be the collection of countable intersections of sets from $\mathcal{A}_\sigma$. Given a premeasure μ_0 on $\mathcal{A}$, prove the following statements.

(a) If $E \subseteq X$ and $\varepsilon > 0$, then there exists a set $A \in \mathcal{A}_\sigma$ such that $E \subseteq A$ and $\mu^*(A) \le \mu^*(E) + \varepsilon$, and there exists a set $H \in \mathcal{A}_{\sigma\delta}$ such that $E \subseteq H$ and $\mu^*(H) = \mu^*(E)$.

(b) If $\mu^*(E) < \infty$, then E is μ^*-measurable if and only if there exists a set $H \in \mathcal{A}_{\sigma\delta}$ such that $E \subseteq H$ and $\mu^*(H \setminus E) = 0$.

(c) If μ_0 is σ-finite, then an arbitrary set $E \subseteq X$ is μ^*-measurable if and only if there exists a set $H \in \mathcal{A}_{\sigma\delta}$ such that $E \subseteq H$ and $\mu^*(H \setminus E) = 0$.

1.6.9. Let μ^* be an outer measure on X induced from a finite premeasure μ_0. If $E \subseteq X$, define the *inner measure* of E to be $\mu_*(E) = \mu_0(X) - \mu^*(E^{\mathrm{C}})$. Prove that E is μ^*-measurable if and only if $\mu^*(E) = \mu_*(E)$.

1.7 Lebesgue–Stieltjes Measures on the Real Line

We will show how the machinery of outer measures and premeasures can be used to construct Lebesgue measure on the real line. In fact, we will construct an entire class of measures, each of which is defined on at least the Borel σ-algebra (and possibly larger σ-algebras, as is the case for Lebesgue measure). These measures will be called *Lebesgue–Stieltjes measures*.

For simplicity, in this section we will abbreviate the Borel σ-algebra on $\mathbb{R}$ as

$$\mathcal{B} \;=\; \mathcal{B}_{\mathbb{R}}.$$

A measure that is defined on $\mathbb{R}$ with respect to this σ-algebra is called a *Borel measure* on $\mathbb{R}$. Examples include Lebesgue measure, the δ measure, and counting measure. However, the measures that we will construct will be σ-finite, so we will not obtain counting measure by this method. In fact, the measures that we will construct will have the property that $\mu(E)$ is finite for every bounded set E. Both Lebesgue measure and the δ measure have this property, and we will see that each of these measures are Lebesgue–Stieltjes measures.

1.7.1 The Distribution Function

First we motivate the construction. Suppose for simplicity that we have in hand a *bounded* Borel measure μ on $\mathbb{R}$. Then we define its *distribution function* to be

$$F(x) \;=\; \mu(-\infty,x], \qquad x \in \mathbb{R}.$$

Note that, depending on the measure, there can be a difference between $\mu(-\infty,x)$ and $\mu(-\infty,x]$.

The distribution function has the following properties.

Lemma 1.7.1. *If μ is a finite Borel measure on $\mathbb{R}$, then the following statements hold.*

(a) *F is monotone increasing.*

(b) *F right-continuous, i.e.,*

$$F(x) \;=\; \lim_{y\to x^+} F(y), \qquad \textit{all } x \in \mathbb{R}.$$

(c) $\displaystyle\lim_{x\to\infty} F(x) = \mu(\mathbb{R})$.

(d) $\displaystyle\lim_{x\to-\infty} F(x) = 0$.

Proof. (a) If $x \le y$, then monotonicity implies that

$$F(x) \;=\; \mu(-\infty,x] \;\le\; \mu(-\infty,y] \;=\; F(y).$$

Therefore F is increasing on $\mathbb{R}$.

(b) Fix any point $x \in \mathbb{R}$, and suppose that $x_1 \ge x_2 \ge \cdots$ is any monotonically decreasing sequence that converges to x from the right, i.e., $x_n \searrow x$. Applying continuity from above, it follows that

$$F(x) \;=\; \mu(-\infty, x] \;=\; \mu\Biggl(\bigcap_{n=1}^{\infty} \mu(-\infty, x_n]\Biggr)$$
$$= \lim_{n\to\infty} \mu(-\infty, x_n]$$
$$= \lim_{n\to\infty} F(x_n).$$

Since this is true for every sequence that decreases to x from the right, it follows that F is continuous from the right.

(c), (d) These properties follow in a similar manner, by applying continuity from above or below. $\square$

Example 1.7.2. The distribution function for the δ measure is

$$F(x) \;=\; \delta(-\infty, x] \;=\; \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases} \qquad \diamondsuit$$

An important property of the distribution function is that it determines the measure.

Lemma 1.7.3. *Let μ and ν be bounded Borel measures on $\mathbb{R}$. Let F be the distribution function for μ and let G be the distribution function for ν. Then*

$$\mu \;=\; \nu \qquad \Longleftrightarrow \qquad F \;=\; G.$$

Proof. $\Rightarrow$. This direction is trivial.

$\Leftarrow$. Suppose that $F = G$. Given any $a < b$, we have

$$(a, b] \;=\; (-\infty, b] \setminus (-\infty, a].$$

As we are dealing with finite measures, it follows that

$$\begin{aligned} \mu(a, b] \;&=\; \mu(-\infty, b] \;-\; \mu(-\infty, a] \\ &=\; F(b) - F(a) \\ &=\; G(b) - G(a) \\ &=\; \nu(-\infty, b] \;-\; \nu(-\infty, a] \\ &=\; \nu(a, b]. \end{aligned}$$

Applying continuity from below, it therefore follows that

$$\mu(a, b) \;=\; \mu\Biggl(\bigcup_{n=1}^{\infty} (a, b - \tfrac{1}{n}]\Biggr)$$
$$= \lim_{n\to\infty} \mu(a, b - \tfrac{1}{n}]$$

$$
\begin{aligned}
&= \lim_{n\to\infty} \nu(a, b - \tfrac{1}{n}] \\
&= \nu\biggl(\bigcup_{n=1}^{\infty} (a, b - \tfrac{1}{n}]\biggr) \\
&= \nu(a,b).
\end{aligned}
$$

Thus μ and ν coincide on a class of sets that determine the Borel σ-algebra. Continuing in this way, we can show that μ and ν agree on every Borel set, and therefore are the same measure (DETAILS needed here; similar to Exercise 1.7.13). □

Note that, as defined, the distribution function for Lebesgue measure would be identically ∞. We can deal with unbounded Borel measures, as long as those measures have the property that $\mu(a,b]$ is finite for every $a < b$. Every bounded measure automatically has this property, and Lebesgue measure is an example of an unbounded measure that also has this property. For these types of measures, we can define the distribution function by

$$
F(x) \;=\; \begin{cases} \mu(0,x], & x > 0, \\ 0, & x = 0, \\ \mu(x,0], & x < 0. \end{cases}
$$

If μ is a bounded measure, then the two distribution functions we have defined simply differ by a constant. The properties of this new distribution function are similar to the old one. In particular, we have

$$
\lim_{x\to\infty} F(x) \;=\; \mu(0,\infty).
$$

However, while this quantity is finite for a bounded measure, it will be infinite for an unbounded measure.

Example 1.7.4. The distribution function for Lebesgue measure is

$$
F(x) \;=\; x, \qquad x \in \mathbb{R}. \qquad \diamondsuit
$$

1.7.2 An Algebra for Lebesgue–Stieltjes Measures

We have seen that every bounded Borel measure (and some unbounded ones) determine a monotone increasing, right-continuous function on $\mathbb{R}$. Conversely, we will now show how each monotone increasing, right-continuous function determines a Borel measure. These are the *Lebesgue–Stieltjes measures* on $\mathbb{R}$.

Essentially, we want to create a measure μ_F by first defining a premeasure and then going through the process to extend to an outer measure, which is then restricted to the measurable sets to form a measure. We first need a class of elementary sets. These are essentially the half-open intervals $(a, b]$, except that these alone will not satisfy the requirements of elementary sets.

Notation 1.7.5. Our collection of elementary sets $\mathcal{E}$ will consist of all sets of the following forms:

$$\begin{aligned} &(a, b], && a < b \in \mathbb{R}, \\ &(a, \infty), && a \in \mathbb{R}, \\ &(-\infty, b], && b \in \mathbb{R}, \\ &\varnothing. && \diamondsuit \end{aligned}$$

This class contains the empty set, and its union is $\mathbb{R}$. Further, this collection is an *elementary family* in the following sense.

Lemma 1.7.6. *$\mathcal{E}$ is an* elementary family*. That is, it satisfies the following three requirements.*

(a) $\varnothing \in \mathcal{E}$.

(b) $E,\, F \in \mathcal{E} \implies E \cap F \in \mathcal{E}$.

(c) *If $E \in \mathcal{E}$, then E^{C} is a union of* finitely many disjoint elements *of $\mathcal{E}$.*

Proof. By inspection. □

All elementary families lead to algebras, as follows.

Lemma 1.7.7. *If $\mathcal{E}$ is an elementary family of subsets of a set X, then the collection*

$$\mathcal{A} = \left\{ \bigcup_{k=1}^{n} E_k : n \in \mathbb{N},\ \textit{disjoint } E_1, \dots, E_n \in \mathcal{E} \right\},$$

is an algebra *of subsets of X. That is, $\mathcal{A}$ is nonempty, $\mathcal{A}$ is closed under complements, and $\mathcal{A}$ is closed under* finite unions*.*

Proof. See Folland for this rather technical proof. □

Since $\mathcal{E}$ contains the empty set, so does $\mathcal{A}$.

1.7.3 Lebesgue–Stieltjes Measures

Suppose that F is a monotone increasing, right-continuous function on $\mathbb{R}$. We want to define a premeasure, which we will call ρ_F, use this premeasure to construct an outer measure μ_F^*, and then restrict μ_F^* to the μ_F^*-measurable sets to obtain a measure μ_F.

We define $\rho_F \colon \mathcal{A} \to [0, \infty]$ as follows. For the empty set, we declare that

$$\rho_F(\varnothing) \;=\; 0.$$

If we consider the relation between Borel measures and distribution functions discussed earlier, we see that for the elementary sets $(a, b]$ with $a < b$ finite, we should define

$$\rho_F(a, b] \;=\; F(b) - F(a).$$

To extend this to the algebra $\mathcal{A}$, suppose that

$$E \;=\; \bigcup_{k=1}^{n} (a_k, b_k]$$

is a union of disjoint elementary sets, where we allow any type of elementary set in this union. Then we define

$$\rho_F(E) \;=\; \sum_{k=1}^{n} \Bigl(F(b_k) - F(a_k)\Bigr).$$

Note that $F(b_k)$ could be infinite, but $F(a_k)$ is always finite. Therefore

$$0 \;\le\; F(b_k) - F(a_k) \;\le\; \infty,$$

and so the sum defining $\rho_F(E)$ exists in the extended real sense.

There is a well-definedness issue. For example, if $a < c < b$ then we have

$$(a, b] \;=\; (a, c] \cup (c, b].$$

However,

$$\Bigl(F(b) - F(c)\Bigr) \;+\; \Bigl(F(c) - F(a)\Bigr) \;=\; F(b) - F(a),$$

so this case can be dealt with. In fact, every case of ambiguity has this type of form, i.e., one half-open interval $(a, b]$ is partitioned into finitely many smaller disjoint half-open intervals. This produces telescoping sums that sum to the appropriate value.

Thus ρ_F is defined on the entire algebra $\mathcal{A}$, and it takes the "correct" value $\rho_F(a, b] = F(b) - F(a)$ on the half-open intervals $(a, b]$. Next, we claim that ρ_F is a premeasure on $\mathcal{A}$.

Theorem 1.7.8. *ρ_F is a premeasure on $\mathcal{A}$. That is,*

(a) $\rho_F(\emptyset) = 0$, *and*

(b) *if $E_1, E_2, \ldots \in \mathcal{A}$ are disjoint* and $\bigcup_{k=1}^{\infty} E_k \in \mathcal{A}$, *then*

$$\rho_F\Biggl(\bigcup_{k=1}^{\infty} E_k\Biggr) \;=\; \sum_{k=1}^{\infty} \rho_F(E_k). \qquad \diamondsuit$$

The proof of Theorem 1.7.8 is not trivial. Since we have already been through the entire construction of Lebesgue measure and since this proof is essentially of the same flavor, we are going to accept the theorem as given (for a complete proof, see Proposition 1.5 in [Fol99]).

Now that we know that we have a premeasure, Theorems 1.6.1 and 1.6.5 tell us that we can create an associated outer measure μ_F^* and then create a true measure by restricting μ_F^* to the μ_F^*-measurable sets. The following corollary gives the details and some additional properties of the resulting measure.

Corollary 1.7.9. *Let F be a monotone increasing, right-continuous function on $\mathbb{R}$, and define $\mu_F^*\colon \mathcal{P}(\mathbb{R}) \to [0,\infty]$ by*

$$\mu_F^*(E) \;=\; \inf\biggl\{\sum_k \rho_F(E_k) : E_k \in \mathcal{A},\, E \subseteq \bigcup_k E_k\biggr\}. \tag{1.9}$$

Then following statements hold.

(a) *μ_F^* is an outer measure on $\mathbb{R}$, and*

$$\mu_F^*(E) \;=\; \inf\biggl\{\sum_k \Bigl(F(b_k) - F(a_k)\Bigr) : E \subseteq \bigcup_k (a_k, b_k],\;\, a_k < b_k \in \mathbb{R}\biggr\}. \tag{1.10}$$

(b) *The σ-algebra Σ_F that consists of all μ_F^*-measurable sets contains every half-open interval $(a,b]$. Consequently*

$$\mathcal{B} \;\subseteq\; \Sigma_F,$$

i.e., Σ_F contains every Borel set.

(c) *$\mu_F = \mu_F^*|_{\Sigma_F}$ is a complete measure with respect to the σ-algebra Σ_F.*

(c) *μ_F is a Borel measure when restricted to the Borel σ-algebra $\mathcal{B}$.*

(d) *Given any $a < b \in \mathbb{R}$, we have*

$$\mu_F(a,b] \;=\; \mu_F^*(a,b] \;=\; F(b) - F(a).$$

(e) *If F, G are both monotone increasing and right-continuous, then*

$$\mu_F \;=\; \mu_G \qquad\Longleftrightarrow\qquad F - G = \textit{constant}.$$

Proof. (a) Theorem 1.6.1 tells us if $\mathcal{E}$ is *any* collection of sets, then the function μ_F^* defined by equation (1.9) is an outer measure. For this we do not even need to know that ρ_F is a premeasure.

To show that equations (1.9) and (1.10) are equivalent, set

$$\nu_F^*(E) \;=\; \inf\biggl\{\sum_k \Bigl(F(b_k) - F(a_k)\Bigr) : E \subseteq \bigcup_k (a_k, b_k],\ a_k < b_k \in \mathbb{R}\biggr\}.$$

Fix any set $E \subseteq \mathbb{R}$, and suppose that

$$E \;\subseteq\; \bigcup_k (a_k, b_k]$$

for some choice of $a_k < b_k \in \mathbb{R}$. Let $E_k = (a_k, b_k]$. Then $E \subseteq \cup E_k$, and we have

$$\sum_k \Bigl(F(b_k) - F(a_k)\Bigr) \;=\; \sum_k \rho_F(a_k, b_k] \;=\; \sum_k \rho_F(E_k).$$

Thus every covering of E in the sense given in the definition of $\nu_F^*(E)$ is a covering of E in the sense of the definition of $\mu_F^*(E)$ as given in equation (1.9). There may be more coverings in the μ_F^*-sense than in the ν_F^*-sense, so by taking the inf over all ν-type coverings we see that

$$\mu_F^*(E) \;\le\; \nu_F^*(E).$$

The converse direction is not quite as pleasant, but has the same flavor. The basic point is that every set $E_k \in \mathcal{A}$ can be written as a union of countably many finite half-open intervals $(a_k, b_k]$. (DETAILS needed here.)

(b) Because ρ_F is a premeasure, Theorem 1.6.5 tells us that every one of our elementary sets is μ_F^*-measurable. In particular, every half-open interval $(a, b]$ is measurable. Therefore the σ-algebra Σ_F contains every interval $(a, b]$. However, the Borel σ-algebra is generated by the collection of all such half-open intervals. That is, $\mathcal{B}$ is the *smallest* σ-algebra that contains every interval $(a, b]$. Consequently $\mathcal{B}$ must be a subset of Σ_F.

(c) Because μ_F^* is an outer measure, Carathéodory's Theorem tells us that the restriction of μ_F^* to Σ_F is a complete measure. That is, $\mu_F = \mu_F^*|_{\Sigma_F}$ is a complete measure.

(d) Since μ_F is a measure with respect to the σ-algebra Σ_F, it is also a measure with respect to any smaller σ-algebra. Since $\mathcal{B} \subseteq \Sigma_F$, we conclude that μ_F is still a measure when restricted to $\mathcal{B}$ (though it need not be complete any longer).

(e) Because ρ_F is a premeasure, Theorem 1.6.5 implies that $\mu_F^*(E) = \rho_F(E)$ for every set $E \in \mathcal{A}$. Further, since μ_F is simply the restriction of μ_F^* to Σ_F (which includes $\mathcal{A}$), we have $\mu_F(E) = \mu_F^*(E)$ for every $E \in \mathcal{A}$. In particular, taking $E = (a, b]$, we see that

$$\mu_F(a, b] \;=\; \mu_F^*(a, b] \;=\; \rho_F(a, b] \;=\; F(b) - F(a).$$

(f) $\Rightarrow$. Suppose that F and G are both monotone increasing and right-continuous and $\mu_F = \mu_G$. Then given any $0 < b$ we have

$$F(b) - F(0) \;=\; \mu_F(0,b] \;=\; \mu_G(0,b] \;=\; G(b) - G(0).$$

After a similar calculation using intervals of the form $(a,0]$ with $a < 0$, we see that F and G differ by a constant.

$\Leftarrow$. Suppose that F and G differ by a constant, say $c = F - G$. Considering equation (1.10), it follows that the outer measures μ_F^* and μ_G^* generated from F and G are identical. The σ-algebras they generate are likewise equal, and therefore the restrictions μ_F and μ_G are equal. $\square$

We give the following name to these types of measures.

Definition 1.7.10 (Lebesgue–Stieltjes Measure). Let F is a monotone increasing, right-continuous function on $\mathbb{R}$. Then the complete measure μ_F on Σ_F defined in Corollary 1.7.9 is called the *Lebesgue–Stieltjes measure determined by* F. $\diamondsuit$

Every Lebesgue–Stieltjes measure is a Borel measure (although it can be shown that $\mathcal{B}$ is *always* a proper subset of Σ_F).

Exercise 1.7.11. Suppose that $0 \le f \in L^1(\mathbb{R})$, i.e., f is nonnegative and integrable on $\mathbb{R}$ (with respect to Lebesgue measure). Define

$$F(x) \;=\; \int_{-\infty}^{x} f(t)\,dt.$$

Then F is monotone and continuous. In fact, we proved in Volume 1 that F is *absolutely continuous* on $\mathbb{R}$. Therefore F is differentiable a.e., and $F' = f$ a.e.

(a) Let μ_F be the Lebesgue–Stieltjes measure determined by F. Then Corollary 1.7.9 implies that

$$\mu_F(a,b] \;=\; F(b) - F(a) \;=\; \int_a^b f(t)\,dt \;=\; \int_{-\infty}^{\infty} f(t)\,\chi_{(a,b]}.$$

Use continuity from above or below to prove that

$$\mu_F(a,b) \;=\; \int_a^b f(t)\,dt \;=\; \int_{-\infty}^{\infty} f(t)\,\chi_{(a,b)}.$$

(b) Given any open set U, prove that

$$\mu_F(U) \;=\; \int_{-\infty}^{\infty} f(t)\,\chi_U(t)\,dt \;=\; \int_U f(t)\,dt.$$

Given any Borel set E, prove that

$$\mu_F(E) \;=\; \int_{-\infty}^{\infty} f(t)\,\chi_E(t)\,dt \;=\; \int_E f(t)\,dt.$$

(c) Prove that Σ_F includes every *Lebesgue measurable* set, and extend part (b) to Lebesgue measurable sets E.

(d) Prove that
$$|E| = 0 \qquad \Longrightarrow \qquad \mu_F(E) = 0.$$
Using terminology that we will introduce later, this says that

μ_F is absolutely continuous with respect to Lebesgue measure.

We denote this by writing $\mu_F \ll dx$. Observe that we also have that

F is an absolutely continuous function. $\diamondsuit$

1.7.4 Regularity of Lebesgue–Stieltjes Measures

Suppose that F is a monotone increasing, right-continuous function on $\mathbb{R}$. Since μ_F, μ_F^*, and ρ_F all agree on the half-open intervals $(a,b]$, we can rewrite the definition of μ_F^* given in equation (1.10) as

$$\mu_F^*(E) \;=\; \inf\biggl\{\sum_k \mu_F(a_k,b_k] : E \subseteq \bigcup_k (a_k,b_k],\ a_k < b_k \in \mathbb{R}\biggr\}. \tag{1.11}$$

Now, if μ_F is Lebesgue measure (obtained by taking $F(x) = x$), then $\mu_F(a,b] = \mu_F(a,b)$ for every $a < b$. However, this equality does not hold for every Lebesgue–Stieltjes measure (for example, it fails if $\mu_F = \delta$). Even so, the following theorem shows that we can replace half-open intervals in equation (1.11) by open intervals.

Theorem 1.7.12. *Let F be a monotone increasing, right-continuous function on $\mathbb{R}$, and let μ_F^* be the outer measure determined by F defined in equation* (1.11)*. If for each set $E \subseteq \mathbb{R}$ we define*

$$\nu_F^*(E) \;=\; \inf\biggl\{\sum_k \mu_F(a_k,b_k) : E \subseteq \bigcup_k (a_k,b_k),\ a_k < b_k \in \mathbb{R}\biggr\}, \tag{1.12}$$

then we have $\mu_F^* = \nu_F^*$.

Proof. Suppose that $E \subseteq \cup(a_k,b_k)$. Each interval (a_k,b_k) is a Borel set and therefore is μ_F^*-measurable. Applying monotonicity and countable subadditivity, we compute that

$$\mu_F(E) \;\le\; \mu_F\biggl(\bigcup_k (a_k,b_k)\biggr) \;\le\; \sum_k \mu_F(a_k,b_k).$$

Taking the inf over all such coverings, we obtain

$$\mu_F^*(E) \ \le\ \nu_F^*(E).$$

To prove the converse inequality, fix any $\varepsilon > 0$. Then, by definition of $\mu_F^*(E)$, there exists some covering

$$E \ \subseteq\ \bigcup_k (a_k, b_k]$$

such that

$$\mu_F^*(E) \ \le\ \sum_k \mu_F(a_k, b_k] \ \le\ \mu_F^*(E) + \varepsilon.$$

Since F is right-continuous, for each k there must exist a $\delta_k > 0$ such that

$$F(b_k + \delta_k) \ -\ F(b_k) \ <\ 2^{-k}\varepsilon.$$

Note that

$$E \ \subseteq\ \bigcup_k (a_k, b_k] \ \subseteq\ \bigcup_k (a_k, b_k + \delta_k).$$

Therefore

$$\begin{aligned}
\nu_F^*(E) \ &\le\ \sum_k \mu_F(a_k, b_k + \delta_k) && \text{(by definition of } \nu_F^*\text{)}\\
&\le\ \sum_k \mu_F(a_k, b_k + \delta_k] && \text{(monotonicity)}\\
&\le\ \sum_k \Big(F(b_k + \delta_k) \ -\ F(a_k)\Big)\\
&\le\ \sum_k \Big(F(b_k) \ +\ 2^{-k}\varepsilon \ -\ F(a_k)\Big)\\
&\le\ \sum_k \mu_F(a_k, b_k] \ +\ \varepsilon\\
&\le\ \mu_F^*(E) \ +\ 2\varepsilon.
\end{aligned}$$

Since ε is arbitrary, we conclude that

$$\nu_F^*(E) \ \le\ \mu_F^*(E). \qquad \square$$

As a consequence, we obtain some additional results that show that Lebesgue–Stieltjes measures have regularity properties that are similar to those of Lebesgue measure (and the proofs are similar too, so we assign these as exercises).

Exercise 1.7.13. Let F be a monotone increasing, right-continuous function on $\mathbb{R}$, and let μ_F^* be the corresponding outer measure determined by F. Given $E \subseteq \mathbb{R}$, prove the following statements.

(a) $\mu_F^*(E) = \inf\big\{\mu_F(U) : \text{open } U \supseteq E\big\}$.

(a) There exists a G_δ-set $V \supseteq E$ such that $\mu_F^*(E) = \mu_F(V)$. $\diamondsuit$

The preceding exercise applies to every set $E \subseteq \mathbb{R}$. If E happens to belong to Σ_F (i.e., it is μ_F^*-measurable), then $\mu_F^*(E) = \mu_F(E)$.

The next exercise gives a regularity property of Lebesgue–Stieltjes measures for measurable sets.

Exercise 1.7.14. Let F be a monotone increasing, right-continuous function on $\mathbb{R}$, and let μ_F^* be the corresponding outer measure determined by F. Given $E \subseteq \mathbb{R}$, prove that the following five statements are equivalent.

(a) $E \in \Sigma_F$, i.e., E is μ_F^*-measurable.

(b) For every $\varepsilon > 0$, there exists an open set $U \supseteq E$ such that $\mu_F^*(U \backslash E) < \varepsilon$.

(c) For every $\varepsilon > 0$, there exists a closed set $F \subseteq E$ such that $\mu_F^*(E \backslash F) < \varepsilon$.

(d) There exists a G_δ-set $H \supseteq E$ such that $\mu_F^*(H \backslash E) = 0$.

(d) There exists an F_σ-set $H \subseteq E$ such that $\mu_F^*(E \backslash H) = 0$. $\diamondsuit$

Problems

1.7.15. Let F be a monotone increasing, right-continuous function on $\mathbb{R}$, and let μ_F^* be the corresponding outer measure determined by F. Given $E \in \Sigma_F$, prove that

$$\mu_F(E) \;=\; \sup\bigl\{\mu_F(K) : \text{compact } K \subseteq E\bigr\}.$$

Chapter 2
Measurable Functions

In this chapter we will define integration with respect to an abstract measure. In many respects, integration with respect to an abstract measure is similar to integration with respect to Lebesgue measure, which we studied extensively in Volume 1. Therefore many of the theorems we will encounter in this chapter will have a familiar flavor. Exercises or other results that are marked with a red asterisk (like this: Exercise n.n.*) are very similar to results we have seen for Lebesgue measure.

2.1 Measurable Functions

Just as a measure need not be defined on every subset of X, we usually will not be able to integrate every possible function on X. We will need to restrict our attention to "measurable functions," which we define and study in this section.

2.1.1 Abstract Measurability

To motivate the definition of a measurable function that maps one measurable space into another measurable space, recall the definition of a continuous function. If f maps one metric space X into another metric space Y, then we can use the notion of distance to formulate continuity in terms of preservation of limits. However, there is a more fundamental definition of continuity that does not depend on distance notions but rather on topology. Explicitly, given topological spaces X and Y, a function $f\colon X \to Y$ is continuous if and only if *the inverse image of each open subset of* Y *is open in* X. That is, the following implication must hold:

$$U \text{ is an open subset of } Y \implies f^{-1}(U) \text{ is open subset of } X. \tag{2.1}$$

Just as the open sets are the "building blocks" of a topological space, the measurable sets are the building blocks of a measurable space. Substituting measurable sets for open sets in equation (2.1) leads to the following definition.

Definition 2.1.1 (Measurable Function). Let (X, Σ_X) and (Y, Σ_Y) be measurable spaces. A function $f\colon X \to Y$ is (Σ_X, Σ_Y)*-measurable*, or simply *measurable* for short, if the following implication holds:

$$E \in \Sigma_Y \implies f^{-1}(E) \in \Sigma_X. \tag{2.2}$$

That is, the inverse image of every measurable subset of Y must be a measurable subset of X. $\diamondsuit$

Notation 2.1.2. Sometimes we will write $f\colon (X, \Sigma_X) \to (Y, \Sigma_Y)$ to denote that f is a function from X to Y and Σ_X, Σ_Y are the σ-algebras on X and Y that we have in mind. This notation does not mean that f is measurable, but rather only means that Σ_X and Σ_Y are the σ-algebras that we should use in order to test whether f is a measurable function. $\diamondsuit$

Definition 2.1.1 can be difficult to implement directly. The next exercise gives a useful simplification of the test for measurability: Instead of showing that equation (2.2) holds for *every* set $E \in \Sigma_Y$, it suffices to only consider sets E that belong to a generating family $\mathcal{E}$ for the σ-algebra Σ_Y.

Exercise 2.1.3. Let (X, Σ_X) and (Y, Σ_Y) be measurable spaces, and suppose that Σ_Y is generated by a collection $\mathcal{E} \subseteq Y$, i.e., $\Sigma_Y = \Sigma(\mathcal{E})$. Show that the following two statements are equivalent.

(a) $f\colon X \to Y$ is (Σ_X, Σ_Y)-measurable.

(b) For each set $E \in \mathcal{E}$ we have $f^{-1}(E) \in \Sigma_X$. $\diamondsuit$

As a corollary, we obtain an important result for measurability of functions that are defined on topological spaces.

Corollary 2.1.4. *Let X, Y be topological spaces, and let $\mathcal{B}_X$, $\mathcal{B}_Y$ be the Borel σ-algebras on X and Y, respectively. Then*

$$f\colon X \to Y \text{ is continuous} \implies f \text{ is } (\mathcal{B}_X, \mathcal{B}_Y)\text{-measurable.}$$

Proof. By definition, the open sets generate the Borel σ-algebra $\mathcal{B}_Y$. Also, since f is continuous we have by definition that $f^{-1}(U) \in \mathcal{B}_X$ for every open set $U \subseteq Y$. Applying Exercise 2.1.3, we conclude that f is $(\mathcal{B}_X, \mathcal{B}_Y)$-measurable. $\square$

It is important to note that Corollary 2.1.4 only applies when X and Y are topological spaces and we take the σ-algebras on X and Y to be the Borel

σ-algebras. If Σ_X is a σ-algebra on X such that $\mathcal{B}_X \subseteq \Sigma_X$ and $f\colon X \to Y$ is continuous, then for each open $U \subseteq Y$ we have $f^{-1}(U) \in \mathcal{B}_X \subseteq \Sigma_X$, and so f is $(\Sigma_X, \mathcal{B}_Y)$-measurable.

2.1.2 Real-Valued, Extended Real-Valued, and Complex-Valued Functions

Many of the functions that we encounter in practice, and most of the functions that we hope to integrate, are either real-valued, extended real-valued, or complex-valued. In other words, the spaces Y that we deal with most often are the real line $\mathbb{R}$, the extended real line $\overline{\mathbb{R}} = [-\infty, \infty]$, and the complex plane $\mathbb{C}$.

Let (X, Σ) be a measurable space, and let f be a function on X that takes values in $\mathbb{R}$, $\overline{\mathbb{R}}$, or $\mathbb{C}$. In order to define the meaning of measurability of f, we have to decide which σ-algebra that we will place on the codomains $\mathbb{R}$, $\overline{\mathbb{R}}$, or $\mathbb{C}$. In almost all circumstances, the appropriate algebra turns out to be the Borel σ-algebra, so we take this as our default choice of σ-algebra on the codomains $\mathbb{R}$, $\overline{\mathbb{R}}$, or $\mathbb{C}$. Here is the restatement of Definition 2.1.1 for these particular codomains.

Definition 2.1.5. Let (X, Σ) be a measurable space.

(a) A function $f\colon X \to \mathbb{R}$ is *measurable* if it is $(\Sigma, \mathcal{B}_{\mathbb{R}})$-measurable.

(b) A function $f\colon X \to \overline{\mathbb{R}}$ is *measurable* if it is $(\Sigma, \mathcal{B}_{\overline{\mathbb{R}}})$-measurable.

(b) A function $f\colon X \to \mathbb{C}$ is *measurable* if it is $(\Sigma, \mathcal{B}_{\mathbb{C}})$-measurable. $\diamondsuit$

That is, a function f is measurable if the inverse image of every Borel set in $\mathbb{R}$, $\overline{\mathbb{R}}$, or $\mathbb{C}$ (as appropriate) is a measurable subset of X. Of course, this definition is a bit premature since we have not yet defined the Borel σ-algebras on $\overline{\mathbb{R}}$ and $\mathbb{C}$. We will do this, but for motivation we begin with the Borel algebra that we already know, which is the Borel σ-algebra on $\mathbb{R}$. We will give a much simpler equivalent formulation of measurability for real-valued functions, and once we see this it will be easy to move to the extended real-valued and complex-valued settings.

2.1.3 Real-Valued Functions

Let $f\colon X \to \mathbb{R}$ be a real-valued function on a set X. By Definition 2.1.5(a), measurability of f means $(\Sigma, \mathcal{B}_{\mathbb{R}})$-measurability. Stated explicitly, this means that $f\colon X \to \mathbb{R}$ is measurable if and only if

$$f^{-1}(E) \in \Sigma \text{ for every Borel set } E \subseteq \mathbb{R}. \tag{2.3}$$

Fortunately, Exercise 2.1.3 tells us that we do not need to test *every* Borel set in order to determine measurability—it is enough to just test sets contained in a generating family for the σ-algebra. Since the Borel σ-algebra on $\mathbb{R}$ is generated by the open sets (by definition), instead of taking every Borel set in equation (2.3) it is enough to just consider open sets E. Or we can use sets from any other generating family, such as the families given in Problem 1.1.14. In particular, each of the following is a generating family for $\mathcal{B}_{\mathbb{R}}$:

$$\begin{aligned} \mathcal{E}_1 &= \{(a,\infty) : a \in \mathbb{R}\}, \\ \mathcal{E}_2 &= \{[a,\infty) : a \in \mathbb{R}\}, \\ \mathcal{E}_3 &= \{(-\infty,a) : a \in \mathbb{R}\}, \\ \mathcal{E}_4 &= \{(-\infty,a] : a \in \mathbb{R}\}. \end{aligned}$$

We can use sets from any of these families in equation (2.3) instead of using every Borel set. For example, choosing the family $\mathcal{E}_1$, we see that it is sufficient to just use sets of the form $E = (a,\infty)$ in equation (2.3). This gives us the following lemma. The statement of this lemma uses our standard abbreviations for inverse images, such as

$$\{f > a\} = f^{-1}(a,\infty).$$

Lemma 2.1.6.* *Let (X,Σ) be a measurable space. Given $f\colon X \to \mathbb{R}$, the following statements are equivalent.*

(a) f *is measurable.*

(b) $\{f > a\} \in \Sigma$ *for every* $a \in \mathbb{R}$.

(c) $\{f \ge a\} \in \Sigma$ *for every* $a \in \mathbb{R}$.

(d) $\{f < a\} \in \Sigma$ *for every* $a \in \mathbb{R}$.

(e) $\{f \le a\} \in \Sigma$ *for every* $a \in \mathbb{R}$. $\diamondsuit$

When $X = \mathbb{R}^d$, we often use the following terminology.

Definition 2.1.7. Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be given.

(a) We say that f is *Borel measurable* if f is $(\mathcal{B}_{\mathbb{R}^d}, \mathcal{B}_{\mathbb{R}})$-measurable. By Lemma 2.1.6, f is Borel measurable if and only if

$$\forall\, a \in \mathbb{R}, \quad \{f > a\} \text{ is a Borel set.}$$

(b) We say that f is *Lebesgue measurable* if f is $(\mathcal{L}_{\mathbb{R}^d}, \mathcal{B}_{\mathbb{R}})$-measurable. By Lemma 2.1.6, f is Lebesgue measurable if and only if

$$\forall\, a \in \mathbb{R}, \quad \{f > a\} \text{ is Lebesgue measurable.} \qquad \diamondsuit$$

Since every Borel set is Lebesgue measurable,

$$f \text{ is Borel measurable} \quad \Longrightarrow \quad f \text{ is Lebesgue measurable.}$$

However, the converse implication fails in general. For example, if $E \subseteq \mathbb{R}^d$, then χ_E is a Lebesgue measurable function on $\mathbb{R}^d$ if and only if E is a Lebesgue measurable set, and χ_E is Borel measurable if and only if E is a Borel set.

Remark 2.1.8. Our default choice of σ-algebra for $\mathbb{R}^d$ is the Lebesgue σ-algebra, so if we write "$f\colon \mathbb{R}^d \to \mathbb{R}$ is measurable" without qualification, then we mean that f is Lebesgue measurable. $\diamondsuit$

Here are some examples of measurable functions.

Lemma 2.1.9. *If $f\colon \mathbb{R} \to \mathbb{R}$ is monotone increasing, then f is Borel measurable (and hence is also Lebesgue measurable).*

Proof. Fix $a \in \mathbb{R}$. Since f is monotone increasing, $\{f > a\}$ is either an interval of the form (x, ∞), or it is an interval of the form $[x, \infty)$. Each of these are Borel sets, so f is Borel measurable. $\square$

Lemma 2.1.10.* *If $f\colon \mathbb{R}^d \to \mathbb{R}$ is continuous, then f is Borel measurable (and hence is also Lebesgue measurable).*

Proof. Fix $a \in \mathbb{R}$. Since the interval (a, ∞) is an open subset of $\mathbb{R}$ and f is continuous, the inverse image $f^{-1}(a, \infty)$ is open in $\mathbb{R}^d$ and hence is a Borel set. Since $\{f > a\} = f^{-1}(a, \infty)$, this shows that f is Borel measurable. $\square$

2.1.4 Complex-Valued Functions

Now we consider complex-valued functions on X (which includes real-valued functions as a special case). By definition, the Borel σ-algebra on the complex plane is the σ-algebra that is generated by the open subsets of $\mathbb{C}$. The topology on $\mathbb{C}$ is simply the usual Euclidean topology of a plane, i.e., we identify $\mathbb{C}$ with $\mathbb{R}^2$ and identify the open subsets of $\mathbb{C}$ with the open subsets of $\mathbb{R}^2$. For example, the "open rectangle" $\{x + iy : a < x < b, c < y < d\}$ is an open subset of $\mathbb{C}$. Since the open sets generate the Borel σ-algebra $\mathcal{B}_{\mathbb{C}}$, it follows that a function $f\colon X \to \mathbb{C}$ is measurable if and only if

$$f^{-1}(U) \in \Sigma \text{ for every open set } U \subseteq \mathbb{C}.$$

However, if we like we can use other generating families for $\mathcal{B}_{\mathbb{C}}$ to test for measurability. For example, Problem 1.1.15 gives several families that generate the Borel σ-algebra on $\mathbb{R}^2$, and by making the appropriate identifications this also gives us generating families for $\mathcal{B}_{\mathbb{C}}$. In particular, it is often convenient to work with the generating family $\mathcal{E}$ that consists of all of the horizontal or vertical open strips

$$\{x+iy : a < x < b,\, y \in \mathbb{R}\} \qquad \text{and} \qquad \{x+iy : x \in \mathbb{R},\, c < y < d\}.$$

A function $f\colon X \to \mathbb{C}$ is measurable if and only if the inverse image of any one of these strips is a measurable subset of X. This is precisely the family to use to solve the next exercise.

Exercise 2.1.11. Let (X, Σ) be a measurable space. Given $f\colon X \to \mathbb{C}$, let f_r and f_i be the real and imaginary parts of f, i.e., $f = f_r + i f_i$ where f_r, f_i are real-valued. Show that

$$f \text{ is measurable} \quad \Longleftrightarrow \quad f_r \text{ and } f_i \text{ are both measurable.} \qquad \diamondsuit$$

If $X = \mathbb{R}^d$, then we define "Borel measurability" and "Lebesgue measurability" of a function $f\colon \mathbb{R}^d \to \mathbb{C}$ analogously to how we defined it for real-valued functions in Definition 2.1.7. By Exercise 2.1.11, a complex-valued function f is Borel measurable if and only if its real and imaginary parts are real-valued Borel measurable functions, and similarly for Lebesgue measurable functions.

Example 2.1.12. Suppose that $f\colon \mathbb{R}^d \to \mathbb{C}$ is continuous. Then the inverse image of any open subset of $\mathbb{C}$ is open in $\mathbb{R}^d$ and hence is a Borel set. Since the open sets generate the Borel σ-algebra, this tells us that f is Borel measurable, and hence is also Lebesgue measurable. Alternatively, we can simply observe that the real and imaginary parts of a continuous complex-valued function are continuous, and every continuous real-valued function is measurable by Lemma 2.1.10. $\diamondsuit$

By identifying $\mathbb{R}^2$ with $\mathbb{C}$, we can also define Borel measurability and Lebesgue measurability for functions of the form $f\colon \mathbb{C} \to \mathbb{C}$. In particular, Example 2.1.12 tells us that every continuous function $f\colon \mathbb{C} \to \mathbb{C}$ is Borel measurable. The next exercise constructs a square-root function on $\mathbb{C}$ that is not continuous, but is still Borel measurable. One way to solve this exercise is to show directly that the inverse image of every open subset of $\mathbb{C}$ is a Borel set in $\mathbb{C}$ (or, it suffices to just consider sets in a generating family instead of every open set).

Exercise 2.1.13. Define a square root function $Sz = z^{1/2}$ on $\mathbb{C}$ by declaring that

$$(re^{i\theta})^{1/2} \;=\; r^{1/2} e^{i\theta/2}, \qquad r > 0,\; 0 \le \theta < 2\pi.$$

Show that $S\colon \mathbb{C} \to \mathbb{C}$ is Borel measurable (but not continuous). $\diamondsuit$

2.1.5 Extended Real-Valued Functions

In order to characterize measurability of extended real-valued functions, we need to understand the Borel σ-algebra on $\overline{\mathbb{R}}$, or we at least need to know a generating set for $\mathcal{B}_{\overline{\mathbb{R}}}$. We will give a detailed discussion of $\mathcal{B}_{\overline{\mathbb{R}}}$ in a moment, but for most purposes it is sufficient just to know that $\mathcal{B}_{\overline{\mathbb{R}}}$ is generated by the family of intervals of the form $(a, \infty]$. This seems quite reasonable in light of the fact that the family of intervals (a, ∞) generates the Borel σ-algebra on $\mathbb{R}$. So for now let us simply take this fact as an axiom, and see what it implies about measurability of extended real-valued functions.

Axiom 2.1.14. Each of the following families of subsets of $\overline{\mathbb{R}}$ generates the Borel σ-algebra on $\overline{\mathbb{R}}$:

(a) $\mathcal{E}_1 = \{(a, \infty] : a \in \mathbb{R}\}$,

(b) $\mathcal{E}_2 = \{[a, \infty] : a \in \mathbb{R}\}$.

(c) $\mathcal{E}_3 = \{[-\infty, a) : a \in \mathbb{R}\}$.

(d) $\mathcal{E}_4 = \{[-\infty, a] : a \in \mathbb{R}\}$. $\diamondsuit$

For an extended real-valued function,

$$\{f > a\} \;=\; \{x \in X : f(x) > a\} \;=\; f^{-1}(a, \infty],$$
$$\{f \geq a\} \;=\; \{x \in X : f(x) \geq a\} \;=\; f^{-1}[a, \infty],$$

and so forth. Therefore, Axiom 2.1.14 immediately implies the following characterization of measurability of extended real-valued functions.

Lemma 2.1.15. *Let (X, Σ) be a measurable space. Given $f \colon X \to \overline{\mathbb{R}}$, the following statements are equivalent.*

(a) *f is measurable.*

(b) *$\{f > a\} \in \Sigma$ for every $a \in \mathbb{R}$.*

(c) *$\{f \geq a\} \in \Sigma$ for every $a \in \mathbb{R}$.*

(d) *$\{f < a\} \in \Sigma$ for every $a \in \mathbb{R}$.*

(e) *$\{f \leq a\} \in \Sigma$ for every $a \in \mathbb{R}$.* $\diamondsuit$

Note the similarity between Lemma 2.1.6 and Lemma 2.1.15—the only difference is that the functions considered in Lemma 2.1.15 can take the values $\pm\infty$ in addition to real values.

If $X = \mathbb{R}^d$, then we define "Borel measurability" and "Lebesgue measurability" of a function $f \colon \mathbb{R}^d \to \overline{\mathbb{R}}$ analogously to how we defined it for real-valued functions in Definition 2.1.7.

We introduce a notational convention that applies when we have an extended real-valued function f on a measurable space (X, Σ) and additionally have a measure μ that is defined on this space.

Notation 2.1.16. Let (X, Σ, μ) be a measure space. We often encounter extended real-valued functions that are finite except on a set with μ-measure zero. If $f\colon X \to \overline{\mathbb{R}}$ is such that $\{f = \pm\infty\}$ is μ-measurable and

$$\mu\{f = \pm\infty\} \;=\; 0,$$

then we say that f *is finite μ-a.e.* $\diamondsuit$

As usual, we sometimes use the abbreviated phrase "f is finite a.e." if the measure μ is understood.

2.1.6 The Topology of the Extended Real Line

In practice, Lemma 2.1.15 contains all that we usually need to know about measurability of extended real-valued functions. However, for the interested reader we will give a more detailed discussion of the Borel σ-algebra $\mathcal{B}_{\overline{\mathbb{R}}}$, and explain why Axiom 2.1.14 holds. This discussion can be skipped without much loss, but the exercises in the discussion do give a good insight into the similarities and differences between topologies and σ-algebras.

By definition, the Borel σ-algebra $\mathcal{B}_{\overline{\mathbb{R}}}$ is the smallest σ-algebra on $\overline{\mathbb{R}}$ that contains all of the open subsets of $\overline{\mathbb{R}}$. So the real issue in defining $\mathcal{B}_{\overline{\mathbb{R}}}$ is this question: What are the open subsets of $\overline{\mathbb{R}}$? Or, in more technical jargon: What is the topology of $\overline{\mathbb{R}}$? Of course, there are many possible topologies on $\overline{\mathbb{R}}$, but since we are thinking of $\overline{\mathbb{R}}$ as being an extension of $\mathbb{R}$, we seek a topology on $\overline{\mathbb{R}}$ that extends the topology of $\mathbb{R}$ in a natural way. Every subset of $\mathbb{R}$ that is open with respect to the usual topology on $\mathbb{R}$ should still be an open subset of $\overline{\mathbb{R}}$. Also, sets like $(a, \infty]$ should be open, but we do not want a singleton like $\{\infty\}$ to be open.

Now, the topology $\mathcal{T}_{\mathbb{R}}$ of $\mathbb{R}$ is generated by the open intervals (a, b). In fact, every open subset of $\mathbb{R}$ can be written as a union of open intervals (a, b), so the collection $\beta_{\mathbb{R}}$ of all open intervals (a, b) is called a *base* for the topology on $\mathbb{R}$. The next definition extends this idea to $\overline{\mathbb{R}}$ by combining the infinite intervals $(a, \infty]$ and $[-\infty, b)$ with the finite open intervals (a, b).

Definition 2.1.17 (Topology of $\overline{\mathbb{R}}$). A subset of $\overline{\mathbb{R}}$ is open if it can be written as a union of intervals or the form (a, b), $(a, \infty]$, or $[-\infty, b)$, where a, b belong to $\mathbb{R}$. We set

$$\beta_{\overline{\mathbb{R}}} \;=\; \bigl\{(a, b) : a < b\bigr\} \;\cup\; \bigl\{(a, \infty] : a \in \mathbb{R}\bigr\} \;\cup\; \bigl\{[-\infty, b) : b \in \mathbb{R}\bigr\}.$$

The topology of $\overline{\mathbb{R}}$ is the collection $\mathcal{T}_{\overline{\mathbb{R}}}$ containing all of open subsets of $\overline{\mathbb{R}}$, i.e., $\mathcal{T}_{\overline{\mathbb{R}}}$ is the set of all possible unions of elements of $\beta_{\overline{\mathbb{R}}}$. $\diamondsuit$

Remark 2.1.18. An alternative way to motivate the topology for $\overline{\mathbb{R}}$ is think of $\overline{\mathbb{R}} = [-\infty, \infty]$ in the same way that we think of a closed interval $[a, b]$. The topologies of $\overline{\mathbb{R}}$ and $[a, b]$ should be identical in some sense (in topological terms, they should be *homeomorphic* topological spaces). If we approach the topology of $\overline{\mathbb{R}}$ this way, we obtain exactly the same open sets as those obtained by applying Definition 2.1.17.

One way to explicitly define the open sets using this viewpoint is to make $\overline{\mathbb{R}}$ into a metric space by defining the distance between points $x, y \in \overline{\mathbb{R}}$ to be

$$\mathrm{d}(x, y) \;=\; |\arctan(x) - \arctan(y)|,$$

where we take $\arctan(\infty) = \pi/2$ and $\arctan(-\infty) = -\pi/2$. $\diamondsuit$

In the language of topology, because every open subset of $\overline{\mathbb{R}}$ is a union of elements of $\beta_{\overline{\mathbb{R}}}$, the collection $\beta_{\overline{\mathbb{R}}}$ is called a *base* for the topology $\mathcal{T}_{\overline{\mathbb{R}}}$. Of course, we are jumping a little ahead of ourselves here, because we have not yet shown that $\mathcal{T}_{\overline{\mathbb{R}}}$ satisfies the requirements of a topology. This is done in the next exercise, which also shows that the open subsets of $\overline{\mathbb{R}}$ can be written as countable unions of elements of $\beta_{\overline{\mathbb{R}}}$.

Exercise 2.1.19. (a) Show that the family $\beta_{\overline{\mathbb{R}}}$ is closed under finite intersections.

(b) Show that $\mathcal{T}_{\overline{\mathbb{R}}}$ is a topology on $\overline{\mathbb{R}}$, i.e., it is nonempty, closed under finite intersections, and closed under arbitrary unions.

(c) By definition, a set $U \subseteq \overline{\mathbb{R}}$ is open if and only if U can be written as a union of intervals of the form (a, b), $(a, \infty]$, or $[-\infty, b)$. Show that U is open if and only if it can be written as a *countable* union of such intervals. $\diamondsuit$

Remark 2.1.20. (a) By construction, every open subset of $\mathbb{R}$ is an open subset of $\overline{\mathbb{R}}$. In particular, the real line $\mathbb{R}$ is an open subset of $\overline{\mathbb{R}}$.

(b) If U is an open subset of $\overline{\mathbb{R}}$ then $U \cap \mathbb{R}$ is a union of intervals of the form (a, b), (a, ∞), or $(-\infty, b)$, and therefore is an open subset of $\mathbb{R}$. However, the converse need not hold, i.e., $U \cap \mathbb{R}$ may be an open subset of $\mathbb{R}$ even though U is not an open subset of $\overline{\mathbb{R}}$. For example, $U = (0, 1) \cup \{\infty\}$ is not an open subset of $\overline{\mathbb{R}}$, but its intersection with $\mathbb{R}$ is the open interval (a, b). $\diamondsuit$

So, we have defined the topology of $\overline{\mathbb{R}}$, and we have determined a generating family (in fact, a base) for this topology. By definition, the Borel σ-algebra $\mathcal{B}_{\overline{\mathbb{R}}}$ is the smallest σ-algebra on $\overline{\mathbb{R}}$ that contains all of the open subsets of $\overline{\mathbb{R}}$. The next exercise gives some generating families for $\mathcal{B}_{\overline{\mathbb{R}}}$, and thereby justifies Axiom 2.1.14.

Exercise 2.1.21. Show that $\mathcal{B}_{\overline{\mathbb{R}}}$ is the smallest σ-algebra on $\overline{\mathbb{R}}$ that contains all intervals of the form $(a,\infty]$. That is, show that if we set

$$\mathcal{E} \;=\; \bigl\{(a,\infty] : a \in \mathbb{R}\bigr\},$$

then $\mathcal{B}_{\overline{\mathbb{R}}} = \Sigma(\mathcal{E})$. Also show that each of the other families given in Axiom 2.1.14 generate $\mathcal{B}_{\overline{\mathbb{R}}}$. $\diamondsuit$

Finally, the next exercise shows that the Borel subsets of $\overline{\mathbb{R}}$ are simply the Borel subsets of $\mathbb{R}$ with $\pm\infty$ possibly adjoined. In particular, every set that is a Borel set in $\mathbb{R}$ is a Borel set in $\overline{\mathbb{R}}$.

Exercise 2.1.22. Given $E \subseteq \overline{\mathbb{R}}$, show that the following statements are equivalent.

(a) E is a Borel set in $\overline{\mathbb{R}}$.

(b) $E \cap \mathbb{R}$ is a Borel set in $\mathbb{R}$.

(c) E has one of the following forms:

$$A, \quad A \cup \{\infty\}, \quad A \cup \{-\infty\}, \quad A \cup \{-\infty,\infty\},$$

where A is a Borel set in $\mathbb{R}$. $\diamondsuit$

2.1.7 Functions Equal Almost Everywhere

If (X,Σ,μ) is a complete measure space, then every subset of a null set is measurable. This gives us the following important lemma, which states that if we change the values of a measurable function f on a set with μ-measure zero, then the new function is still measurable.

Lemma 2.1.23.* *Let (X,Σ,μ) be a complete measure space, and let f be a measurable function on X (either extended real-valued or complex-valued). If g is a function on X such that $f = g$ μ-a.e., then g is measurable.*

Proof. Assume f and g are extended real-valued functions, and $f = g$ μ-a.e. This means that the set $Z = \{f \ne g\}$ is measurable and $\mu(Z) = 0$.

Fix $a \in \mathbb{R}$. By hypothesis, the set $A = \{f < a\}$ is measurable, and our goal is to show that $B = \{g > a\}$ is also measurable. However, $A\backslash B \subseteq Z$ and μ is complete, so we know that $A\backslash B$ is measurable. Similarly, $B\backslash A$ is measurable, so

$$B \;=\; A \,\cup\, (B\backslash A) \,\backslash\, (A\backslash B)$$

is measurable, and therefore g is a measurable function.

The complex case follows by splitting into real and imaginary parts. $\square$

Here is a useful corollary for functions on $\mathbb{R}^d$. Although we do not explicitly specify a measure on $\mathbb{R}^d$ in this corollary, following Notation 1.2.8 we implicitly assume that it is our default choice, i.e., Lebesgue measure.

Corollary 2.1.24.* *Let $f\colon \mathbb{R}^d \to \mathbb{C}$ be given. If there is a continuous function $g\colon \mathbb{R}^d \to \mathbb{C}$ such that $f = g$ almost everywhere, then f is measurable.* $\diamondsuit$

In summary, as far as measurability goes (and in many other contexts as well), when μ is complete we can usually ignore sets of measure zero. Here are two situations where this can be important.

First, suppose that (X, Σ, μ) is a measure space and $f\colon X \to \overline{\mathbb{R}}$ is an extended real-valued function on X. If f is measurable and finite μ-a.e., then the set on which f takes the values $\pm\infty$ has measure zero. By changing the values of f on a null set, we can therefore create a new function g that is measurable, equals f almost everywhere, and is finite at every point of X.

Second, if we are given a function that is defined on all of X except for a subset Z that has measure zero then we can assign any values we like to $f(x)$ for $x \in Z$ without affecting the measurability of f. In this way, functions defined on $X \setminus Z$ can usually be implicitly assumed to be defined on all of X.

Notation 2.1.25. Let (X, Σ, μ) be a measure space. If Z is a μ-null set, then we often say that a function f whose domain is $X \setminus Z$ is *defined almost everywhere on X*. $\diamondsuit$

Here is a substitute for Lemma 2.1.23 that can be used even when the measure space is not complete. Actually, all that is needed in this exercise is that Z is a measurable set, it doesn't have to be a null set.

Exercise 2.1.26. Let (X, Σ, μ) be a measure space, and let f be a measurable function on X (either extended real-valued or complex-valued). Show that if Z is a μ-null set (i.e., $Z \in \Sigma$ and $\mu(Z) = 0$), then

$$g(x) \;=\; \begin{cases} f(x), & x \notin Z, \\ 0, & x \in Z, \end{cases}$$

is a measurable function on X. $\diamondsuit$

Problems

2.1.27. Let (X, Σ_X) be a measurable space, and let Y be any set. Given a function $f\colon X \to Y$, show that

$$\Sigma_Y \;=\; \bigl\{A \subseteq Y : f^{-1}(A) \in \Sigma_X\bigr\}$$

is a σ-algebra on Y.

2.1.28. Let (X, Σ) be a measure space and fix $f: X \to \overline{\mathbb{R}}$. Show that if A is a dense subset of $\mathbb{R}$, then f is measurable if and only if $\{f > a\}$ is measurable for each $a \in A$.

2.1.29. Let (X, Σ) be a measure space. Show that if $f: X \to \overline{\mathbb{R}}$ is measurable then $\{f = a\}$ is measurable in X for every $a \in \overline{\mathbb{R}}$, but the converse statement can fail.

2.1.30. Let X be a set. Characterize the functions $f: X \to \overline{\mathbb{R}}$ that are measurable with respect to the following σ-algebras on X.

(a) $\Sigma = \mathcal{P}(X)$.

(b) $\Sigma = \{\emptyset, X\}$.

(c) Σ is the σ-algebra constructed in Exercise 1.1.6.

2.1.31. Let μ be a bounded measure on a measurable space (X, Σ). Assume that $f: X \to [0, \infty]$ is measurable. Show that f is finite μ-a.e. if and only if for every $\varepsilon > 0$ there exists an $M > 0$ such that $\mu\{f > M\} < \varepsilon$.

2.2 Compositions, Sums, and Limits

2.2.1 Compositions of Measurable Functions

The next lemma shows that measurability is preserved under compositions if we have the appropriate match in the σ-algebras placed on our spaces.

Lemma 2.2.1. *Let (X, Σ_X), (Y, Σ_Y), and (Z, Σ_Z) be measurable spaces. If $f: X \to Y$ and $g: Y \to Z$ are measurable with respect to the given σ-algebras, then $g \circ f: X \to Z$ is measurable.*

Proof. We simply have to verify that if E is a measurable subset of Z then its inverse image under $g \circ f$ is a measurable subset of X. This follows from the individual measurability of f and g, for if $E \in \Sigma_Z$ then $g^{-1}(E) \in \Sigma_Y$ since g is measurable, and hence

$$(g \circ f)^{-1}(E) \;=\; f^{-1}\big(g^{-1}(E)\big) \;\in\; \Sigma_X$$

since f is measurable. $\square$

To emphasize, the hypotheses of Lemma 2.2.1 are that

$$f \text{ is } (\Sigma_X, \Sigma_Y)\text{-measurable} \qquad \text{and} \qquad g \text{ is } (\Sigma_Y, \Sigma_Z)\text{-measurable}.$$

Despite the naturalness of these hypotheses, we must be careful when attempting to apply Lemma 2.2.1 to extended real-valued or complex-valued functions. For example, suppose that we have a measurable real-valued function $f\colon X \to \mathbb{R}$, and we know that $g\colon \mathbb{R} \to \mathbb{R}$ is also measurable. Following the conventions of Definition 2.1.7, if we write "g is measurable" without qualification, it means that g is *Lebesgue measurable*. Thus, we are assuming that

$$f \text{ is } (\Sigma, \mathcal{B}_\mathbb{R})\text{-measurable} \qquad \text{and} \qquad g \text{ is } (\mathcal{L}_\mathbb{R}, \mathcal{B}_\mathbb{R})\text{-measurable.}$$

We do not have the required matchup in σ-algebras that is required in order to apply Lemma 2.2.1. In general, the composition of a measurable function $f\colon X \to \mathbb{R}$ with a measurable function $g\colon \mathbb{R} \to \mathbb{R}$ *need not be measurable*, the basic problem being that if $E \in \mathcal{B}_\mathbb{R}$ then we only know that $g^{-1}(E)$ is Lebesgue measurable, whereas we need to know that $g^{-1}(E)$ is Borel measurable in order to conclude that $f^{-1}(g^{-1}(E))$ is measurable in X. Of course, in order to fix this problem we just need to impose a slightly stronger hypothesis on g.

Lemma 2.2.2. *Let (X, Σ) be a measurable space. If $f\colon X \to \mathbb{R}$ is measurable and $g\colon \mathbb{R} \to \mathbb{R}$ is Borel measurable, then the composition $g \circ f\colon X \to \mathbb{R}$ is measurable.* $\diamondsuit$

For example, Lemma 2.1.10 says that every continuous function $g\colon \mathbb{R} \to \mathbb{R}$ is Borel measurable, so we have many ways to obtain new measurable functions from a given measurable function. We state this explicitly as follows.

Corollary 2.2.3.* *Let (X, Σ) be a measurable space. If $f\colon X \to \mathbb{R}$ is measurable and $g\colon \mathbb{R} \to \mathbb{R}$ is continuous, then $g \circ f\colon X \to \mathbb{R}$ is measurable.* $\diamondsuit$

The next corollary gives some of the most common applications of this idea. Here the *positive part* and the *negative part* of a function $f\colon X \to \mathbb{R}$ are defined just as they were in Volume 1, i.e.,

$$f^+(x) \;=\; \max\{f(x), 0\} \qquad \text{and} \qquad f^-(x) \;=\; \max\{-f(x), 0\}.$$

Corollary 2.2.4.* *Assume (X, Σ) is a measurable space and $f\colon X \to \mathbb{R}$ is a measurable function. Then each of the following are real-valued measurable functions on X:*

(a) $|f|$,

(b) $|f|^p$ *for each real number* $p > 0$,

(c) $f^+(x) = \max\{f(x), 0\}$,

(d) $f^-(x) = \max\{-f(x), 0\}$,

(e) $e^{f(x)}$, $\sin f(x)$, $\cos f(x)$.

Proof. To prove statement (b), fix any $p > 0$. The function $g(x) = |x|^p$ is a continuous mapping of $\mathbb{R}$ into itself, so if f is measurable then so is $|f| = g \circ f$. The other statements follow similarly. $\square$

Analogues of these results can be easily formulated for complex-valued functions, e.g., see Problem 2.2.20. However, there is a technicality that can be problematic when dealing with extended real-valued functions. Often, we have a measurable function $f\colon X \to \overline{\mathbb{R}}$ but we wish to compose it with a continuous or Borel measurable function g that is defined on $\mathbb{R}$ rather than $\overline{\mathbb{R}}$. The next exercise shows that as long as f does not take the values $\pm\infty$ on a set of positive measure, and as long as our measure is complete, this does not pose a problem.

Exercise 2.2.5.* Let $f\colon X \to \overline{\mathbb{R}}$ be a measurable function on a complete measure space (X, Σ, μ), and assume f is finite μ-a.e.

(a) Show that if $g\colon \mathbb{R} \to \mathbb{R}$ is Borel measurable, then $g \circ f$ is defined μ-a.e. on X and is measurable.

(b) Show that $|f|^p$ is measurable for each real number $p > 0$, and both f^+ and f^- are measurable. $\diamondsuit$

The "finite almost everywhere" assumption of Exercise 2.2.5 is a commonly encountered hypothesis. Precisely (see Notation 2.1.16), it means that $\{f = \pm\infty\}$ is a μ-measurable set that has μ-measure zero. Often we deal with a measure μ that is induced from an outer measure μ^* (for example, this is the case for Lebesgue measure). In this case all sets Z that satisfy $\mu^*(Z) = 0$ are measurable, so f is finite μ-a.e. if and only if $\mu^*\{f = \pm\infty\} = 0$.

Sometimes we can remove the finite almost everywhere requirement. The next exercise can be done directly, or by thinking about what it means for a function $g\colon \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ to be continuous.

Exercise 2.2.6. Given a measurable function $f\colon X \to \overline{\mathbb{R}}$ on a measurable space (X, Σ), show that $|f|^p$ is measurable for each real number $p > 0$, and both f^+ and f^- are measurable. $\diamondsuit$

2.2.2 Sums and Scalar Multiples

Now we will see how measurability of functions behaves with respect to the usual operations of arithmetic. We will focus on extended real-valued functions. Most of the results then carry over immediately to complex-valued functions by breaking into real and imaginary parts, and at the end of this section we state an exercise that summarizes the extension to the complex case.

Measurability under sums and scalar multiples works just as it does for Lebesgue measurable functions, so we state these results as exercises.

Exercise 2.2.7.* Let (X, Σ) be a measurable space. Show that if $f\colon X \to \overline{\mathbb{R}}$ is measurable and $c \in \mathbb{R}$, then cf and $f + c$ are both measurable.

We need to be careful when we attempt to add extended real-valued functions, because we might encounter the indeterminate form $\infty - \infty$. The following exercise addresses the case where there are no indeterminate forms.

Exercise 2.2.8.* Prove that if $f,\, g\colon X \to \mathbb{R}$ are measurable functions on a measurable space (X,Σ), then $f+g$ is a measurable function. $\diamondsuit$

If f and g are extended real-valued functions, then $f+g$ may not be defined at all points. Suppose that in addition to just having a measurable space (X,Σ), we have a measure μ on (X,Σ). If the measure μ is complete and f and g are measurable extended real-valued functions that are finite almost everywhere, then the indeterminate forms $f(x)+g(x) = \infty - \infty$ or $-\infty + \infty$ can at most occur on a set of measure zero. Hence $f+g$ is defined almost everywhere in the sense of Notation 2.1.25, and $f+g$ is measurable no matter what values we decide to assign it at points where it is undefined. We state this precisely as a theorem.

Theorem 2.2.9.* *Let (X,Σ,μ) be a complete measure space. If $f,\, g\colon X \to \overline{\mathbb{R}}$ are measurable functions that are finite μ-a.e., then $f+g$ is defined μ-a.e. on X and is measurable.* $\diamondsuit$

The next exercise addresses the more general setting where we either do not have a measure at all or do not know that f and g are finite a.e. This exercise shows that if we choose a single fixed constant and redefine $f(x)+g(x)$ to be that constant whenever it takes an indeterminate form, then $f+g$ will be measurable.

Exercise 2.2.10.* Let $f,\, g\colon X \to \overline{\mathbb{R}}$ be measurable functions on a measurable space (X,Σ). Fix $c \in \overline{\mathbb{R}}$, and show that the function

$$h(x) \;=\; \begin{cases} c, & \text{if } f(x) = \infty \text{ and } g(x) = -\infty,\\ c, & \text{if } f(x) = -\infty \text{ and } g(x) = \infty,\\ f(x)+g(x), & \text{otherwise,} \end{cases}$$

is measurable on X. $\diamondsuit$

2.2.3 Products and Quotients

We use a cute trick to show that products of real-valued measurable functions are measurable.

Theorem 2.2.11.* *If $f,\, g\colon X \to \mathbb{R}$ are measurable functions on a measurable space (X,Σ), then fg is measurable.*

Proof. The functions $f+g$ and $f-g$ are measurable by Exercise 2.2.8, so Corollary 2.2.4 implies that $(f+g)^2$ and $(f-g)^2$ are measurable. Applying Exercise 2.2.8 again, we see that

$$fg = \frac{(f+g)^2 - (f-g)^2}{4}$$

is measurable. □

Although a different proof technique is required, the next exercise states that Theorem 2.2.11 extends without change to extended real-valued functions. We do not even need to assume that f, g are finite almost everywhere.

Exercise 2.2.12.* Given measurable functions $f, g \colon X \to \overline{\mathbb{R}}$ on a measurable space (X, Σ), show that fg is measurable. ♢

When dealing with quotients, we do need to avoid the indeterminate form $1/0$. According to the next exercise, quotients of measurable functions are measurable as long as we avoid division by zero.

Exercise 2.2.13.* Let $f, g \colon X \to \overline{\mathbb{R}}$ be measurable functions on a measurable space (X, Σ).

(a) Show that if $g(x) \neq 0$ for every $x \in X$, then f/g is measurable.

(b) Assume that μ is a complete measure on (X, Σ). Show that if $g \neq 0$ μ-a.e., then f/g is measurable. ♢

2.2.4 Complex-Valued Functions

In some ways, complex-valued functions are simpler than extended real-valued functions because there is no complex infinity. If f maps X into $\mathbb{C}$, then $f(x)$ is a complex number for each $x \in X$, and therefore the real and imaginary parts of $f(x)$ are finite real numbers.

The next exercise summarizes the interaction between arithmetic and measurability for complex-valued functions.

Exercise 2.2.14.* Let $f, g \colon X \to \mathbb{C}$ be measurable functions on a measurable space (X, Σ), and prove the following statements.

(a) cf and $f + c$ are measurable for each $c \in \mathbb{C}$.

(b) $f + g$ is measurable.

(c) fg is measurable.

(d) If $g(x) \neq 0$ for every x, then f/g is measurable.

Also make the appropriate modification to part (d) if μ is a complete measure on (X, Σ) and g is nonzero μ-a.e. ♢

Let $\mathcal{F}_{\mathrm{m}}(X)$ (m for "measurable") denote the space of all complex-valued measurable functions on X, i.e.,

$$\mathcal{F}_{\mathrm{m}}(X) \;=\; \bigl\{f\colon X \to \mathbb{C} \,:\, f \text{ is measurable}\bigr\}.$$

It follows from parts (a) and (b) of Exercise 2.2.14 that $\mathcal{F}_{\mathrm{m}}(X)$ is a vector space over the complex field, and part (c) of that exercise also tells us that $\mathcal{F}_{\mathrm{m}}(X)$ is closed with respect to pointwise products of functions.

2.2.5 Suprema, Infima, and Limits

Now we turn to max, min, sup, inf, limsup, liminf, and limits. Of course, max and min are special cases of sup and inf, but it is nice to observe that we can deduce the measurability of the max and min of two real-valued functions by writing

$$\max\{f,g\} \;=\; \frac{f+g}{2} + \frac{|f-g|}{2}, \qquad \min\{f,g\} \;=\; \frac{f+g}{2} - \frac{|f-g|}{2}.$$

The next theorem gives a more general result for suprema and infima that also includes the case of extended real-valued functions.

Theorem 2.2.15.* *If (X,Σ) is a measurable space and $f_k\colon X \to \overline{\mathbb{R}}$ is measurable for each $k \in \mathbb{N}$, then*

$$g(x) \;=\; \sup_k f_k(x) \qquad \textit{and} \qquad h(x) \;=\; \inf_k f_k(x)$$

are measurable functions on X.

Proof. The proof follows by noting that

$$\Bigl\{\sup_k f_k > a\Bigr\} \;=\; \bigcup_{k=1}^{\infty} \{f_k > a\}$$

and

$$\Bigl\{\inf_k f_k > a\Bigr\} \;=\; \bigcup_{k=1}^{\infty} \{f_k < a\}. \qquad \diamondsuit$$

Since a limsup can be written as an "inf sup," we see that measurability is preserved with respect to limsups and liminfs.

Corollary 2.2.16.* *If (X,Σ) is a measurable space and $f_k\colon X \to \overline{\mathbb{R}}$ is measurable for each $k \in \mathbb{N}$, then*

$$g(x) \;=\; \limsup_{k\to\infty} f_k(x) \qquad \textit{and} \qquad h(x) \;=\; \liminf_{k\to\infty} f_k(x)$$

are measurable functions on X.

Proof. We just have to write

$$\limsup_{k\to\infty} f_k(x) \;=\; \inf_{k\in\mathbb{N}} \sup_{n\geq k} f_n(x)$$

and apply Theorem 2.2.15 twice. A similar argument applies to liminf. $\square$

Since $\lim f_k(x)$ exists if and only if $\limsup f_k(x) = \liminf f_k(x)$, we also have a result for limits. In the case where we have a complete measure on X, we only need to assume that the limit exists almost everywhere.

Corollary 2.2.17.* *Let (X,Σ) be a measurable space and let $f_k\colon X \to \overline{\mathbb{R}}$ be measurable for each $k \in \mathbb{N}$.*

(a) *If $g(x) = \lim_{k\to\infty} f_k(x)$ exists for all $x \in X$, then g is measurable.*

(b) *Assume that μ is a complete measure on (X,Σ). If $g(x) = \lim_{k\to\infty} f_k(x)$ exists for μ-a.e. $x \in X$, then g is measurable.* $\diamondsuit$

By breaking into real and imaginary parts, we see that Corollary 2.2.17 carries over without any changes to measurable complex-valued functions $f_k\colon X \to \mathbb{C}$.

Definition 2.2.18 (Pointwise Almost Everywhere Convergence). Let (X,Σ,μ) be a measure space, and let f_k, f be measurable functions on E that are either extended real-valued or complex-valued. We say that f_k *converges pointwise almost everywhere to* f if there exists a measurable set $Z \subseteq X$ such that $\mu(Z) = 0$ and

$$\forall\, x \in X\setminus Z, \quad \lim_{k\to\infty} f_k(x) \;=\; f(x).$$

We often denote pointwise almost everywhere convergence by writing $f_k \to f$ *pointwise* μ*-a.e.* or simply $f_k \to f$ μ*-a.e.* (and we may also omit writing the symbol μ if it is understood). $\diamondsuit$

Often, the functions f_k not only converge pointwise a.e. to f, but are also monotone increasing except on a set of measure zero, i.e.,

$$f_1(x) \;\leq\; f_2(x) \;\leq\; \cdots \qquad \text{for a.e. } x.$$

In this case we say that f_k *increases pointwise a.e. to* f, and denote this by writing $f_k \nearrow f$ μ-a.e., or just $f_k \nearrow f$ a.e. A similar definition is made for decreasing sequences.

*Remark 2.2.19.** Note that if μ is a complete measure, then Corollary 2.2.17 implies that measurability of the limit function f in Definition 2.2.18 follows automatically from the measurability of the functions f_k. $\diamondsuit$

Problems

2.2.20. Let (X,Σ) be a measure space. Show that if $f\colon X \to \mathbb{C}$ is measurable, then so is $|f|^p$ for each real number $p > 0$.

2.2.21. (a) Given $z \in \mathbb{C}$, define

$$\operatorname{sgn} z \;=\; \begin{cases} \dfrac{z}{|z|}, & z \neq 0, \\ 0, & z = 0. \end{cases}$$

Show that $\operatorname{sgn}\colon \mathbb{C} \to \mathbb{C}$ is Borel measurable.

(b) Show that if $f\colon (X,\Sigma) \to \mathbb{C}$ is a measurable function, then we can write f in *polar form* as $f = (\operatorname{sgn} f)\,|f|$, and both $\operatorname{sgn} f$ and $|f|$ are measurable.

2.2.22. Let (X,Σ) be a measurable space and suppose that $f_k\colon X \to \overline{\mathbb{R}}$ are such that $g(x) = \lim_{k\to\infty} f_k(x)$ exists for each $x \in X$. Given $a \in \mathbb{R}$, show that

$$\{g > a\} \;=\; \bigcup_{n=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{k=N}^{\infty} \bigl\{f_k \geq a + \tfrac{1}{n}\bigr\}.$$

Use this to give another proof that if each f_k is measurable, then g is also measurable.

2.2.23. Let (X,Σ) be a measurable space and let $f_k\colon X \to \overline{\mathbb{R}}$ be measurable for each $k \in \mathbb{N}$. Show that

$$\Bigl\{x \in X : \lim_{k\to\infty} f_k(x) \text{ exists}\Bigr\}$$

is a measurable subset of X.

2.3 L^∞ Convergence

Uniform convergence is a stronger requirement than pointwise convergence in that it requires a "simultaneity" of convergence over all of the domain rather just "individual" convergence at each x. A convenient way to view uniform convergence is in terms of the *uniform norm*

$$\|f\|_{\mathrm{u}} \;=\; \sup_{x\in X} |f(x)|. \tag{2.4}$$

Uniform convergence requires that the distance between f_k and f, as measured by the uniform norm, shrinks to zero as k increases. That is,

$$f_k \to f \text{ uniformly} \qquad \Longleftrightarrow \qquad \lim_{k\to\infty} \|f - f_k\|_{\mathrm{u}} \;=\; 0.$$

We can define uniform convergence without needing to have a measure on our domain X. However, if we do have a measure on X then sets with measure zero often "don't matter." In keeping with this philosophy, we modify the definition of uniform convergence by simply replacing the supremum that appears in equation (2.4) with an essential supremum. We introduced the essential supremum for functions on $\mathbb{R}^d$ in Volume 1, and the following definition extends this to functions on an arbitrary measure space.

Definition 2.3.1 (Essential Supremum).* Let (X, Σ, μ) be a measure space. The *essential supremum* of a measurable function $f\colon X \to \overline{\mathbb{R}}$ is

$$\operatorname*{ess\,sup}_{x \in X} f(x) \;=\; \inf\bigl\{M : f(x) \le M \;\; \mu\text{-a.e.}\bigr\}. \qquad \diamondsuit \tag{2.5}$$

We give the following name to the essential supremum of $|f|$.

Definition 2.3.2 (L^∞ norm).* Let (X, Σ, μ) be a measure space, and let f be an measurable function on X, either extended real-valued or complex-valued. The L^∞ *norm* of f is the essential supremum of $|f|$, and it is denoted by

$$\|f\|_\infty \;=\; \operatorname*{ess\,sup}_{x \in X} |f(x)|. \qquad \diamondsuit$$

It follows immediately that

$$|f(x)| \;\le\; \|f\|_\infty \quad \text{for } \mu\text{-a.e. } x \in X.$$

We always have

$$\|f\|_\infty \;\le\; \|f\|_{\mathrm{u}},$$

but the uniform norm and the L^∞ norm of f need not be equal in general. For example, if we take $X = \mathbb{R}$ and define $f(x) = 0$ for $x \ne 0$ and $f(0) = 1$ then we have $\|f\|_\infty = 0$ while $\|f\|_{\mathrm{u}} = 1$.

We introduce some additional terminology associated with the essential supremum and the L^∞ norm.

Definition 2.3.3 (Essentially Bounded Function).* Let (X, Σ, μ) be a measure space, and let f, f_k, g be measurable functions on X, either extended real-valued or complex-valued.

(a) We say that f is *essentially bounded* if $\|f\|_\infty < \infty$.

(b) The quantity $\|f - g\|_\infty$ is called the L^∞ *distance between f and g*.

(c) We say that f_k *converges to f in L^∞ norm* if

$$\lim_{k \to \infty} \|f - f_k\|_\infty \;=\; 0. \qquad \diamondsuit$$

The next exercise shows that the L^∞ norm *almost* satisfies the properties of a norm. However, there is a subtle difference, as we only obtain uniqueness up to sets of measure zero. We will discuss this difference in more detail after the exercise. To avoid technical complications, we will assume that our measure space is complete and scalars are complex in this exercise.

Exercise 2.3.4.* Let (X, Σ, μ) be a measure space, and let $\mathcal{M}_b(X)$ be the set of all measurable, essentially bounded, complex-valued functions on X, i.e.,

$$\mathcal{M}_b(X) \;=\; \bigl\{f\colon X \to \mathbb{C} \,:\, f \text{ is measurable and essentially bounded}\bigr\}.$$

It follows from Exercise 2.2.14 that $\mathcal{M}_b(X)$ is a vector space over the complex field. Prove the following statements.

(a) Nonnegativity: $0 \le \|f\|_\infty \le \infty$ for all $f \in \mathcal{M}_b(X)$.

(b) Homogeneity: $\|cf\|_\infty = |c|\,\|f\|_\infty$ for all $f \in \mathcal{M}_b(X)$ and scalars $c \in \mathbb{C}$.

(c) The Triangle Inequality: $\|f+g\|_\infty \le \|f\|_\infty + \|g\|_\infty$ for all $f, g \in \mathcal{M}_b(X)$.

(d) Almost Everywhere Uniqueness: $\|f\|_\infty = 0$ if and only if $f = 0$ a.e. $\diamondsuit$

Unfortunately, statement (b) in Exercise 2.3.4 is not quite what we need in order to say that $\|\cdot\|_\infty$ is a norm on $\mathcal{M}_b(X)$. To be a norm it must be the case that $\|f\|_\infty = 0$ if and only if f is the zero vector in $\mathcal{M}_b(X)$. The zero vector in $\mathcal{M}_b(X)$ is the zero function, the function that takes the value 0 for every $x \in X$. However, it is not usually true that $\|f\|_\infty = 0$ if and only if f is the zero function. Instead we only have the weaker statement that

$$\|f\|_\infty = 0 \qquad \Longleftrightarrow \qquad f = 0 \text{ a.e.} \tag{2.6}$$

The technical terminology for this is that $\|\cdot\|_\infty$ is a *seminorm* on $\mathcal{M}_b(X)$ rather than a *norm*. Since our philosophy is that sets of measure zero usually "don't matter," this is not really a problem—we simply accept the fact that the right-hand side of equation (2.6) says that f is zero almost everywhere, instead of saying that f is zero everywhere. Problem 2.4.11 shows one way to modify the space on which $\|\cdot\|_\infty$ is defined so that it becomes a true norm on that space.

*Example 2.3.5.** Consider $X = \mathbb{R}$ and $f = \chi_C$, the characteristic function of the Cantor set C. Since the Cantor zero has measure zero, its characteristic function satisfies $\chi_C = 0$ a.e. Therefore $\|\chi_C\|_\infty = 0$, even though χ_C is not the zero function. $\diamondsuit$

Whenever we have a norm (or a seminorm), it is important to know whether Cauchy sequences must converge with respect to that norm. Following the usual terminology for norms, we say that a sequence $\{f_k\}_{k\in\infty}$ in $\mathcal{M}_b(X)$ is *Cauchy in L^∞ norm* or *Cauchy with respect to* $\|\cdot\|_\infty$ if

$$\forall\, \varepsilon > 0, \quad \exists\, N > 0, \quad \forall\, j,k \ge N, \quad \|f_j - f_k\|_\infty < \varepsilon.$$

We will show in the next theorem that every Cauchy sequence in $\mathcal{M}_b(X)$ must converge in $\mathcal{M}_b(X)$. If $\|\cdot\|_\infty$ was a true norm on $\mathcal{M}_b(X)$ then we would describe this by saying that $\mathcal{M}_b(X)$ is a *complete normed space* or a *Banach space*. However, since $\|\cdot\|_\infty$ is not a true norm, we will not use that terminology in connection with $\mathcal{M}_b(X)$.

Theorem 2.3.6.* *Let (X, Σ, μ) be a measure space. If $\{f_k\}_{k\in\mathbb{N}}$ is a sequence in $\mathcal{M}_b(X)$ that is Cauchy with respect to $\|\cdot\|_\infty$, then there exists a function $f \in \mathcal{M}_b(X)$ such that $\|f - f_k\|_\infty \to 0$ as $k \to \infty$.*

Proof. Let $\{f_k\}_{k\in\mathbb{N}}$ be a Cauchy sequence in $\mathcal{M}_b(X)$. This means that

$$\forall\, \varepsilon > 0, \quad \exists\, N > 0, \quad \forall\, j, k \ge N, \quad \|f_j - f_k\|_\infty < \varepsilon.$$

Let $M_{jk} = \|f_j - f_k\|_\infty$. Then $|f_j(x) - f_k(x)|$ can exceed M_{jk} only for a set of x's that lie in a set of measure zero. That is, if we define

$$Z_{jk} \;=\; \bigl\{|f_j - f_k| > M_{jk}\bigr\},$$

then $\mu(Z_{jk}) = 0$. Since a countable union of zero measure sets has measure zero, it follows that

$$Z \;=\; \bigcup_{j,k\in\mathbb{N}} Z_{jk}$$

has measure zero. For each $k \in \mathbb{N}$, create a function g_k that equals f_k almost everywhere but is zero on the set Z:

$$g_k(x) \;=\; \begin{cases} f_k(x), & x \notin Z, \\ 0, & x \in Z. \end{cases}$$

By Exercise 2.1.26, we know that g_k is a measurable function.

Exercise: Show that the uniform norm of $g_j - g_k$ coincides with its L^∞ norm, which also coincides with the L^∞ norm of $f_j - g_k$, i.e.,

$$\|g_j - g_k\|_{\mathrm{u}} \;=\; \|g_j - g_k\|_\infty \;=\; \|f_j - f_k\|_\infty.$$

Consequently, $\{g_k\}_{k\in\mathbb{N}}$ is a uniformly Cauchy in $\mathcal{F}_b(X)$, the space of bounded, complex-valued functions on X. This is a complete normed space, so there exists a bounded function $g \in \mathcal{F}_b(X)$ such that $g_k \to g$ uniformly. This function g is measurable since it is the pointwise limit of measurable functions. Hence g belongs to $\mathcal{M}_b(X)$. Finally, since $g - g_k$ differs from $g - f_k$ only on a set of measure zero, we have

$$\|g - f_k\|_\infty \;=\; \|g - g_k\|_\infty \;\le\; \|g - g_k\|_{\mathrm{u}} \;\to\; 0 \quad \text{as } k \to \infty. \qquad \diamondsuit$$

*Example 2.3.7.** Let $Z_1 \subseteq Z_2 \subseteq \cdots$ be a nested increasing sequence of subsets of $\mathbb{R}$ that each have measure zero. Define

---

Note the two distinct uses of the word "complete." A *complete measure* is a measure such that every subset of a null set is measurable, while a *complete normed space* is a space such that every Cauchy sequence converges.

$$f_k(x) \;=\; \begin{cases} \frac{1}{k}, & x \notin Z_k, \\ k, & x \in Z_k. \end{cases}$$

The functions f_k do not converge pointwise to zero since there points x where $f_k(x) \to \infty$ as k increases. On the other hand, f_k converges to the zero function with respect to $\|\cdot\|_\infty$, because

$$\|0 - f_k\|_\infty \;=\; \|f_k\|_\infty \;=\; \frac{1}{k} \quad \text{as } k \to \infty. \qquad \diamondsuit$$

If we like, we could make analogous versions of Exercise 2.3.4 and Theorem 2.3.6 for extended real-valued functions that are finite almost everywhere.

2.4 Convergence in Measure

Pointwise convergence, pointwise almost everywhere convergence, uniform convergence, and L^∞ norm convergence are only some of the myriad ways in which we might say that functions f_k "converge" to a given function f. For pointwise convergence we require that $f_k(x)$ becomes close to $f(x)$ for each individual x, while for uniform convergence we require that $f_k(x)$ and $f(x)$ become simultaneously close over all x. Pointwise almost everywhere and L^∞ norm convergence are variations where we ignore sets of measure zero. Now we describe another type of convergence criterion that we will call convergence in measure.

Suppose that we have functions f_k and f on a measure space (X, Σ, μ). The idea of convergence in measure is that we require $f_k(x)$ and $f(x)$ to be close except on a set that has smaller and smaller measure. More precisely, given any fixed $\varepsilon > 0$, the set where $f_k(x)$ and $f(x)$ differ by more than ε should become smaller and smaller in measure as k increases. Here is the explicit definition for complex-valued functions.

Definition 2.4.1 (Convergence in Measure).* Let (X, Σ, μ) be a measure space, and let f_k, f be complex-valued measurable functions on E. Then we say that f_k *converges in measure to* f, and write $f_k \overset{\mathrm{m}}{\to} f$, if

$$\forall\, \varepsilon > 0, \qquad \lim_{k\to\infty} \mu\bigl\{|f - f_k| > \varepsilon\bigr\} \;=\; 0. \qquad \diamondsuit \tag{2.7}$$

It is useful to write out equation (2.7) in detail. With all the quantifiers, the requirement for convergence in measure given in (2.7) is that

$$\forall\, \varepsilon, \eta > 0, \quad \exists\, N > 0 \;\text{ such that }\; k > N \implies \mu\{|f - f_k| \geq \varepsilon\} \;<\; \eta. \tag{2.8}$$

Problem 2.4.9 gives an equivalent formulation that requires only the use of a single quantifier ε instead of two quantifiers ε, η.

We can make a similar definition for convergence in measure of extended real-valued functions, but there is a technical complication because $f - f_k$ might not be measurable. However, if our measure space is complete and the functions f_k and f are finite a.e. then this is not a problem, so we extend the definition to this case as follows.

Definition 2.4.2 (Convergence in Measure).* Let (X, Σ, μ) be a complete measure space, and let f_k, f be extended real-valued measurable functions on E that are finite a.e. Then we say that f_k *converges in measure to* f, and write $f_k \xrightarrow{\mathrm{m}} f$, if

$$\forall\, \varepsilon > 0, \quad \lim_{k\to\infty} \mu\bigl\{|f - f_k| > \varepsilon\bigr\} \;=\; 0. \qquad \diamondsuit$$

For the remainder of this section we will state results for the case of complex-valued functions, and assign the reader the task of formulating analogous results for extended real-valued functions.

Since $\{|f - f_k| > \varepsilon\}$ is the set where f and f_k differ by more than ε, these definitions are saying that the measure of this set must decrease to zero as k increases. For example, if $f_k = f$ except on a set E_k of measure $1/k$, then $f_k \xrightarrow{\mathrm{m}} f$. This example suggests that convergence in measure might imply pointwise or pointwise almost everywhere convergence, but the following exercises and examples show that the relationships among these types of convergence are not quite as straightforward as we might hope.

Exercise 2.4.3.* Show that L^∞ norm convergence implies both pointwise a.e. convergence and convergence in measure. That is, if (X, Σ, μ) is a measure space and $\|f - f_k\|_\infty \to 0$, then $f_k \to f$ pointwise a.e. and also $f_k \xrightarrow{\mathrm{m}} f$. $\diamondsuit$

Several examples from Volume 1 give some insight here.

- The "Shrinking Boxes" show that convergence in measure does not imply convergence in L^∞-norm in general.
- The "Boxes Marching to Infinity" show that pointwise a.e. convergence does not imply convergence in measure in general.
- The "Boxes Marching in Circles" show that convergence in measure does not imply pointwise a.e. convergence.

In summary, convergence in L^∞ norm implies pointwise a.e. convergence and convergence in measure, but pointwise a.e. convergence does not imply convergence in measure, and convergence in measure does not imply pointwise a.e. convergence. On the other hand, after we prove Egorov's Theorem (Theorem 2.5.1) we will be able to show that, *in a finite measure space*, pointwise a.e. convergence implies convergence in measure (see Problem 2.5.2).

In the converse direction, the "Boxes Marching in Circles" show us that convergence in measure does not imply pointwise a.e. convergence, even in a finite measure space. However, there is a weaker but still very important relation between convergence in measure and pointwise a.e. convergence. Specifically, the next theorem will show that if f_k converges in measure to f, then there is a *subsequence* of the f_k that converges pointwise almost everywhere to f.

Theorem 2.4.4.* *Let (X, Σ, μ) be a measure space. If $f_k,\ f\colon X \to \mathbb{C}$ are measurable functions such that $f_k \xrightarrow{\mathrm{m}} f$, then there exists a subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ such that $f_{n_k} \to f$ pointwise a.e.*

Proof. Assume $f_k \xrightarrow{\mathrm{m}} f$. Exercise: Show that there exist indices $n_1 < n_2 < \cdots$ such that

$$\mu\bigl\{|f - f_{n_k}| > \tfrac{1}{k}\bigr\} \;\le\; 2^{-k}, \qquad k \in \mathbb{N}.$$

Define

$$E_k \;=\; \bigl\{|f - f_{n_k}| > \tfrac{1}{k}\bigr\} \qquad \text{and} \qquad H_m \;=\; \bigcup_{k=m}^{\infty} E_k.$$

These sets are measurable since f_k and f are measurable functions. By construction we have $\mu(E_k) \le 2^{-k}$, so by subadditivity,

$$\mu(H_m) \;\le\; \sum_{k=m}^{\infty} \mu(E_k) \;\le\; \sum_{k=m}^{\infty} \frac{1}{2^k} \;=\; 2^{-m+1}.$$

Set

$$Z \;=\; \bigcap_{m=1}^{\infty} H_m.$$

For each $m \in \mathbb{N}$ we have $\mu(Z) \le \mu(H_m) \le 2^{-m+1}$, so Z has measure zero (note that $Z = \limsup E_k$ in the terminology of Problem 1.3.12).

If $x \notin Z$, then $x \notin H_m$ for some m. Hence $x \notin E_k$ for all $k \ge m$, which implies that

$$|f(x) - f_{n_k}(x)| \;\le\; \frac{1}{k} \quad \text{for all } k \ge m.$$

Thus $f_{n_k}(x) \to f(x)$ when $x \notin Z$. Since Z has measure zero, this says that the functions f_{n_k} converge pointwise almost everywhere to f. □

Unlike L^∞ convergence, it is not possible to characterize convergence in measure in terms of a norm. Problem 2.4.12 shows that it is possible to create a *metric* that corresponds to convergence in measure, but in practice it is usually more convenient work directly with the definition of convergence in measure instead of trying to use that metric. On the other hand, it is important to know whether there is a notion of sequences that are Cauchy with respect to convergence in measure, and whether such Cauchy sequences must converge in measure. The next definition introduces the Cauchy criterion.

Definition 2.4.5 (Cauchy in Measure).* Let (X, Σ, μ) be a measure space. Given measurable functions $f_k,\ f\colon X \to \mathbb{C}$, we say that $\{f_k\}_{k\in\mathbb{N}}$ is *Cauchy in measure* if

$$\forall\, \varepsilon > 0, \qquad \lim_{j,k\to\infty} \mu\{|f_j - f_k| > \varepsilon\} \;=\; 0. \tag{2.9}$$

Precisely, equation (2.9) means that

$$\forall\, \varepsilon, \eta > 0, \quad \exists\, N > 0 \;\text{ such that }\; j,k > N \;\implies\; \mu\{|f_j - f_k| > \varepsilon\} \;<\; \eta. \quad \diamondsuit$$

Now we prove that if a sequence is Cauchy in measure then it must converge in measure. The usefulness of the Cauchy criterion is that we can test for it without knowing what the limit function is, or even whether it exists.

Theorem 2.4.6 (Cauchy Criterion for Convergence in Measure). * *Let (X, Σ, μ) be a measure space. If $f_k\colon X \to \mathbb{C}$ are measurable functions on X, then the following statements are equivalent.*

(a) *There exists a measurable function f such that $f_k \xrightarrow{\mathrm{m}} f$.*

(b) *$\{f_k\}_{k\in\mathbb{N}}$ is Cauchy in measure.*

Proof. (a) $\Rightarrow$ (b). Assume that $f_k \xrightarrow{\mathrm{m}} f$, and fix ε, $\eta > 0$. Then by equation (2.8), there exists an $N > 0$ such that

$$\mu\{|f - f_k| > \varepsilon\} \;<\; \eta \qquad \text{for all } k > N.$$

By the Triangle Inequality,

$$|f_j(x) - f_k(x)| \;\le\; |f_j(x) - f(x)| + |f(x) - f_k(x)|.$$

It follows from this that

$$\{|f_j - f_k| > 2\varepsilon\} \;\subseteq\; \{|f_j - f| > \varepsilon\} \;\cup\; \{|f - f_k| > \varepsilon\}.$$

Consequently, if $k > N$ then

$$\mu\{|f_j - f_k| > 2\varepsilon\} \;\le\; \mu\{|f_j - f| > \varepsilon\} \;+\; \mu\{|f - f_k| > \varepsilon\} \;<\; \eta + \eta \;=\; 2\eta.$$

Therefore $\{f_k\}_{k\in\mathbb{N}}$ is Cauchy in measure.

(b) $\Rightarrow$ (a). The first part of this proof proceeds much like the proof of Theorem 2.4.4. Suppose that $\{f_k\}_{k\in\mathbb{N}}$ is Cauchy in measure. Then there exist indices $n_1 < n_2 < \cdots$ such that

$$\mu\bigl\{|f_{n_{k+1}} - f_{n_k}| > 2^{-k}\bigr\} \;\le\; 2^{-k}, \qquad k \in \mathbb{N}.$$

For simplicity of notation, let $g_k = f_{n_k}$. Define

$$E_k \;=\; \bigl\{|g_{k+1} - g_k| > 2^{-k}\bigr\} \qquad \text{and} \qquad H_m \;=\; \bigcup_{k=m}^{\infty} E_k.$$

Then $\mu(E_k) \le 2^{-k}$, so

$$\mu(H_m) \;\le\; \sum_{k=m}^{\infty} \mu(E_k) \;\le\; \sum_{k=m}^{\infty} 2^{-k} \;=\; 2^{-m+1}.$$

Therefore, if we set

$$Z \;=\; \bigcap_{m=1}^{\infty} H_m \;=\; \bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} E_k \;=\; \limsup E_k,$$

then Z is measurable and $\mu(Z) = 0$.

If $x \notin Z$ then $x \notin H_m$ for some m, and therefore $x \notin E_k$ for all $k \ge m$. That is, $|g_{k+1}(x) - g_k(x)| \le 2^{-k}$ for all $k \ge m$. This implies that $\{g_k(x)\}_{k\in\mathbb{N}}$ is a Cauchy sequence of complex scalars, and therefore it must converge. The function

$$f(x) \;=\; \begin{cases} \lim\limits_{k\to\infty} g_k(x), & \text{if the limit exists,} \\ 0, & \text{otherwise,} \end{cases}$$

is measurable by Corollary 2.2.17, and we have by construction that $g_k \to f$ pointwise a.e.

Now will show that g_k converges in measure to f. Fix $\varepsilon > 0$, and choose m large enough that $2^{-m} \le \varepsilon$. Fix $m \in \mathbb{N}$, and suppose that $x \notin H_m$. If $n > k > m$ then we have

$$|g_n(x) - g_k(x)| \;\le\; \sum_{i=k}^{n-1} |g_{i+1}(x) - g_i(x)| \;\le\; \sum_{i=k}^{n-1} 2^{-i} \;\le\; 2^{-k+1} \;\le\; 2^{-m} \;\le\; \varepsilon.$$

Taking the limit as $n \to \infty$, this implies that $|f(x) - g_k(x)| \le \varepsilon$ for all $x \notin H_m$ and $k > m$. Hence

$$\bigl\{|f - g_k| > \varepsilon\bigr\} \;\subseteq\; H_m, \qquad k > m,$$

so

$$\limsup_{k\to\infty} \mu\bigl\{|f - g_k| > \varepsilon\bigr\} \;\le\; \mu(H_m) \;\le\; 2^{-m+1}.$$

This is true for every m, so we conclude that $\lim_{k\to\infty} \mu\{|f - g_k| > \varepsilon\} = 0$. Therefore $g_k \overset{\mathrm{m}}{\to} f$. This is not quite enough to complete the proof, because $\{g_k\}_{k\in\mathbb{N}}$ is only a subsequence of $\{f_k\}_{k\in\mathbb{N}}$. However, by Problem 2.4.8, the fact that $\{f_k\}_{k\in N}$ is Cauchy in measure and has a subsequence that converges in measure to f implies that $f_k \overset{\mathrm{m}}{\to} f$. $\square$

Problems

2.4.7. Let μ be counting measure on $\mathbb{N}$, and let $f_n,\ f\colon \mathbb{N} \to \mathbb{R}$. Show that $f_n \to f$ in measure if and only if $f_n \to f$ uniformly.

2.4.8. Let (X, Σ, μ) be a measure space, and let $f_k,\, f\colon X \to \mathbb{C}$ be measurable functions on X. Show that if $\{f_k\}_{k\in\mathbb{N}}$ is Cauchy in measure and there exists a subsequence such that $f_{n_k} \stackrel{\mathrm{m}}{\to} f$, then $f_k \stackrel{\mathrm{m}}{\to} f$.

2.4.9. Let (X, Σ, μ) be a measure space, and let $f_k,\, f\colon X \to \mathbb{C}$ be measurable functions on X. Prove that $f_k \stackrel{\mathrm{m}}{\to} f$ if and only if

$$\forall\, \varepsilon > 0, \quad \exists\, N > 0 \ \text{ such that } \ k > N \implies \mu\{|f - f_k| \ge \varepsilon\} \ < \ \varepsilon.$$

Formulate and prove an analogous Cauchy criterion.

2.4.10. Let (X, Σ, μ) be a measure space. Let $f_k,\, f,\, g_k,\, g\colon X \to \mathbb{C}$ be measurable functions on X, and prove the following statements.

(a) If $f_k \stackrel{\mathrm{m}}{\to} f$ and $f_k \stackrel{\mathrm{m}}{\to} g$, then $f = g$ a.e.

(b) If $f_k \stackrel{\mathrm{m}}{\to} f$ and $g_k \stackrel{\mathrm{m}}{\to} g$, then $f_k + g_k \stackrel{\mathrm{m}}{\to} f + g$.

(c) If μ is a bounded measure, $f_k \stackrel{\mathrm{m}}{\to} f$, and $g_k \stackrel{\mathrm{m}}{\to} g$, then $f_k g_k \stackrel{\mathrm{m}}{\to} fg$.

(d) The conclusion of part (c) can fail if μ is not bounded.

2.4.11. Let (X, Σ, μ) be a measure space.

(a) Show that $f \sim g$ if $f = g$ a.e. is an equivalence relation on $\mathcal{M}_b(X)$.

(b) Let $\widetilde{f}$ denote the equivalence class of f under this relation, i.e.,

$$\widetilde{f} \ = \ \bigl\{g \in \mathcal{M}_b(X) \ : \ g = f \text{ a.e.}\bigr\}.$$

Define $|||\widetilde{f}\,|||_\infty = \|f\|_\infty$, and show that this quantity is independent of the choice of representative f of $\widetilde{f}$.

(c) Let the quotient space $L^\infty(X)$ consist of all the distinct equivalence classes of $f \in \mathcal{M}_b(X)$, i.e.,

$$L^\infty(X) \ = \ \bigl\{\widetilde{f} : f \in \mathcal{M}_b(X)\bigr\}.$$

Define operations of addition and scalar multiplication that make $L^\infty(X)$ into a vector space over the complex field. What is the zero element of $L^\infty(X)$?

(d) Show that $|||\cdot|||_\infty$ is a norm on $L^\infty(X)$.

2.4.12. Let (X, Σ, μ) be a measure space, and let $\mathcal{F}_{\mathrm{m}}(X)$ be the vector space of all complex-valued measurable functions on X. For each $n \in \mathbb{N}$, define

$$\rho_n(f) \ = \ \mu\bigl\{|f| > \tfrac{1}{n}\bigr\}$$

and for $f,\, g \in \mathcal{F}_{\mathrm{m}}(X)$ define

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

Prove that $\rho_n(f+g) \leq \rho_{2n}(f) + \rho_{2n}(g)$, and use this to prove the following statements.

(a) $\mathrm{d}(f,g) \geq 0$,

(b) $\mathrm{d}(f,g) = 0$ if and only if $f = g$,

(c) $\mathrm{d}(f,g) = \mathrm{d}(g,f)$, and

(d) $\mathrm{d}(f,h) \leq \mathrm{d}(f,g) + \mathrm{d}(g,h)$.

Consequently, if we identify functions that are equal almost everywhere then d is a metric on $\mathcal{F}_{\mathrm{m}}(x)$. Prove that convergence with respect to this metric coincides with convergence in measure, i.e.,

$$f_k \xrightarrow{\mathrm{m}} f \qquad \Longleftrightarrow \qquad \lim_{k\to\infty} \mathrm{d}(f, f_k) \;=\; 0.$$

2.5 Egorov's Theorem

We know that pointwise convergence of functions does not imply uniform convergence, and likewise pointwise a.e. convergence does not imply L^∞ norm convergence. A standard counterexample is the Shrinking Boxes: $\chi_{[0,\frac{1}{k}]}$ converges pointwise a.e. to the zero function, but the convergence is not uniform. The functions $\chi_{[0,\frac{1}{k}]}$ are not continuous, but this is not the issue. For example,

$$f_k(x) \;=\; \begin{cases} 0, & x \leq 0, \\ \text{linear}, & 0 < x < \frac{1}{2k}, \\ 1, & x = \frac{1}{2k}, \\ \text{linear}, & \frac{1}{2k} < x < \frac{1}{k}, \\ 0, & x \geq \frac{1}{k}. \end{cases}$$

is a continuous function and $f_k \to 0$ pointwise, but f_k does not converge uniformly to the zero function.

However, if we allow ourselves to reduce the domain of these functions, then we can find a subset on which we have uniform convergence. In particular, if $0 < \delta < 1$ then the functions f_k converge uniformly to 0 on the interval $[\delta, 1]$. Egorov's Theorem essentially states that this example is typical, as long as we are dealing with a finite measure space. So it is important in the example above that the region of interest (the interval $[0,1]$) has finite measure, but according to Egorov, whenever we have pointwise a.e. convergence of a sequence of functions in a finite measure space, the sequence will actually converge uniformly on a "large" part of the domain.

Theorem 2.5.1 (Egorov's Theorem).* *Let (X, Σ, μ) be a finite measure space. If $f_k,\ f\colon X \to \mathbb{C}$ are measurable functions such that $f_k \to f$ point-*

wise a.e., then for every $\varepsilon > 0$ *there exists a measurable set* $E \subseteq X$ *such that*

(a) $\mu(E) < \varepsilon$, *and*

(b) f_k *converges uniformly to* f *on* E^{C}, *i.e.,*

$$\lim_{k\to\infty} \Big(\sup_{x\notin E} |f(x) - f_k(x)|\Big) \;=\; 0.$$

Proof. Let Z be the set of measure zero consisting of all points $x \in X$ such that $f_k(x)$ does not converge to $f(x)$. For each $k, n \in \mathbb{N}$, define the measurable sets

$$E_k(n) \;=\; \bigcup_{m=k}^{\infty} \Big\{|f - f_m| \ge \frac{1}{n}\Big\} \qquad \text{and} \qquad Z_n \;=\; \bigcap_{k=1}^{\infty} E_k(n).$$

Fix n, and suppose that $x \in Z_n$. Then $x \in E_k(n)$ for every k, so for each k there must exist some integer $m \ge k$ such that $|f(x) - f_m(x)| > \frac{1}{n}$. Therefore $f_k(x)$ does not converge to $f(x)$ as k increases, so this point x belongs to Z. This shows that

$$Z_n \;\subseteq\; Z,$$

and therefore $\mu(Z_n) = 0$ by monotonicity. With n fixed, the sets $E_k(n)$ are nested decreasing, and their intersection is Z_n by definition. Therefore, it follows from continuity from above that

$$\forall\, n \in \mathbb{N}, \quad \lim_{k\to\infty} \mu\big(E_k(n)\big) \;=\; \mu(Z_n) \;=\; 0. \tag{2.10}$$

Fix $\varepsilon > 0$. Applying equation (2.10), for each integer n there is some integer $k_n > 0$ such that

$$\mu\big(E_{k_n}(n)\big) \;<\; \frac{\varepsilon}{2^n}.$$

Define

$$E \;=\; \bigcup_{n=1}^{\infty} E_{k_n}(n).$$

Subadditivity implies that $\mu(E) \le \varepsilon$. Further, if $x \notin E$ then $x \notin E_{k_n}(n)$ for any n, so $|f(x) - f_m(x)| < \frac{1}{n}$ for all $m \ge k_n$.

In summary, $\mu(E) \le \varepsilon$ and for each $n \in \mathbb{N}$ there exists an integer $k_n > 0$ such that

$$m \;\ge\; k_n \quad \Longrightarrow \quad \sup_{x\notin E} |f(x) - f_m(x)| \;\le\; \frac{1}{n}.$$

This says that f_k converges uniformly to f on E^{C}. $\square$

Egorov's Theorem (also known as Egoroff's Theorem) is named for the Russian mathematician Dmitri Egorov (1869–1931). The theorem was also proved independently by Carlo Severini (1872–1951).

Problems

2.5.2. Let (X, Σ, μ) be a finite measure space, and let $f_k,\ f\colon X \to \mathbb{R}$ be measurable function on X. Show that if $f_k \to f$ pointwise a.e., then $f_k \xrightarrow{\mathrm{m}} f$.

2.5.3. Let D be a Lebesgue measurable subset of $\mathbb{R}^d$ such that $|D| < \infty$. Show that if $f_k,\ f\colon D \to \mathbb{C}$ are measurable and $f_k \to f$ a.e. on D, then there exists a closed set $F \subseteq D$ such that $|D \backslash F| < \varepsilon$ and $f_k \to f$ uniformly on F.

2.5.4. Suppose that μ is a σ-finite measure on X, and $f_k,\ f\colon X \to \mathbb{C}$ are measurable functions such that $f_k \to f$ a.e. Show that there exist measurable sets $E_n \subseteq E$ such that

(a) $\mu\big(X \setminus \cup E_n\big) = 0$, and

(b) $f_k \to f$ uniformly on each set E_n.

2.6 Approximation by Simple Functions

Often, the easiest way to deal with a generic measurable function is to approximate it by simpler functions. Of course, the meaning of "simpler" is in the eye of the beholder, but one way in which a function φ can be "simple" is if it only takes finitely many different values. We will not require that the set on which φ takes a particular value have any special structure aside from being measurable. All we will require of a "simple function" is that it is measurable and takes only finitely many real or complex values (infinity is not allowed). The precise definition is as follows.

Definition 2.6.1.* Let (X, Σ) be a measurable space. A *simple function* on X is a measurable function $\varphi\colon X \to \mathbb{C}$ that takes only finitely many distinct values. $\diamondsuit$

A simple function can be real-valued, but it *cannot* take the values $\pm\infty$. In order for φ to be called a simple function, φ must be measurable, $\varphi(x)$ must be a real or complex scalar for each $x \in X$, and the set of all values of φ must be a finite set. The set of all values is just another name for

$$\operatorname{range}(\varphi) \;=\; \big\{\varphi(x) : x \in X\big\},$$

so a simple function is a measurable function whose range is a finite subset of $\mathbb{C}$.

*Example 2.6.2.** Aside from the zero function, the simplest example of a simple function is the characteristic function of a measurable set $A \subseteq X$, which is defined explicitly as

$$\chi_A(x) \;=\; \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

A characteristic function takes only two values, and we see that χ_A is a measurable function if and only if A is a measurable subset of X. $\diamondsuit$

We can make more "interesting" simple functions by forming finite linear combinations of characteristic functions. Specifically, if $E_1, \dots, E_N$ are measurable subsets of X and $c_1, \dots, c_N$ are complex scalars, then the function $\varphi = \sum_{k=1}^{N} c_k \, \chi_{E_k}$ takes only finitely many values and is measurable, so φ is a simple function on X. In fact, the next lemma (whose proof essentially follows "from inspection") states that every simple function has this form.

Lemma 2.6.3.* *Let φ be a simple function on a measurable space (X, Σ). If $c_1, \dots, c_N$ are the distinct values taken by φ and we set*

$$E_k \;=\; \varphi^{-1}\{c_k\} \;=\; \{\varphi = c_k\}, \qquad k = 1, \dots, N, \tag{2.11}$$

then

$$\varphi \;=\; \sum_{k=1}^{N} c_k \, \chi_{E_k}.$$

Moreover, the sets $E_1, \dots, E_N$ defined in equation (2.11) partition X into disjoint measurable sets. $\diamondsuit$

There may be many ways to write a given simple function as a linear combination of characteristic functions, but we encounter the particular form given in Lemma 2.6.3 often enough that we give it a special name.

Definition 2.6.4.* The *standard representation* of a simple function φ is the representation given by Lemma 2.6.3, i.e., $\varphi = \sum_{k=1}^{N} c_k \, \chi_{E_k}$ where $c_1, \dots, c_N$ are the distinct values taken by φ and $E_k = \{\varphi = c_k\}$. $\diamondsuit$

For example, the standard representation of $\varphi = \chi_{[0,2]} + \chi_{[1,3]}$ is

$$\varphi \;=\; 0\, \chi_{E_1} + 1\, \chi_{E_2} + 2\, \chi_{E_3},$$

where $E_1 = (-\infty, 0) \cup (3, \infty)$, $E_2 = [0,1) \cup (2,3]$, and $E_3 = [1,2]$. Of course, we can also write φ in the form

$$\varphi \;=\; \chi_{E_2} + 2\, \chi_{E_3},$$

but while the sets E_2, E_3 are disjoint, they do not partition $\mathbb{R}$. In general, one of the scalars c_k in the standard representation of a simple function φ might be zero.

If $\varphi = \sum_{j=1}^{M} c_j \, \chi_{A_j}$ and $\psi = \sum_{k=1}^{N} d_k \, \chi_{B_k}$ are the standard representations of the simple functions φ and ψ, then

$$\varphi + \psi \;=\; \sum_{j=1}^{M}\sum_{k=1}^{N} (c_j + d_k)\, \chi_{A_j \cap B_k}.$$

This need not be the standard representation of $\varphi+\psi$, since the scalars c_j+d_k may coincide for different values of j and k. However, it does show that the sum of two simple functions is simple. A similar idea applied to products establishes the next lemma.

Lemma 2.6.5.* *The class of simple functions on a measurable space* (X, Σ) *is closed with respect to addition, scalar multiplication, and products. That is, if* φ *and* ψ *are simple functions on* X *and* $c \in \mathbb{C}$*, then*

$$\varphi + \psi, \qquad c\varphi, \qquad \varphi\,\psi,$$

are all simple functions on X. $\diamondsuit$

Much of the power of simple functions lies in the next theorem, which states that every nonnegative function (including those that take the value ∞) can be written as a pointwise limit of a sequence of simple functions ϕ_k. In fact, we will be able to construct the simple functions ϕ_k so that they increase pointwise to f (which we denote by writing $\phi_k \nearrow f$), and the convergence is uniform on any set where f is bounded.

Theorem 2.6.6.* *Let* (X, Σ) *be a measurable space. If* $f \colon X \to [0, \infty]$ *is measurable, then there exist nonnegative simple functions* $\phi_1, \phi_2, \dots$ *that increase pointwise to* f*, i.e.,*

(a) $0 \le \phi_1 \le \phi_2 \le \cdots \le f$, *and*

(b) $\lim\limits_{k\to\infty} \phi_k(x) = f(x)$ *for each* $x \in X$.

Moreover, if f *is bounded on some set* $E \subseteq X$*, then we can construct the functions* ϕ_k *so that they converge uniformly to* f *on* E*, i.e.,*

$$\lim_{k\to\infty} \left(\sup_{x\in E} |f(x) - \phi_k(x)| \right) \;=\; 0.$$

Proof. The idea of the proof is that we construct ϕ_k by simply rounding f down to the nearest integer multiple of 2^{-k}. However, if f is unbounded then this would give ϕ_k infinitely many values, while a simple function can only take finitely many values. Hence we stop the rounding down process at some finite height. Typical choices for this height are k or 2^k; we will use the former.

Thus, for $k = 1$ we define ϕ_1 by rounding f down to the nearest integer, with the caveat that we stop at height 1 (see the illustration in Figure 2.1).

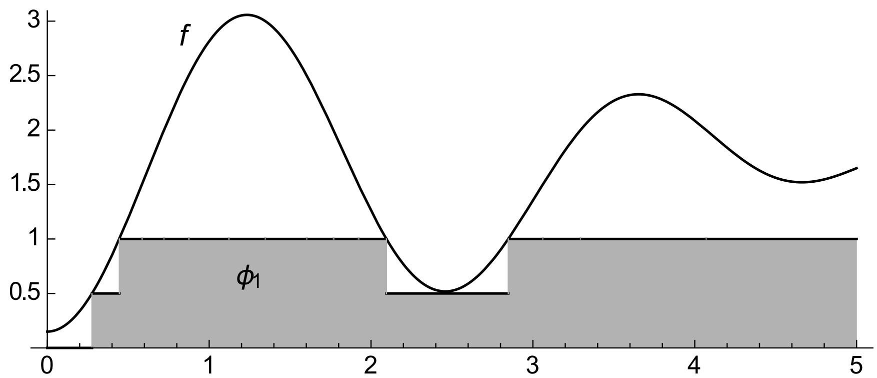


Fig. 2.1 Graphs of a function f and the approximating simple function ϕ_1 (the region under the graph of ϕ_1 is shaded).

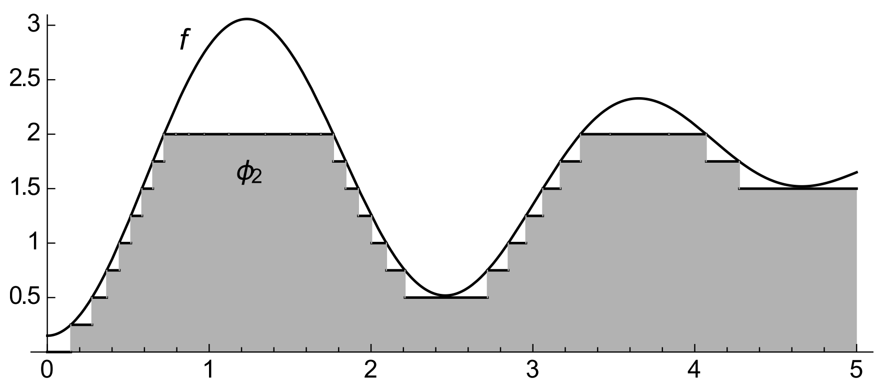


Fig. 2.2 Graphs of a function f and the approximating simple function ϕ_2 (with shading under the graph of ϕ_2).

Specifically,

$$\phi_1(x) = \begin{cases} 0, & 0 \le f(x) < 1, \\ 1, & f(x) \ge 1. \end{cases}$$

For ϕ_2 we round down to the nearest integer multiple of $\frac{1}{2}$, except we never exceed height 2 (see Figure 2.2):

$$\phi_2(x) = \begin{cases} 0, & 0 \le f(x) < \frac{1}{2}, \\ \frac{1}{2}, & \frac{1}{2} \le f(x) < 1, \\ 1, & 1 \le f(x) < \frac{3}{2}, \\ \frac{3}{2}, & \frac{3}{2} \le f(x) < 2, \\ 2, & f(x) \ge 2. \end{cases}$$

Note that if $f(x) \le 2$, then $f(x)$ and $\phi_k(x)$ differ by at most $\frac{1}{2}$.

In general, given a positive integer k we define ϕ_k by

$$\phi_k(x) \;=\; \begin{cases} \dfrac{j-1}{2^k}, & \dfrac{j-1}{2^k} \le f(x) < \dfrac{j}{2^k},\ j = 1, \dots, k2^k, \\ k, & f(x) \ge k, \end{cases} \tag{2.12}$$

The sets

$$\Bigl\{\frac{j-1}{2^k} \le f < \frac{j}{2^k}\Bigr\} \qquad \text{and} \qquad \bigl\{f \ge k\bigr\}$$

are measurable because f is measurable, so it follows that ϕ_k is a measurable function. Further, by construction we have $\phi_k(x) \le \phi_{k+1}(x)$ for every x, and

$$f(x) \;\le\; k \quad\implies\quad |f(x) - \phi_k(x)| \;\le\; 2^{-k}.$$

If $f(x)$ is finite, then k will eventually exceed $f(x)$, so we have $\phi_k(x) \to f(x)$ in this case. In fact, if $f(x) \le M < \infty$ for all x in some set E, then

$$\sup_{x \in E} |f(x) - \phi_k(x)| \;\le\; 2^{-k} \quad \text{for } k \ge M.$$

Hence ϕ_k converges uniformly to f on E in this case. On the other hand, if $f(x) = \infty$ then $\phi_k(x) = k$ for every k, so $\phi_k(x) \to f(x)$ as $k \to \infty$, and thus we still have pointwise convergence in this case. □

We can extend Theorem 2.6.6 to the case of an extended real-valued function f by writing $f = f^+ - f^-$ where f^+ and f^- are the positive and negative parts of f. Complex-valued functions can be treated similarly by breaking into real and imaginary parts. We assign the proof of this as Problem 2.6.8.

Additional Problems

2.6.7. Let (X, Σ) be a measurable space. Suppose that $f \colon X \to \mathbb{R}$ takes only finitely many distinct values, and let $E_1, \dots, E_N$ be the corresponding disjoint subsets of X on which these values are taken. Show that f is measurable if and only if $E_1, \dots, E_N$ are all measurable.

2.6.8. Let (X, Σ) be a measurable space, and let f be a measurable function on X (either extended real-valued or complex-valued). Show that there exist simple functions ϕ_k such that:

(a) $\phi_k \to f$ pointwise as $k \to \infty$,

(b) $|\phi_k(x)| \le |f(x)|$ for every k and x,

(c) the convergence is uniform on every set on which f is bounded.

3.1

~~2.2~~ Integration of Nonnegative Functions

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

Exercise (Exactly the same as Lebesgue measure)

Let ϕ be a simple function on X. Then ϕ takes only finitely many values, let the distinct values be $a_1, \ldots, a_N$. Show that

$$\phi = \sum_{k=1}^{N} a_k \chi_{E_k}$$

where the E_k are disjoint & $\bigcup E_k = X$.
This is the standard representation of ϕ.

Note that ϕ takes the value zero, say $a_1 = 0$, then we can also write

$$\phi = \sum_{k=2}^{N} a_k \chi_{E_k},$$

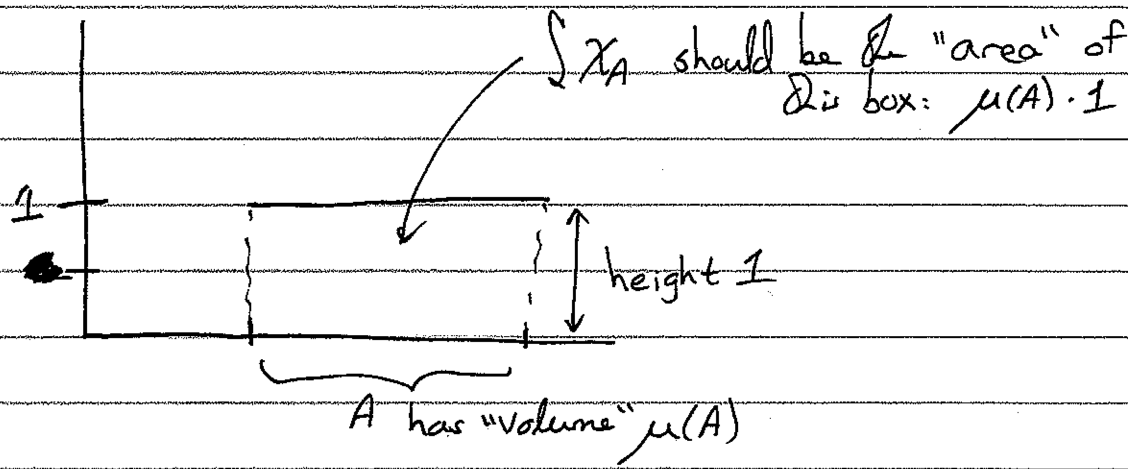
but in this case $\bigcup_{k=2}^{N} E_k \neq X$.

Notation alert

ϕ is the Greek letter "phi". Unfortunately, in my handwriting the empty set symbol looks identical: ϕ. Instead, the empty set should look more like a zero with a line through it, which is different from phi.

Idea of integration (Same as Lebesgue measure)

The integral of a characteristic function χ_A should be the measure of A:



Thus, we declare that $\int \chi_A = \mu(A)$.

Then we can easily extend this to simple functions.

Since general functions are limits of simple functions, hopefully we can extend to them as well.

Definition (Same as Lebesgue measure)

Let

$$\phi = \sum_{k=1}^{N} a_k \chi_{E_k}$$

be the standard representation of a simple function on $(X, \mathcal{M}, \mu)$. If $a_k \geq 0 \ \forall k$ (equivalent to $\phi \geq 0$), then we define

$$\int \phi \, d\mu = \sum_{k=1}^{N} a_k \, \mu(E_k).$$

Note: While each a_k must be finite, we could have $\mu(E_k) = \infty$. We declare that $0 \cdot \infty = 0$.

Notation

We use many abbreviations to denote the integral:

$$\int \phi \, d\mu = \int_X \phi \, d\mu = \int \phi = \int \phi(x) \, d\mu(x)$$

especially when we need to identify the variable of integration, as in a double integral.

Also, if $A \subseteq X$, then we set

$$\int_A \phi \, d\mu = \int \phi \cdot \chi_A \, d\mu$$

We will use the same kind of notations for integrals of general functions (once we have defined them).

Exercise

(same as Lebesgue measure)

Let ϕ, ψ be nonnegative simple functions on X. Prove the following.

a. $\int c\phi = c\int\phi \quad \forall c \geq 0.$

b. $\int \phi + \psi = \int\phi + \int\psi$

c. $\phi \leq \psi \implies \int\phi \leq \int\psi.$

Hint: If

$$\phi = \sum_{k=1}^{N} a_k \chi_{E_k} \quad \& \quad \psi = \sum_{j=1}^{M} b_j \chi_{F_j},$$

then $\{E_k \cap F_j\}_{k,j}$ is a disjoint partition of X.

Hence

$$\phi + \psi = \sum_{k=1}^{N} \sum_{j=1}^{M} (a_k + b_j)\, \chi_{E_k \cap F_j}.$$

Show that

$$\int \phi + \psi = \sum_{k=1}^{N} \sum_{j=1}^{M} (a_k + b_j)\, \mu(E_k \cap F_j)$$

even if some values of $a_k + b_j$ are equal.

Note that as a consequence of this exercise, if

$$\phi = \sum_{j=1}^{N} a_j \chi_{E_j} \quad (*)$$

is a simple function, and $E_j \in \mathcal{M}$, then

$$\int \phi \, d\mu = \sum_{j=1}^{N} a_j \, \mu(E_j),$$

regardless of whether $(*)$ is the standard representation or not.

The following seemingly esoteric result will be quite important to us.

Theorem (We had very similar results for Lebesgue measure, but did not word them in terms of creating a new measure)

If ϕ is a nonnegative simple function on X, then

$$\nu(A) = \int_A \phi \, d\mu = \int \phi \chi_A \, d\mu, \qquad A \in \mathcal{M},$$

is a measure on $\mathcal{M}$.

Remark: Note that $\mu(A) = \int \chi_A \, d\mu = \int_A d\mu$,

so if $\phi = 1$ then $\nu = \mu$.

Proof:

We have $\nu(\emptyset) = \int \phi \chi_\emptyset = \int \chi_\emptyset = \mu(\emptyset) = 0$.

phi

empty set   empty set

Now suppose that $A_1, A_2, \ldots$ are disjoint $\mathcal{M}$ sets, and let $A = \bigcup A_k$.

Let

$$\phi = \sum_{j=1}^{N_1} a_j \chi_{E_j}$$

be the standard representation of ϕ. Then

$$\nu(A) = \int \phi \chi_A$$

$$= \int \left(\sum_{j=1}^{N} a_j \chi_{E_j \cap A} \right)$$

$$= \sum_{j=1}^{N} a_j \int \chi_{E_j \cap A}$$

$$= \sum_{j=1}^{N} a_j \mu(E_j \cap A)$$

$$= \sum_{j=1}^{N} a_j \mu\left(\bigcup_{k=1}^{\infty} (E_j \cap A_k) \right)$$

disjoint union

$$= \sum_{j=1}^{N} a_j \sum_{k=1}^{\infty} \mu(E_j \cap A_k)$$

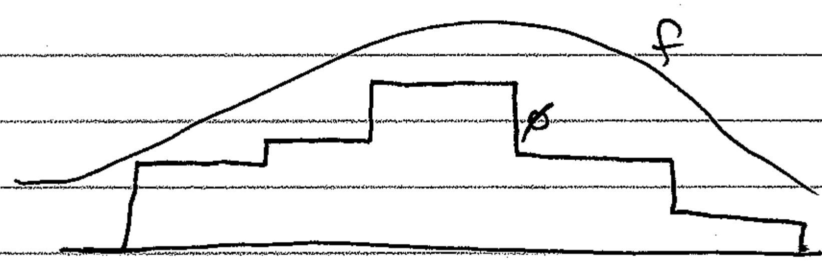
$$= \sum_{k=1}^{\infty} \sum_{j=1}^{N} a_j \mu(E_j \cap A_k) \qquad \text{all terms} \geq 0$$

$$= \sum_{k=1}^{\infty} \int \phi \chi_{A_k} \, d\mu$$

$$= \sum_{k=1}^{\infty} \nu(A_k). \qquad \blacksquare$$

Integration of Nonnegative Functions

We pass to integrals of general functions by approximating them by simple functions.



Thinking of an integral as somehow being the "area under the curve", if $f \geq 0$ is (m) and ϕ is a simple function with $0 \leq \phi \leq f$, then we should have that

$$0 \leq \int \phi \leq \int f$$

And by choosing different simple functions, we should be able to approximate better & better the area under the curve.

This is just motivation however. For one thing, there is no $\int f$ yet, and for another our picture shows f as a continuous function & ϕ as a step function, neither of which need hold in general.

However, this does suggest that $\int f$ should be an upper bound to all the integrals of simple fuctions $0 \le \phi \le f$, and in fact we simply declare $\int f$ to be the supremum of those integrals.

Definition (same as Lebesgue measure)

If $f : X \longrightarrow [0, \infty]$ is m, then

$$\int f \, d\mu = \sup \left\{ \int \phi \, d\mu : 0 \le \phi \le f, \ \phi \text{ simple} \right\}$$

Notation

As before, we write the integral in many ways:

$$\int f \, d\mu = \int f = \int_X f = \int f(x) \, d\mu(x).$$

The Lebesgue Integral

If $X \subseteq \mathbb{R}^d$ is Lebesgue m & μ is Lebesgue measure (restricted to X), then $\int f \, d\mu$ is the Lebesgue Integral of f, which we usually write as

$$\int_X f(x) \, dx \quad \text{or} \quad \int_X f$$

Note (same as Lebesgue measure)

Every nonnegative (m) function f has a uniquely defined integral, & this integral lies in the range

$$0 \leq \int f \leq \infty$$

Exercise (same as Lebesgue measure)

If $f = \sum_{j=1}^{N} a_j \chi_{E_j}$ is a simple function,

then we now have two definitions of the integral:

$$\int f = \sum_{j=1}^{N} a_j \mu(E_j)$$

and

$$\int f = \sup \left\{ \int \phi : 0 \leq \phi \leq f, \ \phi \text{ simple} \right\}$$

Show these are equal.

(specific to Lebesgue measure)

Example: Relation between Riemann & Lebesgue Integral

Suppose $f: [a,b] \longrightarrow [0,\infty)$ is continuous.

The Riemann integral of f is defined by taking the limit of Riemann sums. Each Riemann sum is an approximation by a very simple function, namely, a step function. For example if we choose a partition $\Gamma = \{a = x_0 < x_1 < \cdots < x_n = b\}$

Then the lower Riemann sum is

$$L_\Gamma = \sum_{k=1}^{n} a_k \Delta x_k \quad \text{where} \quad a_k = \min_{x \in [x_{k-1}, x_k]} f(x)$$

$$\Delta x_k = x_k - x_{k-1}$$

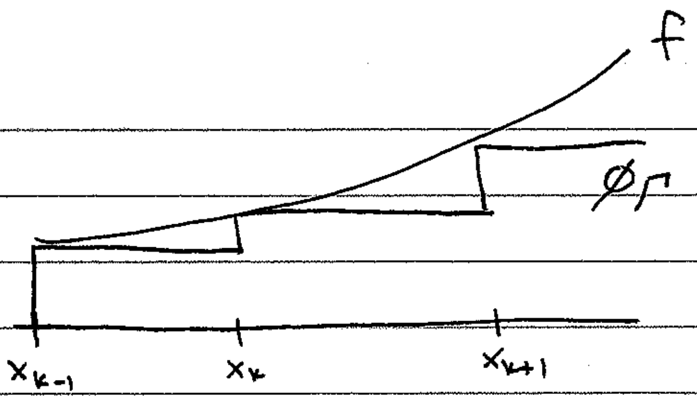
In terms of a simple function,

$$L_\Gamma = \int \phi_\Gamma \quad \text{where} \quad \phi_\Gamma = \sum_{k=1}^{n} a_k \chi_{[x_{k-1}, x_k)}$$

Since ~~the~~ is a lower Riemann sum, we have

$$\int \phi_\Gamma = L_\Gamma \leq \int f(x)\,dx$$

↑ Riemann integral



Further, as the mesh size $|\Gamma| = \max \Delta x_k$ shrinks to zero, $\int \phi_\Gamma$ converges to the Riemann integral $\int f(x)\,dx$ by definition.

So for example we could choose a countable sequence of partitions Γ_n & find corresponding simple functions $0 \le \phi_n \le f$ such that

$$\underbrace{\int \phi_n}_{} \longrightarrow \int f \text{ (Riemann integral)}.$$

Each is $\le \int f$ (Lebesgue), so

$$\int f \text{ (Riemann)} \le \int f \text{ (Lebesgue)}.$$

If we only knew that $\int \phi_n \longrightarrow \int f$ (Lebesgue), then we would conclude that

$$\int f \text{ (Riemann)} = \int f \text{ (Lebesgue)}$$

And in fact, our next big Theorem will tell us this is true, as long as we choose the ϕ_n to be monotone, i.e.,

$$\phi_1 \leq \phi_2 \leq \phi_3 \leq \cdots \longrightarrow f$$

And we can make sure this happens by choosing the partition Γ_{n+1} to be a refinement of Γ_n for each n. So:

$$\begin{array}{l} f \text{ continuous,} \\ \text{nonnegative} \\ \text{on } [a,b] \end{array} \implies \begin{array}{l} \text{Riemann integral} \\ = \text{Lebesgue integral.} \end{array}$$

Thus all the formulas we know & love from calculus are still valid. However, <u>now</u> we can integrate <u>any</u> m function, not just continuous ones.

Example

Let $f = \chi_{\mathbb{Q}}$, the characteristic function of the rationals. This function is not Riemann integrable. However, since it is a simple function, it is Lebesgue integrable:

$$\int_0^1 f(x)\,dx = \int_{[0,1]} \chi_{\mathbb{Q}} = |\mathbb{Q} \cap [0,1]| = 0.$$

Note that $f = 0$ a.e. and its integral is zero.

(same as Lebesgue measure)

Exercises

Suppose $f, g : X \to [0,\infty]$ are (m). Show directly from the definition of the integral that

a. $f \le g \Rightarrow \int f \le \int g$

b. $\int cf = c \int f \quad \forall c \ge 0$

c. $A \subseteq B$ (m) $\Rightarrow \int_A f \le \int_B f$

Try to use the definition to find the relationship between

$$\int (f+g) \quad \text{and} \quad \int f + \int g.$$

You'll find that you can easily prove an inequality, but it is not at all obvious whether equality will hold. It is, but we need better "control" of $\int f$ before we can prove it.

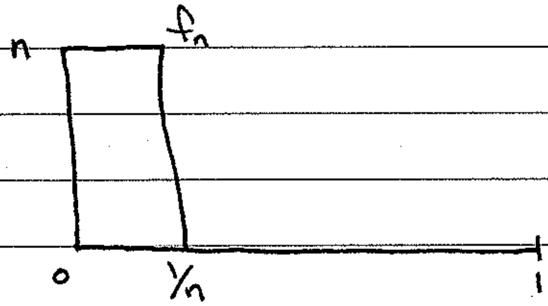
Switching Integrals & Limits: Beware!

In general, we CANNOT interchange integrals and limits!!

Example

$$f(x) = \lim_{n\to\infty} f_n(x) \;\not\Longrightarrow\; \int f = \lim_{n\to\infty} \int f_n \quad !$$

Consider $f_n = n\chi_{[0,1/n]}$:



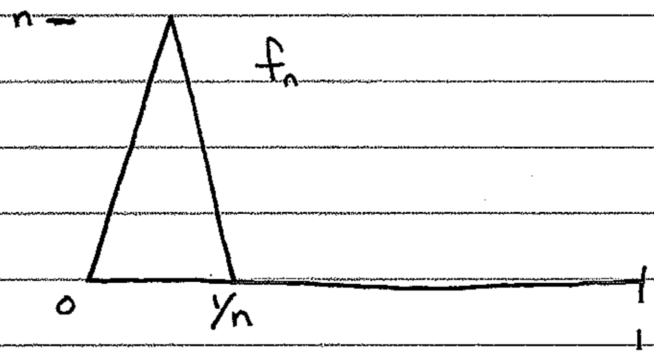
This is a simple function, so we can integrate it:

$$\int_0^1 f_n = n\,\left|[0,\tfrac{1}{n}]\right| = n\cdot\frac{1}{n} = 1 \quad \forall n$$

But
$$\lim_{n\to\infty} f_n(x) = \infty\chi_{\{0\}}(x) = \begin{cases} \infty, & x=0 \\ 0, & x>0 \end{cases}$$

and
$$\int_0^1 \chi_{\{0\}} = \infty\cdot|\{0\}| = 0$$

Note that the problem is <u>NOT</u> that the f_n are discontinuous! If we accept the equality of Riemann & Lebesgue integrals then we can make a continuous version of this example as follows.



We have $f(x) = \lim_{n\to\infty} f_n(x) = 0 \quad \forall x,$ but

$$\int_0^1 f_n = \frac{1}{2} n \cdot \frac{1}{n} = \frac{1}{2} \not\to 0 = \int_0^1 f \quad !$$

Note that if we changed the height to n^2 instead of n, we would have $\int_0^1 f_n \to \infty$ even though $\lim_{n\to\infty} f_n(x) = 0 \;\; \forall x$ still!

We need more that just <u>pointwise</u> convergence in order to draw conclusions about limits of integrals. We will see several versions of this, beginning with the following.

(same as Lebesgue measure)

Monotone Convergence Theorem (Beppo Levi Theorem)

Suppose $f_n : X \longrightarrow [0,\infty]$ is ⓜ $\forall n \in \mathbb{N}$. If

$$0 \leq f_1 \leq f_2 \leq \cdots$$

and we set $f(x) = \lim_{n\to\infty} f_n(x) = \sup_n f_n(x)$, then

$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

Proof:

By previous exercises, $0 \leq f_1 \leq f_2 \leq \cdots \leq f$ implies

$$0 \leq \int f_1 \leq \int f_2 \leq \cdots \leq \int f \quad \forall n \in \mathbb{N}.$$

Consequently $\lim_{n\to\infty} \int f_n$ exists as an extended real number, and we have

$$\lim_{n\to\infty} \int f_n \leq \int f.$$

Now fix any simple function with $0 \leq \phi \leq f$, and any $0 < \alpha < 1$. Consider the set

$$E_n = \{ f_n \geqslant \alpha \phi \}.$$

Since the f_n are increasing, we have

$$E_1 \subseteq E_2 \subseteq \cdots$$

Also, since $f_n(x) \nearrow f(x)$ and $\alpha\phi(x) < \phi(x) \leq f(x)$, we have $f_n(x) \geq \alpha\phi(x)$ for n large enough, & hence

$$\bigcup E_n = X.$$

Now,

$$\int f_n \geq \int_{E_n} f_n \geq \int_{E_n} \alpha\phi = \alpha \int_{E_n} \phi$$

$$\Big\downarrow \ (*)\ \text{CLAIM}$$

$$\alpha \int \phi$$

If we believe the claim, then we conclude that

$\lim_{n\to\infty} \int f_n \geq \alpha \int \phi$ for all $\alpha < 1$, & hence

$\lim_{n\to\infty} \int f_n \geq \int \phi$. As this is true for all simple $0 \leq \phi \leq f$, we conclude that

$$\lim_{n\to\infty} \int f_n \geq \sup\left\{\int \phi : \text{simple } 0 \leq \phi \leq f\right\} = \int f.$$

So, we just have to establish the claim.

By an earlier result,

$$\nu(E) = \int_E \phi \, d\mu, \quad E \in \mathcal{M},$$

is a measure. Since $E_1 \subseteq E_2 \subseteq \cdots$ and $\bigcup E_n = X$, it therefore follows from continuity from below that

$$\int_{E_n} \phi \, d\mu = \nu(E_n) \longrightarrow \nu(X) = \int \phi \, d\mu. \quad \blacksquare$$

<u>Remark</u>

Because of the MCT, in order to compute $\int f$, instead of taking the supremum of $\int \phi$ over <u>all</u> simple $0 \leq \phi \leq f$, we can compute it as a <u>limit</u> of $\int \phi_n$ where ϕ_n is any sequence of functions (simple or otherwise) that increase to f, such as the simple functions constructed in an earlier theorem. This is extremely useful, because limits are linear where suprema are not in general.

<u>Example</u>

Let μ be counting measure on X:

$$\mu(E) = \begin{cases} \# E, & \text{if } E \text{ is finite} \\ \infty, & \text{if } E \text{ is infinite.} \end{cases}$$

If $\phi = \sum_{j=1}^{n} a_j \chi_{E_j}$ is simple, then

$$\int \phi \, d\mu = \sum_{j=1}^{n} a_j \, \# E_j .$$

Suppose $f : X \longrightarrow [0,\infty]$ is nonzero at only countably many points in X, say $\{x_k\}_{k \in \mathbb{N}}$.

Then

$$\phi_n = \sum_{j=1}^{n} f(x_j) \chi_{\{x_j\}} \nearrow \sum_{n=1}^{\infty} f(x_j) \chi_{\{x_j\}} = f,$$

So

$$\int f \, d\mu = \lim_{n \to \infty} \int \phi_n = \sum_{j=1}^{\infty} f(x_j) \quad (\text{could be } \infty).$$

Thus, an infinite series is just an integral with respect to a counting measure.

To make this even simpler, consider $X = \mathbb{N}$.

Again let μ be counting measure. Every sequence $\{a_k\}_{k\in\mathbb{N}}$ of numbers $a_k \in [0,\infty]$ is a function on $\mathbb{N}$, namely

$$f : \mathbb{N} \to [0,\infty]$$
$$k \mapsto a_k.$$

Hence

$$\sum_{k=1}^{\infty}{}' a_k = \sum_{k=1}^{\infty}{}' f(k) = \int_{\mathbb{N}} f \, d\mu.$$

Infinite series are just special cases of integrals.

Exercise

Let δ be the delta measure on $\mathbb{R}^d$. Find a formula for

$$\int f \, d\delta, \qquad f : \mathbb{R}^d \to [0,\infty].$$

Theorem: Additivity of the Integral. (same as Lebesgue measure)

If $f_n : X \to [0,\infty]$ are ⓜ $\forall n \in \mathbb{N}$, then

$$\int\left(\sum_{n=1}^{\infty} f_n\right) d\mu = \sum_{n=1}^{\infty} \int f_n \, d\mu$$

Remarks

a. By taking infinitely many f_n to be the zero function, this result also applies to finite sums.

b. Since $\sum_{n=1}^{\infty} f_n(x)$ is a series of extended real numbers $0 \leq f_n(x) \leq \infty$, it converges at each x to an extended real number.

c. If we rewrite the infinite series as an integral with respect to counting measure, then this theorem is a result about switching the order of integration for nonnegative functions. We will later see Tonelli's Theorem, which says that if $F \geq 0$ is ⓜ with respect to the product measure, then

$$\int\int F(x,y)\, d\mu(x)\, d\nu(y) = \int\int F(x,y)\, d\nu(y)\, d\mu(x),$$

where μ, ν can be any measures.

Proof:

Let us begin with two (m) functions $f, g : X \to [0,\infty]$.

Choose simple functions $0 \le \phi_n \nearrow f$ and $0 \le \psi_n \nearrow g$. Then $\phi_n + \psi_n$ is simple, & $0 \le \phi_n + \psi_n \nearrow f+g$. Hence

$$\int (f+g) = \lim_{n\to\infty} \int (\phi_n + \psi_n)$$

$$= \lim_{n\to\infty} \left(\int \phi_n + \int \psi_n \right) \quad \text{because } \phi_n, \psi_n \text{ are } \underline{\text{simple}}$$

$$= \lim_{n\to\infty} \int \phi_n + \lim_{n\to\infty} \int \psi_n \quad \underline{\text{linearity}} \text{ of the limit}$$

$$= \int f + \int g.$$

By induction, this extends to any finite number of functions.

Now consider countably many (m) $f_n : X \to [0,\infty]$. Set

$$f = \sum_{k=1}^{\infty} f_k \quad \text{and} \quad g_N = \sum_{k=1}^{N} f_k .$$

Then $0 \le g_N \nearrow f$, so by the MCT,

$$\int \sum_{k=1}^{\infty} f_k = \int f$$

$$= \int \lim_{N\to\infty} g_N$$

$$= \lim_{N\to\infty} \int g_N$$

$$= \lim_{N\to\infty} \int \sum_{k=1}^{N} f_k$$

$$= \lim_{N\to\infty} \sum_{k=1}^{N} \int f_k$$

$$= \sum_{k=1}^{\infty} \int f_k. \quad \blacksquare$$

Sets of Measure Zero

Now we show that as far as the integral is concerned, changing values of a function on a set of measure zero has no effect.

Theorem (same as Lebesgue measure)

Let $f: X \to [0,\infty]$ be (m). Then

$$\int f = 0 \iff f = 0 \text{ a.e.}$$

Proof:

$\Leftarrow$ Exercise: Show that if $\phi = \sum_{j=1}^{N} a_j \chi_{E_j}$ is simple and $\phi = 0$ a.e. then $\int \phi = 0$.

Consequently, if $f = 0$ a.e. & $0 \le \phi \le f$ with ϕ simple, then $\phi = 0$ a.e. so

$$\int f = \sup\left\{ \int \phi : 0 \le \phi \le f, \ \phi \text{ simple} \right\} = 0.$$

$\Rightarrow$ Suppose that $\int f = 0$. Set

$$E = \{f > 0\} \quad \text{and} \quad E_n = \left\{f > \frac{1}{n}\right\}$$

For each $n \in \mathbb{N}$, we have

$$0 = \int f \geq \int_{E_n} f \geq \int_{E_n} \frac{1}{n} = \frac{1}{n} \mu(E_n) \geq 0.$$

Consequently $\mu(E_n) = 0$ for every n. Since $E = \bigcup E_n$, it follows that

$$\mu\{f \neq 0\} = \mu(E) = 0.$$

Thus $f = 0$ a.e. $\blacksquare$

<u>Corollary</u> (a.e. version of the MCT)

(same as Lebesgue measure, except that since \mu might not be complete we have to assume the measurability of f)

[redacted]

If $f_n, f : X \to [0, \infty]$ are (m) and $f_n(x) \nearrow f(x)$ for a.e. $x \in X$, then

$$\lim_{n\to\infty} \int f_n = \int f.$$

<u>Proof:</u>

Let $E = \{x \in X : f_n(x) \nearrow f(x)\}$, so $\mu(E^c) = 0$. Set

$$g_n(x) = \begin{cases} f_n(x), & x \in E \\ 0, & x \notin E \end{cases} \qquad g(x) = \begin{cases} f(x), & x \in E \\ 0, & x \notin E \end{cases}$$

Then g_n, g are (m) (why?), and $g_n(x) \nearrow g(x)$ for <u>every</u> x, so the MCT implies

$$\lim_{n\to\infty} \int f_n = \lim_{n\to\infty} \int g_n = \int g = \int f. \quad \blacksquare$$

<u>Remark</u>

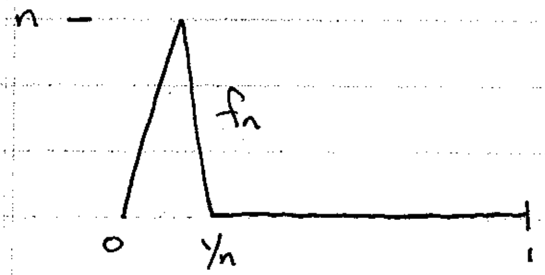
If μ is complete, then by one of our homework exercises, we don't need to assume that f is (m) in the preceding corollary:

$$\left.\begin{array}{l} \mu \text{ complete, } f_n : X \to [0,\infty) \text{ (m)}, \\ f(x) = \lim_{n\to\infty} f_n(x) \text{ for a.e. } x \end{array}\right\} \implies f \text{ is (m)}.$$

Fatou's Lemma

Although pointwise convergence of functions does not imply convergence of their integrals, we will see that at least an inequality does hold.

Example
Consider again this example:



We have $f_n(x) \longrightarrow 0$ pointwise. Although $\int f_n \not\to \int 0$, we do have an inequality:

$$\int 0 = 0 \leq \frac{1}{2} = \lim_{n\to\infty} \int f_n .$$

Fatou's lemma says this inequality will always hold for nonnegative functions. In fact, even if the f_n do not converge pointwise, we still obtain a useful inequality.

Theorem: Fatou's Lemma (same as Lebesgue measure)

If $f_n : X \longrightarrow [0,\infty]$ are any (m) functions, then

$$\int \left(\liminf_{n\to\infty} f_n\right) \leq \liminf_{n\to\infty} \int f_n$$

In particular, if $f(x) = \lim_{n\to\infty} f_n(x)$ exists $\forall x$, then

$$\int f \leq \liminf_{n\to\infty} \int f_n$$

Proof:

Note that the function

$$f(x) = \liminf_{n\to\infty} f_n(x) = \lim_{k\to\infty} \inf_{n\geq k} f_n(x) \quad \text{is (m).}$$

For each $k \in \mathbb{N}$, define

$$g_k = \inf_{n\geq k} f_n.$$

For any $m \geq k$, we have $g_k \leq f_m$, so

$$\int g_k \leq \int f_m \quad \forall m \geq k$$

and therefore

$$\int g_k \leq \inf_{m\geq k} \int f_m.$$

On the other hand, $g_k(x) \nearrow f(x)$, so the MCT implies that

$$\int f = \lim_{k\to\infty} \int g_k \quad \text{(MCT)}$$

$$\leq \lim_{k\to\infty} \inf_{m\geq k} \int f_m$$

$$= \liminf_{m\to\infty} \int f_m. \quad \blacksquare$$

<u>Exercise</u>

We deduced Fatou's Lemma from the MCT. Show that these theorems are actually equivalent, i.e., show that Fatou's Lemma implies the MCT.

3.2 ~~2.3~~ Integration of Complex Functions

So far we have only integrated nonnegative functions. Now we will extend to general extended real-valued or complex-valued functions.

Recall that if $f: X \to \overline{\mathbb{R}}$, then the functions

$$f^+ = \max\{f, 0\} \quad \& \quad f^- = -\min\{f, 0\}$$

are (m) & nonnegative. Further,

$$f = f^+ - f^- \quad \& \quad |f| = f^+ + f^-.$$

Definition (same as Lebesgue measure)

If $f: X \to \overline{\mathbb{R}}$ is (m) and at least one of $\int f^+$, $\int f^-$ is finite, then we define

$$\int f = \int f^+ - \int f^-.$$

If both $\int f^+ = \infty$ & $\int f^- = \infty$, then $\int f$ is undefined.

<u>Example</u>

Even though the improper Riemann integral

$$\int_0^\infty \frac{\sin x}{x}\,dx = \lim_{b\to\infty}\int_0^b \frac{\sin x}{x}\,dx = \frac{\pi}{2}$$

exists, the <u>Lebesgue</u> integral of $\frac{\sin x}{x}$ on $[0,\infty)$ does not exist, because

$$\int_0^\infty \left(\frac{\sin x}{x}\right)^+ dx = \infty = \int_0^\infty \left(\frac{\sin x}{x}\right)^- dx.$$

<u>Notation</u>

a. Given $f: X \to \overline{\mathbb{R}}$, if both

$$\int f^+ < \infty \quad \& \quad \int f^- < \infty$$

then we say that f is <u>integrable</u>. (same as Lebesgue measure)

b. Although this terminology is not used as commonly, it is convenient to say that f is an <u>extended integrable</u> function if

$$\text{EITHER} \int f^+ < \infty \quad \text{OR} \quad \int f^- < \infty.$$

That is, f is extended integrable $\iff \int f$ exists.

Exercise (same as Lebesgue measure)

Show that

$$f \text{ is integrable} \iff \int |f| < \infty.$$

Definition (similar to Lebesgue measure)

We let $L^1(X)$ denote the space of integrable functions on X:

$$L^1(X) = \left\{ f: X \to \overline{\mathbb{R}} \;:\; \int |f| < \infty \right\}$$

Note the implicit dependence on μ in the notation. There are many other common notations for the space, e.g.,

$$L^1,\ L^1(\mu),\ L^1(X,\mu),\ L^1(X,\mathcal{M},\mu),$$

$$L^1(d\mu),\ L^1(X,d\mu),\ \text{etc.}$$

Definition (same as Lebesgue measure)

If $f: X \to \overline{\mathbb{R}}$ is integrable, then we define its <u>L^1-norm</u> to be

$$\|f\|_1 = \int |f| \, d\mu.$$

Again note the implicit dependence on μ in the notation.

Despite its name, the L^1 norm is not a norm.

Exercise (same as Lebesgue measure)

Show that $\|\cdot\|_1$ is a seminorm on $L^1(X)$, i.e.,

a. $0 \le \|f\|_1 < \infty \quad \forall f \in L^1(X)$,

b. $\|cf\|_1 = |c|\,\|f\|_1 \quad \forall f \in L^1(X),\ \forall c \in \mathbb{R}$

c. Triangle Inequality:

$$\|f+g\|_1 \le \|f\|_1 + \|g\|_1 \quad \forall f,g \in L^1(X).$$

Show that, as a consequence, $L^1(X)$ is a vector space under the operations of function addition & scalar multiplication.

Remark

In order to be called a norm, $\|\cdot\|_1$ would have to have the additional property that

$$\|f\|_1 = 0 \implies f = 0.$$

Unfortunately, this is not true in general. Instead, show that we only have

$$\|f\|_1 = 0 \implies f = 0 \text{ a.e.}$$

The standard way to turn a seminorm into a norm is to form equivalence classes.

Exercise (same as Lebesgue measure)

a. Show that the relation

$$f \sim g \quad \text{if} \quad f = g \text{ a.e.}$$

is an equivalence relation on $L^1(X)$.

b. Let

$$\tilde{f} = \{ g \in L^1(X) : f = g \text{ a.e.} \},$$

i.e., $\tilde{f}$ is the equivalence class of f under the relation $\sim$. Define

$$|||\tilde{f}|||_1 = \|f\|_1.$$

Show that this quantity is well-defined, i.e., it is independent of the choice of representative f.

c. Let $\widetilde{L^1(X)}$ be the quotient space with respect to this relation, i.e.,

$$\widetilde{L^1}(X) = L^1(X)/\sim \;=\; \{\tilde{f} : f \in L^1(X)\}.$$

Show that $\widetilde{L^1}(X)$ is a normed space with respect to $|||\cdot|||_1$. Note that the zero vector in $\widetilde{L^1}(X)$ is

$$\tilde{0} = \{g \in L^1(X) : g = 0 \text{ a.e.}\}$$

Remark

In effect, in passing from $L^1(X)$, whose elements are functions, to $\widetilde{L^1}(X)$, whose elements are equivalence classes of functions, we are "identifying" all functions that are equal a.e. In other words, if f & g are equal a.e., then they define the same equivalence class: $\tilde{f} = \tilde{g}$. Typically, we abuse notation and let the symbol "f" denote both the function f &

the equivalence class $\tilde{f}$ of all functions equal to f a.e., and we write $L^1(X)$ instead of $\widetilde{L^1}(X)$. We will do exactly this from now on, i.e., whenever we write $L^1(X)$ we really mean the space $\widetilde{L^1}(X)$. With this abuse of notation, any two functions that are equal a.e. "are" the same element of $L^1(X)$. In particular, any function that is zero a.e. "is" the zero vector in $L^1(X)$.

As long as we take a little care, ignoring the distinction between a function & its equivalence class is not usually much of a problem. One thing we have to careful of is that a "function" f in $L^1(X)$ does not have function values, as we can change the values on any set of measure zero and still have the <u>same</u> element of $L^1(X)$!

<u>Examples</u> (same as Lebesgue measure)

If we say "$f \in L^1(X)$ is continuous", we mean that f is equal a.e. to some continuous function, i.e., one of the representatives of the equivalence class of f is continuous.

On the other hand, if g is a continuous & integrable function, then, as an element of $L^1(X)$, g and any function equal to it a.e. determine the same element of $L^1(X)$.

<u>Exercise</u> (same as Lebesgue measure)

Show that $\quad f \in L^1(X) \implies f$ is finite a.e.

Give an example showing that the converse need not hold.

Complex-Valued Functions

Definition (same as Lebesgue measure)
Given $f: X \to \mathbb{C}$ (m), write $f = \operatorname{Re} f + i \operatorname{Im} f$.

If $\int \operatorname{Re} f$ and $\int \operatorname{Im} f$ both exist and are finite

then we define

$$\underbrace{\int f = \int \operatorname{Re} f + i \int \operatorname{Im} f.}_{\text{Note this must be a complex number}}$$

Note that $\int f$ exists if & only if both $\operatorname{Re} f, \operatorname{Im} f$

are integrable. In this case, we say f is integrable

Exercise (same as Lebesgue measure)
f is integrable $\iff \int |f| < \infty$.

Definition (same as Lebesgue measure)

$$L^1(X) = \left\{ f : X \to \mathbb{C} \ : \ \int |f| < \infty \right\}$$

Note that we use the same symbols, $L^1(X)$, to denote the space of integrable extended-real functions and integrable complex functions. Which one is meant is usually clear from context.

All the same remarks apply: $L^1(X)$ is a seminormed space w.r.t. $\|f\|_1 = \int |f|$, and "becomes" a normed space when we identify functions that are equal a.e.

Exercise (same as Lebesgue measure)

Prove the linearity of the integral on $L^1(X)$, in both the extended real & complex settings:

$$\int (af+bg) = a\int f + b\int g \quad \forall f,g \in L^1(X),\ \forall \text{ scalars } a,b$$

Scalars are either real or complex according to context.

Basic Properties of the Integral

Theorem (same as Lebesgue measure)

If $f \in L^1(X)$, then $\left|\int f\right| \le \int |f|$.

Proof:

Exercise: Do the real case.

For the complex case, Write

$$\left|\int f\right| = \alpha \int f \quad \text{where } |\alpha| = 1.$$

Then

$$\int \alpha f = \alpha \int f = \left|\int f\right| \quad \text{all are } \underline{\text{real}}.$$

Write

$$\alpha f = f_1 + i f_2 \quad \text{where } f_1, f_2 \text{ are real.}$$

Then

$$\left|\int f\right| = \int \alpha f = \int f_1 + i \int f_2 \quad \text{is } \underline{\text{real}}$$

so we must have $\int f_2 = 0$ (does not imply $f_2 = 0$ a.e.)

Therefore,

$$\left|\int f\right| = \int f_1 \le \int |f_1| \le \int |f|$$

real case $\qquad$ $|f_1| \le |\alpha f| = |f|.$

∎

<u>Theorem</u> (same as Lebesgue measure)

If $f, g \in L^1(X)$, then TFAE:

a. $\int_E f = \int_E g \quad \forall E \in \mathcal{M}$

b. $\|f-g\|_1 = \int |f-g| = 0$.

c. $f = g$ a.e.

<u>Proof:</u>

$a \Rightarrow c$. Set

$$u = \operatorname{Re}(f-g), \qquad v = \operatorname{Im}(f-g).$$

Suppose that $E = \{u^+ \neq 0\}$ has positive measure.

Then since $u^-(x) = 0$ whenever $u^+(x) \neq 0$,

$$\int_E \operatorname{Re}(f-g) = \int_E u^+ - \cancelto{0}{\int_E u^-} > 0,$$

contradicting the fact that $\int_E (f-g) = 0$.

Thus $\{u^+ \neq 0\}$ has measure zero, & similarly so do $\{u^- \neq 0\}$, $\{v^+ \neq 0\}$, and $\{v^- \neq 0\}$. Hence

$f - g = 0$ a.e. ■

$c \Rightarrow b$. If $f = g$ a.e. then $|f-g| = 0$ a.e., so

$\int |f-g| = 0$.

$b \Rightarrow a$. If $\int |f-g| = 0$ & $E \in \mathcal{M}$, then

$$0 \le \left| \int_E f - \int_E g \right| \le \int_E |f-g| \le \int_X |f-g| = 0,$$

so $\int_E = \int_E g$. $\blacksquare$

Exercise: Reverse Triangle Inequality

$$\big| \|f\|_1 - \|g\|_1 \big| \le \|f-g\|_1$$

(True for any norm.)

Definition: Convergence (same as Lebesgue measure)

If $f_n, f \in L^1(X)$, then we say that fn converges to f in L^1-norm if

$$\lim_{n\to\infty} \|f - f_n\|_1 = \lim_{n\to\infty} \int |f - f_n| = 0.$$

In this case, we write

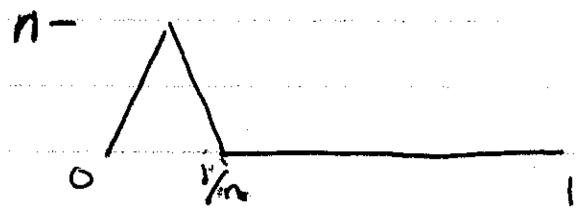
$$f_n \longrightarrow f \quad \text{in } L^1(X).$$

__Example__

a. Show that $\|f-f_n\|_1 \to 0 \implies \|f_n\|_1 \to \|f\|_1$

b. Show that $\|f_n\|_1 \to \|f\|_1 \not\Rightarrow \|f-f_n\|_1 \to 0$ in general

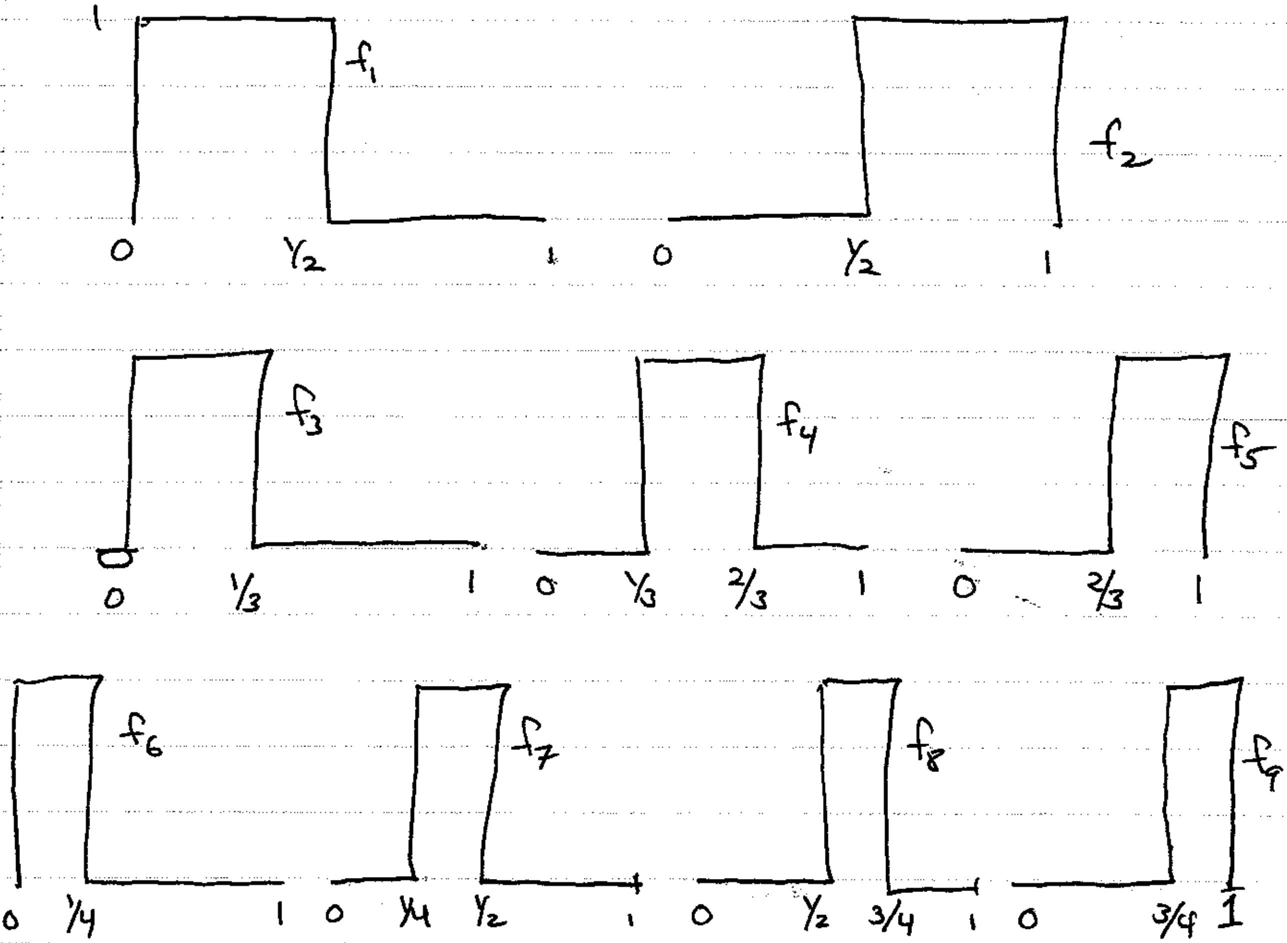
c. Show that $f_n \longrightarrow f$ pointwise $\not\Rightarrow f_n \to f$ in L^1.

Hint: 

d. Show $f_n \longrightarrow f$ in $L^1 \not\Rightarrow f_n \to f$ pointwise.

Hint: Consider the "Marching Boxes" example (see next page).

Marching Boxes



etc.

e. Find a <u>subsequence</u> $\{f_{n_k}\}_{k\in\mathbb{N}}$ of the marching boxes that <u>does</u> converge pointwise a.e.

<u>Remark</u>
We'll see later that

$f_n \to f$ in L^1 $\Longrightarrow$ $\exists$ subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ that converges pointwise a.e. to f

Some additional remarks on $L^1(X)$

For a complete measure, we know that if f is Ⓜ & $g = f$ a.e., then g is Ⓜ. Thus if $f \in L^1(X)$, then every representative of the equivalence class of f is Ⓜ. What if μ is not complete? Fortunately, not much goes wrong if we replace μ by its completion $\bar{\mu}$ (defined in HW 2).

Theorem (New)

Let $(X, \mathcal{M}, \mu)$ be a measure space, and let $(X, \overline{\mathcal{M}}, \bar{\mu})$ be its completion. If f is an extended real-valued or complex-valued $\overline{\mathcal{M}}$-measurable function on X, then $\exists$ an $\mathcal{M}$-measurable g on X s.t. $f = g$ $\bar{\mu}$-a.e.

Proof

Suppose that $A \in \overline{\mathcal{M}}$. Then, by definition,

$A = E \cup F$ where $E \in \mathcal{M}$ and $F \subseteq N$ for some μ-null set N. Hence $\chi_A = \chi_E$ $\bar{\mu}$-a.e., and χ_E is $\mathcal{M}$-measurable.

Now let f be an arbitrary $\overline{\mathcal{M}}$-m function.

Exercise: Although we only did it for nonnegative functions, so $\exists$ simple functions ϕ_k that are $\overline{\mathcal{M}}$-m and converge pointwise to f.

Since each ϕ_k is a ~~finear~~ linear combination of characteristic functions, write

$\phi_k = \sum_{j=1}^{N} a_j \chi_{A_j}$ with A_j $\overline{\mathcal{M}}$-m. Then $\exists$ $\mathcal{M}$-m E_j s.t. $\chi_{A_j} = \chi_{E_j}$ $\bar{\mu}$-a.e. Hence, if we set $\psi_k = \sum_{j=1}^{N} a_j \chi_{E_j}$ then ψ_k is $\mathcal{M}$-m and $\psi_k = \phi_k$ $\bar{\mu}$-a.e. In particular, if

$$E_k = \{\psi_k \neq \phi_k\}$$

Then $\bar{\mu}(E_k) = 0$.

Therefore $E = \bigcup E_k$ also satisfies

$\bar{\mu}(E) = 0$. By definition,

$$E = H \cup F$$

where $H \in \mathcal{M}$ and $F \subseteq N$ for some μ-null set N. Since $\mu(H) = \bar{\mu}(E) = 0$, we have that

$$M = H \cup N$$

is an $\mathcal{M}$-m null set that contains E. Set

$$f_k = \varphi_k \cdot \chi_{M^c}.$$

Each f_k is $\mathcal{M}$-m, so

$$g(x) = \lim_{k\to\infty} f_k(x) \quad \text{(exists } \forall x \text{, why?)}$$

is $\mathcal{M}$-m as well. Finally, $g = f$ $\bar{\mu}$-a.e. (why?)

$\blacksquare$

<u>Remark.</u>

In the converse direction, note that

$$f \text{ is } \mathcal{M}\text{-m} \Rightarrow f \text{ is } \overline{\mathcal{M}}\text{-m}.$$

Convention

As a consequence of these results, whenever dealing with $L^1(X)$, we assume μ is complete (or replace μ by its completion).

Now we come to the "workhorse" of the convergence theorems!

Not the discrete cosine transform!!

Dominated Convergence Theorem (DCT)

Suppose that $f_n \in L^1(X)$ are such that

a. $f_n(x) \longrightarrow f(x)$ pointwise a.e.

b. $\exists g \in L^1(X)$ s.t. $|f_n| \leq g$ a.e. $\forall n \in \mathbb{N}$

Then $f \in L^1(X)$ and

$$\lim_{n\to\infty} \int f_n = \int f.$$

Proof:

If μ is complete, as in our convention, then f is (m) (and even if μ is not complete, if we redefine f on a null set then it will be m).

Further,

$$|f(x)| = \lim_{n\to\infty} |f_n(x)| \le g(x) \text{ a.e.},$$

so

$$\int |f| \le \int g < \infty.$$

Thus $f \in L^1(X)$.

Suppose that f is extended-real-valued. Then

$$g \pm f_n \geq 0 \text{ a.e.}$$

Hence, by Fatou's Lemma,

$$\int g + \int f = \int (g+f)$$

$$= \int \liminf_{n\to\infty} (g+f_n)$$

$$\le \liminf_{n\to\infty} \int (g+f_n) \quad \text{Fatou.}$$

constant!

$$= \liminf_{n\to\infty} \left(\int g + \int f_n \right)$$

$$= \int g + \liminf_{n\to\infty} \int f_n .$$

Consequently,

$$\int f \leq \liminf_{n\to\infty} \int f_n .$$

Similarly,

$$\int g - \int f = \int (g-f) \qquad \text{note all are } \underline{\text{finite}}$$

$$= \int \liminf_{n\to\infty} (g - f_n)$$

$$\leq \liminf_{n\to\infty} \int (g - f_n) \qquad \text{Fatou}$$

$$= \liminf_{n\to\infty} \left(\int g - \int f_n \right)$$

$$= \int g + \liminf_{n\to\infty} \left(-\int f_n \right)$$

$$= \int g - \limsup_{n\to\infty} \int f_n$$

Hence

$$\limsup_{n\to\infty} \int f_n \leq \int f.$$

Combining all the inequalities, we see that

$$\liminf_{n\to\infty} \int f_n = \int f = \limsup_{n\to\infty} \int f_n. \quad \blacksquare$$

Exercise: An extension.

Show that with the same hypotheses we actually get the better conclusion that $f_n \to f$ in L^1-norm, i.e.,

$$\lim_{n\to\infty} \|f - f_n\|_1 = \lim_{n\to\infty} \int |f - f_n| = 0.$$

(refer to earlier exercises to see why this is better).

Hint: Apply the DCT to $|f - f_n|$.

Exercise: Another extension.

This is the Generalized DCT. Suppose $f_n \in L^1(X)$ and

a. $f_n(x) \longrightarrow f(x)$ pointwise a.e.

b. $\exists\, g_n \in L^1(X)$ s.t. $|f_n| \leq g_n$ a.e.

c. $\exists\, g \in L^1(X)$ s.t. $g_n(x) \longrightarrow g(x)$ pointwise a.e.

Then $\lim_{n\to\infty} \int f_n = \int f.$

Hint: Closely follow the proof of the DCT.

As a corollary of the DCT, we get a result about switching integrals and infinite series. Recall that we had a similar theorem before, a special case of Tonelli's Theorem, but that only applied to nonnegative functions. The following result, which is a special case of Fubini's Theorem, applies to general functions.

Corollary

If $f_n \in L^1(X)$ and

$$\sum_{n=1}^{\infty} \|f_n\|_1 < \infty$$

(in which case we say that $\sum f_n$ converges absolutely in L^1-norm)

then

a. $f(x) = \sum_{n=1}^{\infty} f_n(x)$ converges for a.e. x,

b. $f \in L^1(X)$,

c. $\int \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int f_n$.

Proof:

Set
$$g_N(x) = \sum_{n=1}^{N} f_n(x) \quad \& \quad g(x) = \sum_{n=1}^{\infty} |f_n(x)|.$$

Converges in $[0,\infty]$

By Tonelli's Theorem, since all the terms are nonnegative,

$$\int g = \int \sum_{n=1}^{\infty} |f_n|$$
$$= \sum_{n=1}^{\infty} \int |f_n| \qquad \text{Tonelli}$$
$$= \sum_{n=1}^{\infty} \|f_n\|_1$$
$$< \infty$$

Therefore $g \in L^1(X)$ since $g \geq 0$.

Consequently, g must be finite a.e., i.e.,

$$g(x) = \sum_{n=1}^{\infty} |f_n(x)| < \infty \quad \text{except for } x \in Z$$
with $\mu(Z) = 0$.

Thus, for $x \notin Z$, as a series of <u>scalars</u>, the series

$$f(x) \stackrel{\text{def}}{=} \sum_{n=1}^{\infty} f_n(x) = \lim_{N\to\infty} g_N(x)$$

exists & is finite. That is, $g_N(x) \to f(x)$ pointwise a.e.

Since we also have

$$|g_N(x)| \le \sum_{n=1}^{N} |f_n(x)| \le \sum_{n=1}^{\infty} |f_n(x)|$$
$$= g(x) \in L^1(X)$$

we can apply the DCT to g_N to obtain that $f \in L^1(X)$ and

$$\int f = \lim_{N\to\infty} \int g_N$$

$$= \lim_{N\to\infty} \int \sum_{n=1}^{N} f_n$$

$$= \lim_{N\to\infty} \sum_{n=1}^{N} \int f_n \quad \text{linearity of the integral}$$

$$= \sum_{n=1}^{\infty} \int f_n \quad \text{definition of infinite series.}$$

Application: Switching Derivatives and Integrals

Since a derivative is a limit, we might expect that, with suitable hypotheses, we can interchange an integral & a derivative (this is known as "differentiating under the integral sign").

The following result is typical. We will omit the proof, which can be found in Folland's text.

Theorem

Let $f: X \times [a,b] \to \mathbb{C}$ be given. Suppose that

a. $f_t(x) = f(x,t) \in L^1(X) \quad \forall t \in [a,b]$

b. $\frac{\partial f}{\partial t}$ exists

c. $\exists\, g \in L^1(X)$ with $\left|\frac{\partial f}{\partial t}(x,t)\right| \leq g(x) \quad \forall x, t$

Then

$$F(t) = \int f_t(x)\, d\mu(x), \qquad t \in [a,b]$$

is differentiable, and

$$F'(t) = \int \frac{\partial f}{\partial t}(x,t)\, d\mu(x).$$

Relation between Riemann & Lebesgue integrals

We sketched this relation before, assuming continuity of f. Let us give this a little more precisely now, without assuming continuity.

Definition (Darboux's definition of the Riemann integral).

Let $f: [a,b] \longrightarrow \mathbb{R}$ be bounded. Given a partition

$\Gamma = \{a = x_0 < x_1 < \cdots < x_n = b\}$ of $[a,b]$,

define

$$S_\Gamma f = \sum_{j=1}^{n} M_j (x_j - x_{j-1}) \quad \text{upper Riemann sum}$$

$$s_\Gamma f = \sum_{j=1}^{n} m_j (x_j - x_{j-1}) \quad \text{lower Riemann sum}$$

Where

$$m_j = \inf_{[x_{j-1}, x_j]} f \quad \& \quad M_j = \sup_{[x_{j-1}, x_j]} f.$$

Set

$$\overline{I}_a^b(f) = \inf\{S_\Gamma f : \text{all partitions } \Gamma\}$$

$$\underline{I}_a^b(f) = \sup\{s_\Gamma f : \text{all partitions } \Gamma\}$$

Then the <u>Riemann Integral</u> of f exists if

$$\underline{I}_a^b(f) = \overline{I}_a^b(f),$$

and in this case we denote this number by

$$(R) \int_a^b f(x)\,dx$$

<u>Theorem</u>

Let $f: [a,b] \longrightarrow \mathbb{R}$ be bounded.

a. If f is Riemann integrable, then it is Lebesgue (m) and integrable, and

$$(R) \int_a^b f(x)\,dx = \int_{[a,b]} f(x)\,dx.$$

b. f is Riemann integrable $\iff$ f is continuous a.e.

<u>Note</u>

Continuous a.e. means the set of discontinuities has measure zero. It does <u>not</u> mean that $\exists$ continuous function that equals f a.e.

Proof:

a. Assume f is Riemann integrable (and note we are not assuming f is continuous). Define simple functions

$$g_\Gamma = \sum_{j=1}^{n} m_j \chi_{[x_{j-1}, x_j]}, \quad G_\Gamma = \sum_{j=1}^{n} M_j \chi_{(x_{j-1}, x_j]}$$

Corresponding to a partition Γ. Then

$$s_\Gamma f = \int g_\Gamma \quad \& \quad S_\Gamma f = \int G_\Gamma \quad \text{(Lebesgue integrals!)}$$

Note that if Γ' refines Γ (includes all the points in Γ plus more), then

$$g_\Gamma \le g_{\Gamma'} \le G_{\Gamma'} \le G_\Gamma .$$

Now let

$$I = (R)\int_a^b f(x)\,dx = \sup\left\{ \int g_\Gamma : \text{all partitions } \Gamma \right\}$$

Then there must exist a partition Γ_1 s.t.

$$I - 1 \le \int g_{\Gamma_1} \le I$$

Also, $\exists$ partition Γ_1' s.t.

$$I - \tfrac{1}{2} \le \int g_{\Gamma_1'} \le I$$

Set

$$\Gamma_2 = \Gamma_1 \cup \Gamma_1'.$$

Then Γ_2 is a refinement of both Γ_1 & Γ_1'. In particular,

$$g_{\Gamma_1'} \leq g_{\Gamma_2},$$

so

$$I - \tfrac{1}{2} \leq \int g_{\Gamma_1'} \leq \int g_{\Gamma_2} \leq I.$$

Continuing in this way, we can obtain refinements

$$\Gamma_1 \subseteq \Gamma_2 \subseteq \cdots$$

such that

$$g_{\Gamma_1} \leq g_{\Gamma_2} \leq \cdots$$

and

$$\int g_{\Gamma_k} \longrightarrow I.$$

Let

$$g = \lim_{k \to \infty} g_{\Gamma_k}$$

This converges since the g_{Γ_k} are increasing, and furthermore each $g_{\Gamma_k} \leq f$, which is a bounded function. Since we are on a finite domain & f is bounded, it is integrable. Also

$$|g_{\Gamma_k}| \leq |f| \in L^1[a,b] \quad \forall k.$$

Therefore, by the DCT,

$$\int g_{\Gamma_k} \longrightarrow \int g \text{ (Lebesgue integral)}$$

But also we know that $\int g_{\Gamma_k} \to I$,

so

$$\int g = I.$$

Note that, by construction, $g \leq f$, and g is (m) since it is a pointwise limit of the simple functions g_{Γ_k}.

Operating similarly from above, we can create a (m) function $G \geq f$ that satisfies

$$\int G = I.$$

Hence $G-g \geq 0$ & is (m), and

$$\int (G-g) = \int G - \int g = I - I = 0.$$

Since $G-g$ is nonnegative, we conclude that

$G-g = 0$ a.e. And since $g \leq f \leq G$,

$$f = g = G \text{ a.e.}$$

Therefore f is (m) since Lebesgue measure is complete,

and

$$\underbrace{\int f}_{\text{Lebesgue integral}} = \int g = I = \mathcal{R}\int f.$$

b. $\Longrightarrow$ Suppose that f is Riemann integrable.

Using all the same notation as in part a, set

$$E = \{ x \in X : f(x) = g(x) = G(x) \}.$$

Since $f = g = G$ a.e., $Z = E^C$ has measure zero.

Also, since each partition Γ_k contains only finitely many points,

$$\Gamma = \bigcup_k \Gamma_k$$

is a countable set.

Suppose that $x \notin \Gamma$ and f is discontinuous at x.

Exercise: Show $\exists\, \varepsilon > 0$ s.t.

$$G_{\Gamma_k}(x) - g_{\Gamma_k}(x) \geq \varepsilon \quad \forall\, k \in \mathbb{N}.$$

Letting $k \to \infty$, we see that

$$G(x) - g(x) \geq \varepsilon > 0,$$

so $x \notin E$, and therefore $x \in Z$. Hence f is continuous for all $x \notin \Gamma \cup Z$. This is a set of measure zero, so f is continuous a.e.

$\Leftarrow$ Suppose now that f is continuous a.e.

Choose <u>any</u> sequence of partitions Γ_k such that $|\Gamma_k| \to 0$. As in the proof of Problem #4 on HW 4,

$g_{\Gamma_k}(x),\ G_{\Gamma_k}(x) \longrightarrow f(x)$ at all points of continuity of f. By their definition, if we set

$$M = \sup_{x \in [a,b]} |f(x)|,$$

then

$$|g_{\Gamma_k}(x)|,\ |G_{\Gamma_k}(x)| \leq M \quad \forall x \in [a,b].$$

Since $[a,b]$ is a finite interval, the constant function M is integrable, i.e., $M \in L^1[a,b]$. The DCT therefore implies that

$$s_{\Gamma_k} = \int g_{\Gamma_k} \longrightarrow \int f$$

$$S_{\Gamma_k} = \int G_{\Gamma_k} \longrightarrow \int f$$

Therefore

$$\underline{I}_a^b(f) = \sup_{\Gamma} \{s_\Gamma\} \geq \int f$$

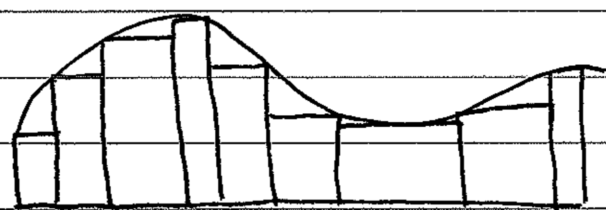
$$\overline{I}_a^b(b) = \inf_{\Gamma} \{S_\Gamma\} \leq \int f$$

The opposite inequalities follow immediately from the definition, so f is Riemann integrable, &

$$\int f = \underline{I}_a^b(f) = \overline{I}_a^b(f) = (R)\int f. \blacksquare$$

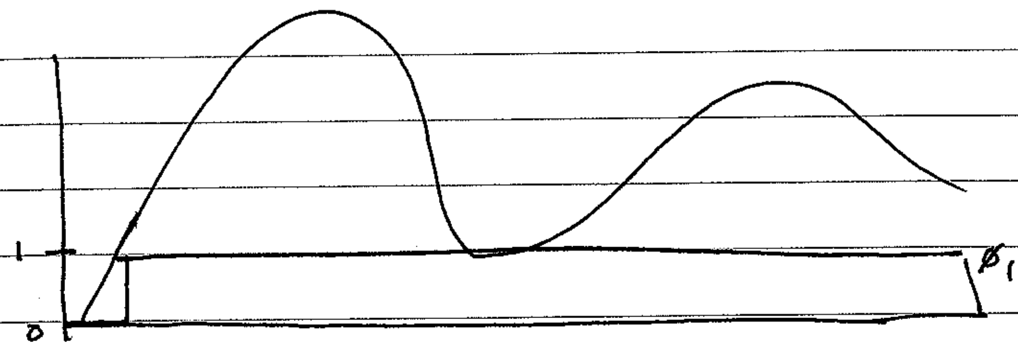
Comparison of Riemann & Lebesgue approaches

Riemann works by subdividing the x-axis:

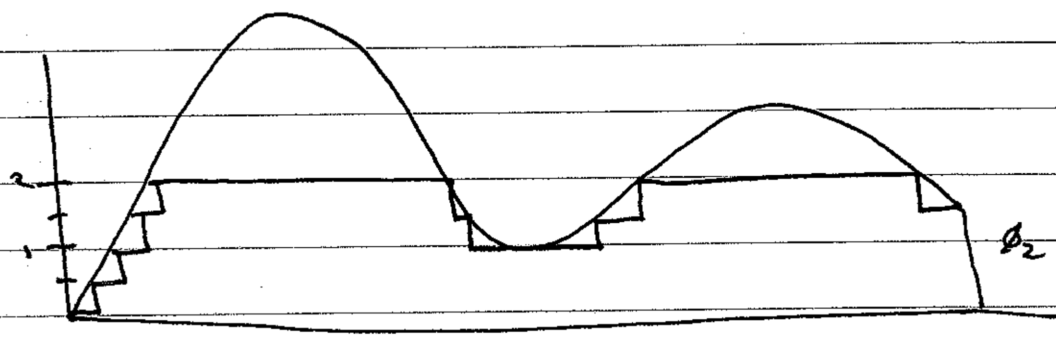


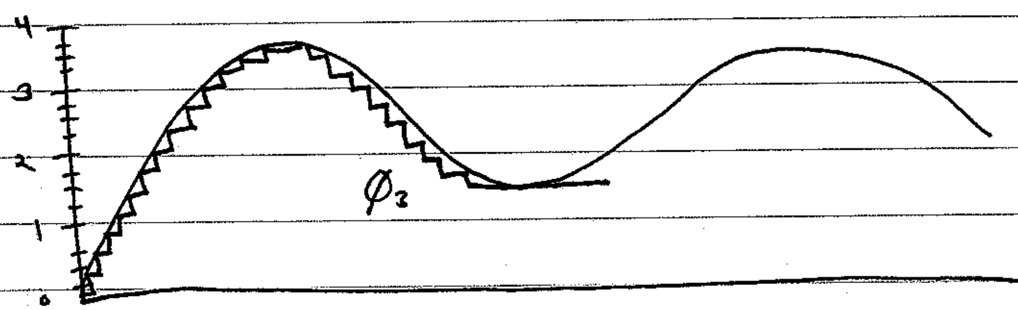
refine the partition

Lebesgue subdivides the y-axis (see picture next page — keep in mind that although it looks similar, that's because the illustration shows a continuous function).



Step 1





Disadvantages: More "complicated" def

Advantages: Wide class of fns are integrable,

Convergence Thms hold

Some Exercises

1. Prove the Bounded Convergence Theorem

Suppose μ is finite measure, and suppose

a. f_n, f are m functions on X (either extended-real or complex-valued),

b. $f_n(x) \longrightarrow f(x)$ a.e.

c. $\exists M$ s.t. $|f_n(x)| \leq M$ a.e. $\forall n \in \mathbb{N}$

Then $\lim\limits_{n\to\infty} \int f_n = \int f$.

2. Prove the Uniform Convergence Theorem:

Let μ be a finite measure, and suppose

a. f_n, f are m functions on X (either extended-real or complex valued)

b. $f_n \longrightarrow f$ uniformly on X.

Then $\lim\limits_{n\to\infty} \int f_n = \int f$.

3. Suppose $f: \mathbb{R}^d \longrightarrow \mathbb{R}$ is Lebesgue m. Show that if $T: \mathbb{R}^d \longrightarrow \mathbb{R}^d$ is a [redacted] nonsingular linear transformation, then $f \circ T$ is m.

Hint: Show $\{ f \circ T > a \} = T^{-1}(\{f > a\})$. Then show that any Lipschitz function (such as T^{-1}) maps m sets to m sets. (Recall we showed on Exam 1 that a Lipschitz function maps zero measure sets to zero measure sets. It is also true that a continuous function maps compact sets to compact sets. An arbitrary m function is the union of an F_σ set and a set of zero measure.)

4. Let $f_n, f: X \to [0,\infty]$ be m, with $f_n \leq f$ a.e. Show that if $f_n(x) \to f(x)$ for a.e. x, then $\int f_n \longrightarrow \int f$ (might be infinite).

5. Show that if $f_n \to f$ in $L^1(X)$, then $f_n \xrightarrow{m} f$ (see HW 4 for the definition of convergence in measure).

6. Let $f_n \in L^1(X)$ be given. Show that

$\sum_{k=1}^{\infty} f_k$ converges absolutely in $L^1(X)$, i.e., $\sum_{k=1}^{\infty} \|f_k\|_1 < \infty$,

implies $\sum_{k=1}^{\infty} f_k(x)$ converges absolutely for a.e. x,

i.e., $\sum_{k=1}^{\infty} |f_k(x)| < \infty$ a.e.

Definition

The essential supremum of an extended real-valued function $f: X \to \overline{\mathbb{R}}$ is the smallest constant that dominates f almost everywhere:

$$\operatorname*{ess\,sup}_{x \in X} f(x) = \inf\{M : f \le M \text{ a.e.}\}$$

The L^∞-norm of $f: X \to \overline{\mathbb{R}}$ or $f: X \to \mathbb{C}$ is

$$\|f\|_\infty = \operatorname*{ess\,sup}_{x \in X} |f(x)|.$$

7. Show that if μ is a finite measure, then

$$\|f\|_1 \leq \mu(X) \cdot \|f\|_\infty$$

8. Show that $\nexists\, C > 0$ s.t.

$$\forall f \in L^1(\mathbb{R}), \quad \|f\|_1 \leq C\, \|f\|_\infty .$$

9. Show that if $f : [a,b] \longrightarrow \mathbb{R}$ is continuous, then

$$[redacted] \quad \operatorname*{ess\,sup}_{x \in [a,b]} f(x) = \sup_{x \in [a,b]} f(x)$$

10. Let

$$L^\infty(X) = \left\{ f : \underset{m}{X} \to \mathbb{C} \;:\; \|f\|_\infty < \infty \right\}$$

Show that $\|\cdot\|_\infty$ is a seminorm on $L^\infty(X)$, & is a norm if we identify functions that are equal a.e.

11. Show that if μ is a finite measure, then $L^\infty(X) \subseteq L^1(X)$.

Show that $L^\infty(\mathbb{R}) \not\subseteq L^1(\mathbb{R})$.

2.5 Product Measures

We want to develop measures & integration on $X \times Y$.

Suppose that $(X, \mathcal{M}, \mu)$ & $(Y, \mathcal{N}, \nu)$ are measure spaces.

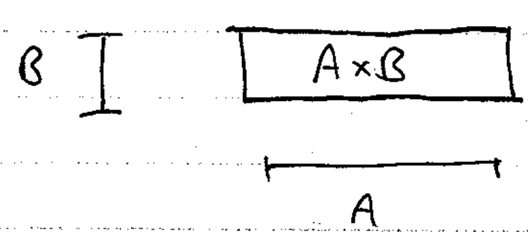
We want to construct an appropriate measure space $(X \times Y, \mathcal{M} \otimes \mathcal{N}, \mu \times \nu)$.

What sets will be m & what will their measures be? We'll sketch the construction.

Start with what we certainly want to be m:

If $A \in \mathcal{M}$, $B \in \mathcal{N}$, then we certainly want $A \times B$ to be m. So $\mathcal{M} \otimes \mathcal{N}$ should include all such sets.

<u>Notation</u>
We call $A \times B$ a <u>rectangle</u>, even though this is clearly an abuse of terminology.



We declare that

$$(\mu \times \nu)(A \times B) = \mu(A)\,\nu(B).$$

So now we know how to measure rectangles, and we have to figure out how to extend to arbitrary m sets.

If we set

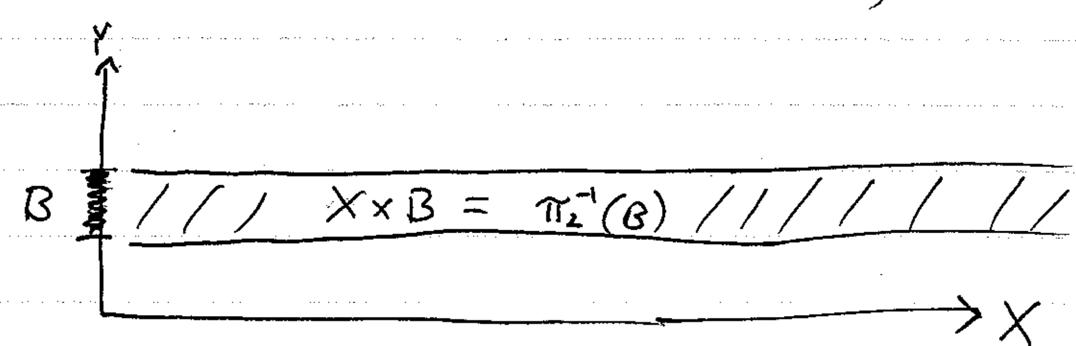
$$\mathcal{E} = \{A \times B : A \in \mathcal{M},\ B \in \mathcal{N}\},$$

then we define

$$\mathcal{M} \otimes \mathcal{N} = \mathcal{M}(\mathcal{E}),$$ the σ-algebra generated by $\mathcal{E}$.

<u>Exercise</u>
We don't even need all the rectangles. Show that $\mathcal{M} \otimes \mathcal{N}$ is generated by

$$\{A \times Y : A \in \mathcal{M}\} \cup \{X \times B : B \in \mathcal{N}\}$$



Now we construct the measure $\mu \times \nu$ by going through a premeasure construction.

First we form an algebra by taking all finite unions of disjoint ~~algebras~~ rectangles

$$\mathcal{A} = \left\{ \bigcup_{j=1}^{n} A_j \times B_j : A_j \in \mathcal{M}, B_j \in \mathcal{N}, \quad A_j \times B_j \text{ disjoint} \right\}$$

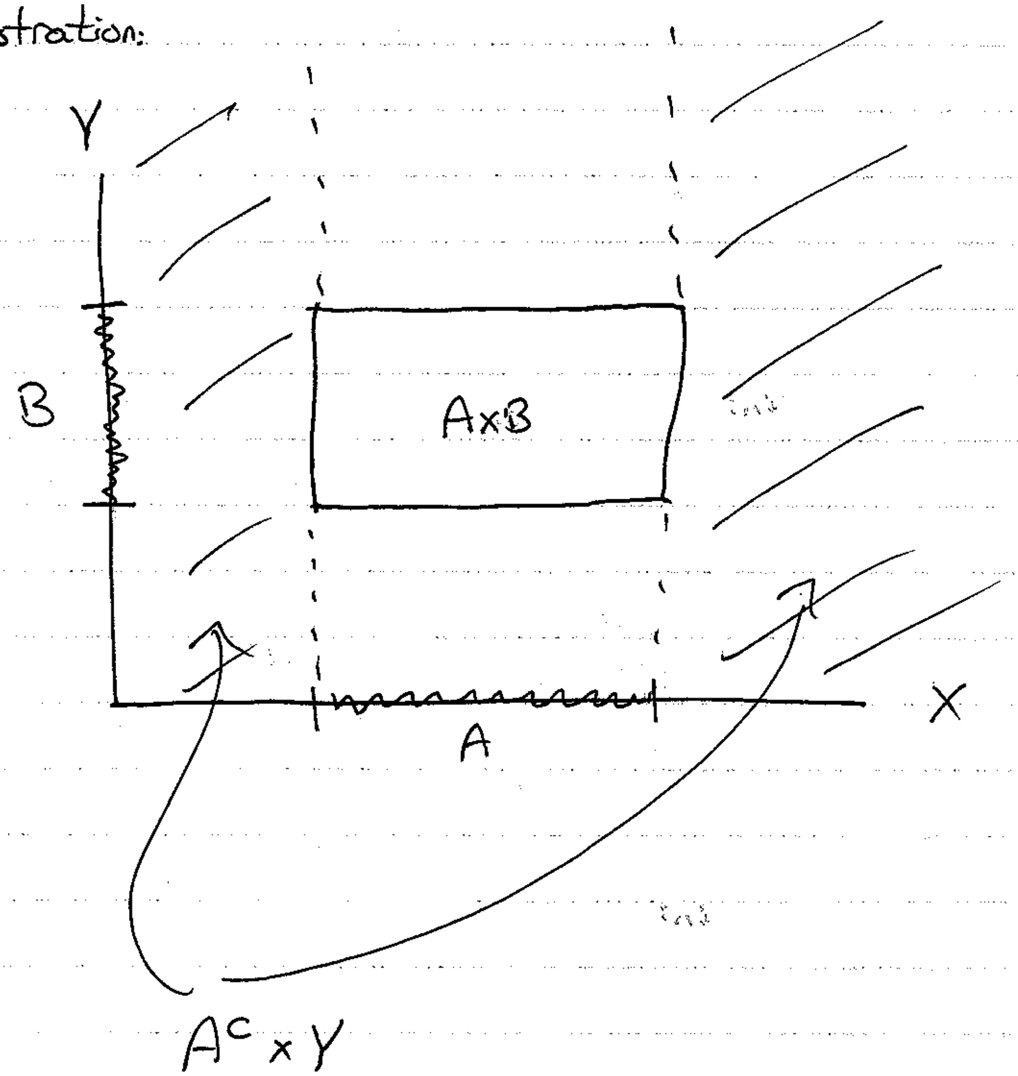
To show this is an algebra, just show that $\mathcal{E}$ is an elementary family, ie.

a. $\emptyset \in \mathcal{E}$

b. $E, F \in \mathcal{E} \implies E \cap F \in \mathcal{E}$

c. $E \in \mathcal{E} \implies E^c = \bigcup_{j=1}^{n} E_j, \quad E_j \in \mathcal{M}$ disjoint

Illustration:



$$(A \times B)^c = (A^c \times Y) \cup (A \times B^c) \text{ disjointly.}$$

If $E = \bigcup_{j=1}^{n} A_j \times B_j \in \mathcal{A}$ (disjoint sum of rectangles), then we set

$$\rho(E) = \sum_{j=1}^{n} \mu(A_j)\nu(B_j)$$

This defines a premeasure. We then have an outer measure:

$$\rho^*(E) = \inf\left\{\sum_k \rho(E_k) : E_k \in \mathcal{A},\ E \subseteq \bigcup E_k\right\}$$

Since we've gone through the premeasure process, every set in $\mathcal{M}\otimes\mathcal{N}$ is ρ^*-(m), and

$$\rho^*(A\times B) = \mu(A)\nu(B).$$

We then define

$$\mu\times\nu = \rho^*|_{\mathcal{M}\otimes\mathcal{N}},$$

this is a measure with σ-algebra $\mathcal{M}\otimes\mathcal{N}$.

Good news: $(\mu\times\nu)(A\times B) = \mu(A)\nu(B)$
$\forall A\in\mathcal{M},\ B\in\mathcal{N}.$

In fact, $\mu \times \nu$ is the unique measure on $\mathcal{M} \otimes \mathcal{N}$ for which $(\mu \times \nu)(A \times B) = \mu(A)\nu(B)$ holds.

By induction, we can extend to higher-fold products, & show that associativity holds, e.g.,

$$\mathcal{M}_1 \otimes (\mathcal{M}_2 \otimes \mathcal{M}_3) = (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3$$

$$\mu_1 \times (\mu_2 \times \mu_3) = (\mu_1 \times \mu_2) \times \mu_3$$

We'll take the construction of $\mu \times \nu$ as given from now on.

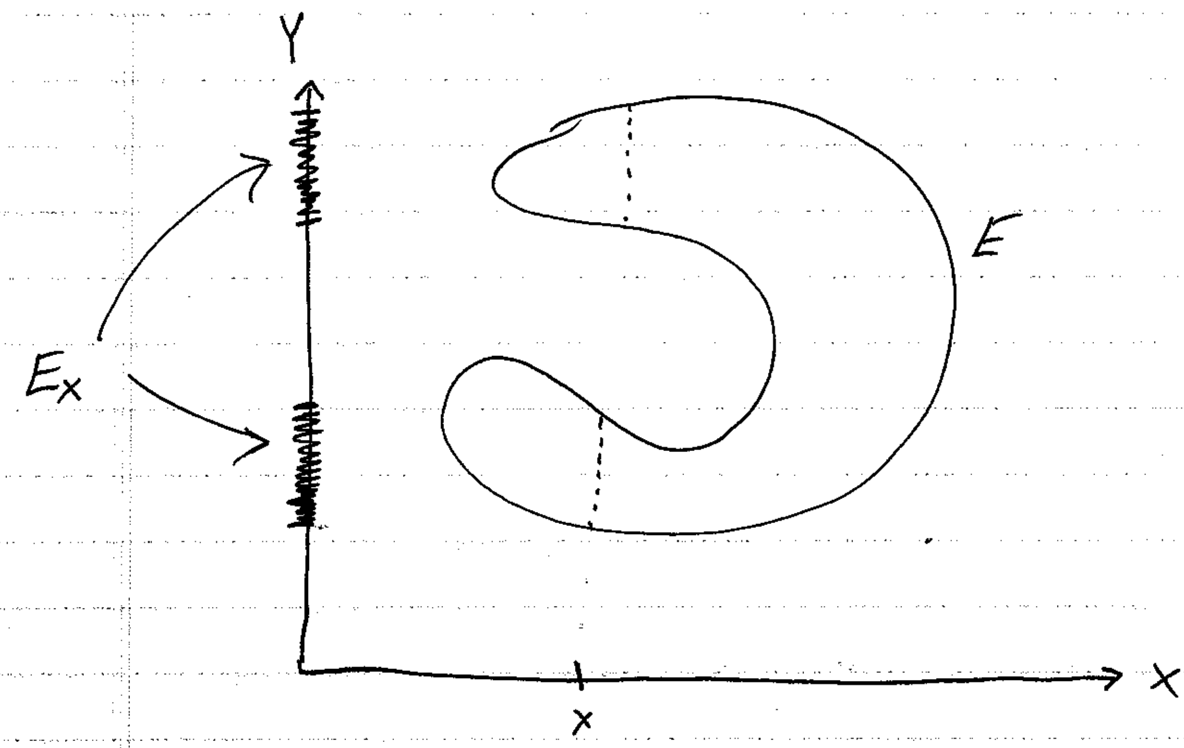
Now we proceed towards theorems involving repeated integration.

Sections of a Set

Given $E \subseteq X \times Y$, $x \in X$, and $y \in Y$,

the x-section of E is $E_x = \{y \in Y : (x,y) \in E\}$

the y-section of E is $E^y = \{x \in X : (x,y) \in E\}$



Note that $E_x \subseteq Y$ & $E^y \subseteq X$.

Exercises

a. $(E^c)_x = (E_x)^c$, $\quad (E^c)^y = (E^y)^c$

b. $\left(\bigcup E_j\right)_x = \bigcup (E_j)_x$, $\quad \left(\bigcup E_j\right)^y = \bigcup (E_j)^y$

c.
$$(A \times B)_x = \begin{cases} B, & x \in A \\ \emptyset, & x \notin A \end{cases}$$

$$(A \times B)^y = \begin{cases} A, & y \in B \\ \emptyset, & y \notin B \end{cases}$$

Now we show that sections of a m set are m.

Theorem
If $E \in \mathcal{M} \otimes \mathcal{N}$ then $E_x \in \mathcal{N}$ & $E^y \in \mathcal{M}$ $\forall x \in X, y \in Y$.

Proof:
Let
$$\mathcal{R} = \{E \subseteq X \times Y : E_x \in \mathcal{N} \ \& \ E^y \in \mathcal{M} \ \ \forall x \in X, y \in Y\}$$

Exercise: Use the exercise above to show that $\mathcal{R}$ is a σ-algebra, and that $A \times B \in \mathcal{R}$ whenever $A \in \mathcal{M}$, $B \in \mathcal{N}$. Hence $\mathcal{R}$ contains all rectangles and therefore contains the σ-algebra generated by the rectangles, which is $\mathcal{M} \otimes \mathcal{N}$.

That is, $\mathcal{M} \otimes \mathcal{N} \subseteq \mathcal{R}$, which proves the theorem. $\blacksquare$

<u>Sections of a function</u>

Given $f : X \times Y \to Z$, $x \in X$, $y \in Y$:

the x-section of f is

$$f_x : Y \to Z \qquad \text{i.e., } f_x(y) = f(x,y)$$
$$y \mapsto f(x,y)$$

the y-section of f is

$$f^y : X \longrightarrow Z \qquad \text{i.e., } f^y(x) = f(x,y)$$
$$x \longmapsto f(x,y)$$

Note that

$$(\chi_E)_x = \chi_{E_x} \quad \& \quad (\chi_E)^y = \chi_{E^y}.$$

Exercise

Show that if f is $\mathcal{M}\otimes\mathcal{N}$-(m), then

f_x is $\mathcal{N}$-(m) $\forall x \in X$ and

f^y is $\mathcal{M}$-(m) $\forall y \in Y$.

Motivation: Characteristic Functions

Suppose for the moment that we could switch integrals at will. Then we would have

$$\begin{aligned}(\mu\times\nu)(E) &= \iint \chi_E(x,y)\, d(\mu\times\nu)(x,y) \\ &= \iint \chi_{E_x}(y)\, d\mu(x)\, d\nu(y) \\ &= \int \left(\int \chi_{E_x}(y)\, d\nu(y) \right) d\mu(x) \\ &= \int \nu(E_x)\, d\mu(x).\end{aligned}$$

The next theorem will show directly that

$(\mu\times\nu)(E) = \int \nu(E_x)\, d\mu(x)$ is valid.

We begin with finite measures.

Theorem

Suppose $(X, \mathcal{M}, \mu)$ & $(Y, \mathcal{N}, \nu)$ are finite measures.

If $E \in \mathcal{M} \otimes \mathcal{N}$, then the following statements hold:

a. $x \longmapsto \nu(E_x)$ is $\mathcal{M}$-m

b. $y \longmapsto \mu(E^y)$ is $\mathcal{N}$-m

c. $(\mu \times \nu)(E) = \int \nu(E_x)\, d\mu(x) = \int \mu(E^y)\, d\nu(y)$

Proof:

Let $\mathcal{C}$ be the set of all $E \in \mathcal{M} \otimes \mathcal{N}$ for which a, b, & c hold.

Claim 1: $\mathcal{C}$ contains all measurable rectangles.

To see this, suppose $E = A \times B$ where $A \in \mathcal{M}$, $B \in \mathcal{N}$.

Then E_x is either B or $\emptyset$, which are ⓜ, & likewise E^y is ⓜ. Also,

$$\nu(E_x) = \chi_A(x) \cdot \nu(B)$$

so

$$\int \nu(E_x)\, d\mu(x) = \int \chi_A(x)\, \nu(B)\, d\mu(x)$$

$$= \mu(A)\, \nu(B)$$

$$= (\mu \times \nu)(E)$$

and similarly $\int \mu(E^y)\, d\nu(y) = (\mu \times \nu)(E)$.

Hence $E \in \mathcal{C}$.

<u>Exercise:</u> Extend this to show that $\mathcal{A} \subseteq \mathcal{C}$.

Claim 2: $\mathcal{C}$ is closed under increasing unions

Suppose $E_1 \subseteq E_2 \subseteq \cdots$ belong to $\mathcal{C}$, and set $E = \bigcup E_k$. Define

$$f_k(y) = \mu(E_k^y), \quad y \in Y.$$

Since $E_k \in \mathcal{C}$, we have that f_k is m. Since $E_1^y \subseteq E_2^y \subseteq \cdots$ and $\bigcup E_k^y = E^y$, we have by continuity from below that the f_k converge:

$$f(y) = \mu(E^y) = \lim_{k \to \infty} \mu(E_k^y) = \lim_{k \to \infty} f_k(y).$$

Hence f is m. Also,

$$\int \mu(E^y)\, d\nu(y) = \lim_{k \to \infty} \int \mu(E_k^y)\, d\nu(y) \quad \text{MCT}$$

$$= \lim_{k \to \infty} (\mu \times \nu)(E_k) \quad \text{since } E_k \in \mathcal{C}$$

$$= (\mu \times \nu)(E) \quad \text{by continuity from below.}$$

The remaining properties are symmetric, so $E \in \mathcal{C}$.

Claim 3: $\mathcal{C}$ is closed under decreasing intersections

Suppose $E_1 \supseteq E_2 \supseteq \cdots$ belong to $\mathcal{C}$, and set $E = \bigcap E_k$. As before, $f_k(y) = \mu(E_k^y)$ is ⓜ. Also $E_1^y \supseteq E_2^y \supseteq \cdots$ & $E^y = \bigcap E_k^y$. Since our measures are finite, we have by continuity from above that

$$f(y) = \mu(E^y) = \lim_{k\to\infty} \mu(E_k^y) = \lim_{k\to\infty} f_k(y)$$

converges, & therefore is ⓜ. Also, $f_k(y) \searrow f(y) \geq 0$ and all these functions are integrable, since

$$\begin{aligned}\int f_1(y)\, d\nu(y) &= \int \mu(E_1^y)\, d\nu(y)\\ &\leq \int \mu(X)\, d\nu(y)\\ &= \mu(X)\,\nu(Y) < \infty.\end{aligned}$$

Exercise: $f_k \searrow f \geq 0$ & $f_1 \in L^1(Y) \Rightarrow \int f_k \to \int f$.

Therefore we have

$$\int \mu(E^y)\, d\nu(y) = \lim_{k\to\infty} \int \mu(E_k^y)\, d\nu(y) \quad \text{by the exercise}$$

$$= \lim_{k\to\infty} (\mu\times\nu)(E_k^y) \quad \text{since } E_k \in \mathcal{C}$$

$$= (\mu\times\nu)(E) \quad \text{by continuity from above}$$

Thus $E \in \mathcal{C}$.

Now we appeal to the <u>Monotone Class Lemma</u>: because

i. $\mathcal{A}$ is an algebra & $\mathcal{A} \subseteq \mathcal{C}$,
ii. $\mathcal{C}$ is closed under increasing unions,
iii. $\mathcal{C}$ is closed under decreasing intersections,

it follows that $\mathcal{C}$ contains the σ-algebra generated by $\mathcal{A}$, which is $\mathcal{M}\otimes\mathcal{N}$. ∎

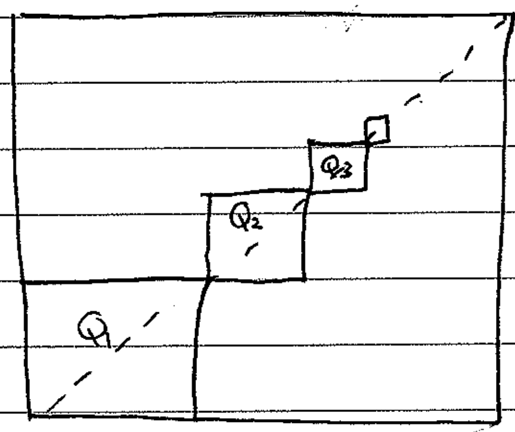
<u>Remark</u>

For a proof of the Monotone Class Lemma, see Folland's text. It actually says that if $\mathcal{S}$ is the <u>smallest</u> monotone class (ie, satisfying i, ii, iii) that contains $\mathcal{A}$, then $\mathcal{S}$ <u>is</u> the σ-algebra generated by $\mathcal{A}$. Hence we conclude that $\mathcal{C} \supseteq \mathcal{S} = \mathcal{M}\otimes\mathcal{N}$.

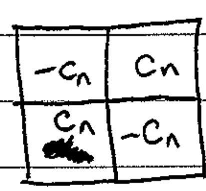
<u>Exercise</u>: Show that the preceding theorem remains valid if we only assume that μ, ν are σ-finite.

<u>Example</u> (Zygmund)

Let $X = Y = [0,1]$. Place countably many squares on the diagonal of the unit square:



Set $c_n = \frac{1}{|Q_n|}$. Define $f(x,y)$ to be zero outside $\bigcup Q_n$. On each Q_n let f have values as follows:



Q_n

Then f is ⓜ &

$$\int_0^1 f(x,y)\,dy = 0 = \int_0^1 f(x,y)\,dx.$$

Hence the following <u>iterated integrals</u> exist:

$$\int_0^1 \left(\int_0^1 f(x,y)\,dy \right) dx = 0 = \int_0^1 \left(\int_0^1 f(x,y)\,dx \right) dy .$$

However,

$$\iint_{[0,1]^2} f^+(x,y)\ (dx \times dy) = \sum_{k=1}^{\infty} \frac{1}{2} \frac{1}{|Q_n|} |Q_n| = \infty$$

and likewise $\iint f^- = \infty$. Hence the <u>double integral</u> $\iint f(x,y)\ (dx \times dy)$ <u>does not exist</u>.

Exercise

Show that

$$\int_1^\infty \left(\int_1^\infty \frac{x^2-y^2}{(x^2+y^2)^2}\, dy \right) dx = -\frac{\pi}{4}$$

while

$$\int_1^\infty \left(\int_1^\infty \frac{x^2-y^2}{(x^2+y^2)^2}\, dx \right) dy = \frac{\pi}{4}$$

Hint: Use the fact that $\frac{d}{dy} \frac{y}{x^2+y^2} = \frac{x^2-y^2}{(x^2+y^2)^2}$.

Thus, in general, the iterated integrals

$$\int \left(\int f(x,y)\, d\mu(x) \right) d\nu(y)$$

$$\int \left(\int f(x,y)\, d\nu(y) \right) d\mu(x)$$

and the double integral

$$\iint f(x,y)\, d(\mu\times\nu)(x,y)$$

need not be equal. Tonelli's & Fubini's Theorems will provide hypotheses under which these are all equal.

Tonelli's Theorem says that interchange is allowed for <u>nonnegative</u> functions (if the measure spaces are σ-finite).

<u>Tonelli's Theorem</u>

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces.

If $f: X \times Y \to [0, \infty]$ is (m), then the following statements hold.

We already know these

a. $f_x(y) = f(x,y)$ is (m) for $x \in X$.

b. $f^y(x) = f(x,y)$ is (m) for $y \in Y$.

c. $g(x) = \int f_x(y)\, d\nu(y)$ is (m)

d. $h(y) = \int f^y(x)\, d\mu(x)$ is (m)

e. As extended real numbers,

$$\iint f(x,y)\, d(\mu \times \nu)(x,y) = \int \left(\int f(x,y)\, d\mu(x) \right) d\nu(y)$$
$$= \int \left(\int f(x,y)\, d\nu(y) \right) d\mu(x).$$

<u>Proof:</u>

Exercise: Verify that our previous results establish Tonelli's Theorem for the case where f is a characteristic function, and extend this by linearity to simple functions.

Given an arbitrary m $f: X \times Y \to [0,\infty]$, let ϕ_n be simple functions $0 \le \phi_n \nearrow f$.

Then $g_n(x) = \int \phi_n(x,y)\, d\nu(y)$ is m, and by the Monotone Convergence Theorem, for each x,

$$g_n(x) = \int \phi_n(x,y)\, d\nu(y) \nearrow \int f(x,y)\, d\nu(y) = g(x),$$

so g is m, & similarly h is m. Also,

$$\iint f\, d(\mu\times\nu) = \lim_{n\to\infty} \iint \phi_n\, d(\mu\times\nu) \quad (\text{MCT})$$

$$= \lim_{n\to\infty} \int\left(\int \phi_n(x,y)\, d\nu(y)\right) d\mu(x)$$

$$= \lim_{n\to\infty} \int g_n\, d\mu$$

$$= \int g\, d\mu \quad (\text{MCT})$$

$$= \iint f(x,y)\, d\nu(y)\, d\mu(x)$$

and the equality of the other iterated integral is similar. ∎

The following corollary is extremely useful: to test whether $f \in L^1(X \times Y)$, we can check any <u>one</u> of three integrals for finiteness.

<u>Corollary</u>
Let $(X, \mathcal{M}, \mu)$ & $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces.

If $f : X \times Y \to \overline{\mathbb{R}}$ or $\mathbb{C}$ is ⓜ, then (as extended real numbers)

$$\iint |f| \, d(\mu \times \nu) = \iint |f(x,y)| \, d\mu(x) \, d\nu(y)$$
$$= \iint |f(x,y)| \, d\nu(y) \, d\mu(x)$$

Consequently, if any one of these integrals is finite, then $f \in L^1(X \times Y)$.

Fubini's Theorem allows the interchange of integrals if f is integrable (thereby again avoiding the ambiguity that is $\infty - \infty$).

<u>Fubini's Theorem</u>

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces.

If $f \in L^1(X \times Y)$, then the following statements hold.

a. $f_x \in L^1(Y)$ for μ-a.e. $x \in X$,

b. $f^y \in L^1(X)$ for ν-a.e. $y \in Y$

c. $g(x) = \int f_x(y)\, d\nu(y)$ belongs to $L^1(X)$

d. $h(y) = \int f^y(x)\, d\mu(x)$ belongs to $L^1(Y)$

e. $$\iint f(x,y)\, d(\mu \times \nu)(x,y) = \int \left(\int f(x,y)\, d\mu(x) \right) d\nu(y)$$

$$= \int \left(\int f(x,y)\, d\nu(y) \right) d\mu(x)$$

$\blacksquare$

Proof:

Suppose first that $f \geq 0$. Then by Tonelli's Theorem, f_x, f^y, g, & h are m, and

$$0 \leq \int g \, d\mu = \int h \, d\nu = \iint f \, d(\mu \times \nu) < \infty.$$

since $f \in L^1$

Hence $g \in L^1(X)$ & $h \in L^1(Y)$, since they are nonnegative. Consequently, g & h must be finite a.e., so

$$0 \leq g(x) = \int f_x \, d\nu < \infty \quad \text{a.e. } x$$

$$0 \leq h(y) = \int f^y \, d\mu < \infty \quad \text{a.e. } y$$

Therefore f_x is integrable for a.e. x, & f^y is integrable for a.e. y. This completes the proof for the case $f \geq 0$.

For general $f \in L^1(X)$, write

$$f = (f_1 - f_2) + i(f_3 - f_4)$$

where f_1, f_2, f_3, f_4. Then (exercise) apply the preceding case to each f_i, & show that the result follows. ∎

<u>Remark</u>

Technically, elements of L^1 are equivalence classes of functions that are equal a.e. Fubini's Theorem applies to any of the representatives of f.

<u>Remark</u>

Again, remember the Corollary to Tonelli's Theorem, which tells us that to check whether $f \in L^1(X \times Y)$, we can check any <u>one</u> of three integrals, whichever is more convenient.

Unfortunately, $\mu \times \nu$ is rarely complete, so sometimes we need a version of Tonelli/Fubini for the completion of $\mu \times \nu$. For the general case, see Folland's text.

For example, Lebesgue measure on $\mathbb{R}^n$ is the completion of the n-fold product of Lebesgue measure on $\mathbb{R}$. The statement of Tonelli & Fubini for Lebesgue measure is as follows

Tonelli's Theorem

Let $E \subseteq \mathbb{R}^m$ & $F \subseteq \mathbb{R}^n$ be Lebesgue measurable. If $f: E \times F \to [0,\infty]$ is (m), then:

a. f_x is (m) $\forall x \in E$ (Lebesgue measurable)

b. f^y is (m) $\forall y \in F$

c. $g(x) = \int_F f_x(y)\,dy$ is (m) on E

d. $h(y) = \int_E f^y(x)\,dx$ is (m) on F

e. $\displaystyle\iint_{E\times F} f(x,y)\,(dx\,dy) = \int_F \left(\int_E f(x,y)\,dx\right) dy$

$$= \int_E \left(\int_F f(x,y)\,dy\right) dx$$

Fubini's Theorem

Let $E \subseteq \mathbb{R}^m$ & $F \subseteq \mathbb{R}^n$ be Lebesgue measurable.

If $f \in L^1(E \times F)$, then:

a. $f_x \in L^1(F)$ for a.e. $x \in E$

b. $f^y \in L^1(E)$ for a.e. $y \in F$

c. $g(x) = \int_F f(x,y)\,dy$ belongs to $L^1(E)$

d. $h(y) = \int_E f(x,y)\,dx$ belongs to $L^1(F)$

e. $$\iint_{E \times F} f(x,y)\,(dx\,dy) = \int_F \left(\int_E f(x,y)\,dx \right) dy$$

$$= \int_E \left(\int_F f(x,y)\,dy \right) dx$$

The Integral as the Area under the Graph

Exercise (See Folland 2.5 #50)
Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space, & let $f: X \to [0,\infty]$ be ⓜ.

a. The region under the graph of f is

$$G_f = \{(x,y) \in X \times [0,\infty] : 0 \le y \le f(x)\}$$

Show that G_f is $\mathcal{M} \times \mathcal{B}$ - measurable
($\mathcal{B}$ = Borel σ-algebra in $\mathbb{R}$)

and that

$$(\mu \times m)(G_f) = \int f \, d\mu$$

(m = Lebesgue measure).

See Folland for a hint.

b. The graph of f is

$$\Gamma_f = \{(x, f(x)) : x \in X\} \subseteq X \times [0,\infty].$$

Show that Γ_f is ⓜ, & $(\mu \times m)(\Gamma_f) = 0$.

Hint: Consider

$$E_n = \{\varepsilon n \le f < \varepsilon(n+1)\} \quad \& \quad E_\infty = \{f = \infty\}$$

<u>Convolution</u>

Let $f, g : \mathbb{R} \longrightarrow \overline{\mathbb{R}}$ be Lebesgue ⓜ.

a. Show that $f(x)g(y)$ is Lebesgue ⓜ on $\mathbb{R}^2$.

b. Show $f(x-y)g(y)$ is ⓜ on $\mathbb{R}^2$.

Hint: Composition of part a with a linear transformation

c. Given $f, g \in L^1(\mathbb{R})$, show their <u>convolution</u>

$$(f * g)(x) = \int f(x-y)\, g(y)\, dy$$

is defined a.e. & is ⓜ.

Hint: Show $\iint |f(x-y)\, g(y)|\, dx\, dy < \infty$,
apply Fubini.

d. Show $f * g \in L^1(\mathbb{R})$ & $\|f*g\|_1 \le \|f\|_1 \|g\|_1$

(Thus $L^1(\mathbb{R})$ is a <u>Banach algebra</u> under convolution).

3. Signed Measures & Differentiation

3.1 Signed Measures

Motivation

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

One way to construct other measures on X is to take a (m) $f: X \to [0,\infty]$ & define

$$\nu(E) = \int_E f \, d\mu, \quad E \in \mathcal{M}.$$

Later we will see that this procedure constructs the measures that are "absolutely continuous" w.r.t. μ.

Not all measures on X will have this form, in general. For example, let μ be Lebesgue measure on $\mathbb{R}$. Is the δ measure absolutely continuous w.r.t. μ? That is, is there a (m) function $f: \mathbb{R} \to [0,\infty]$ s.t. $\delta(E) = \int_E f$ for all (m) E? If there was, then we would have

$$1 = \delta\{0\} = \int_{\{0\}} f(x)\, dx = 0,$$

which is a contradiction.

On the other hand, the absolutely continuous measures do provide us with a large class of measures that we can use for motivation.

For example, what goes wrong if we do not require f to be nonnegative? If f is an extended integrable function, then we can still define a function $\nu : \mathcal{M} \to \overline{\mathbb{R}}$ by

$$\nu(E) = \int_E f \, d\mu, \qquad E \in \mathcal{M}.$$

This function has the following properties.

a. $\nu(\emptyset) = 0$

b. if $E_1, E_2, \ldots \in \mathcal{M}$ are disjoint, then

$$\nu\left(\bigcup E_k\right) = \sum_{k=1}^{\infty} \nu(E_k)$$

These are the same properties satisfied by a measure, except that ν need not be nonnegative.

Exercise

Show ν is well-defined, that properties a & b hold, and ν can take at most one of the values $+\infty$ or $-\infty$.

Such a ν is an example of a <u>signed measure</u>.

<u>Definition.</u>

Let $(X, \mathcal{M})$ be a measurable space.

A <u>signed measure</u> on $(X, \mathcal{M})$ is a function $\nu : X \to \overline{\mathbb{R}}$ that satisfies

a. $\nu(\emptyset) = 0$

b. ν assumes at most one of the values $\pm\infty$.

c. If $E_1, E_2, \ldots \in \mathcal{M}$ are disjoint, then

$$\nu\Big(\bigcup_k E_k\Big) = \sum_k{}' \nu(E_k)$$

If $\nu(E) \geq 0$ $\forall E \in \mathcal{M}$ then ν is a measure on X, which for emphasis we sometimes call a <u>positive measure</u> & write $\nu \geq 0$.

If $|\nu(E)| < \infty$ $\forall E \in \mathcal{M}$, then we say that ν is a <u>bounded measure</u> or a <u>finite measure</u>.

Remark

If $E_1, E_2, \ldots$ are disjoint & $\left|\nu\left(\bigcup E_k\right)\right| < \infty$,

Then the series $\sum_k' \nu(E_k) = \nu\left(\bigcup E_k\right)$ converges to a finite real value. Since the union $\bigcup E_k$ is independent of the ordering of the E_k, this implies that $\sum_k' \nu(E_k)$ converges regardless of ordering, i.e., $\sum_k' \nu(E_{\sigma(k)})$ converges for every bijection $\sigma : \mathbb{N} \longrightarrow \mathbb{N}$. Hence $\sum_k' \nu(E_k)$ converges unconditionally, which for a series of scalars is equivalent to absolute convergence.

Hence

$$E_k \text{ disjoint } \& \left|\nu\left(\bigcup E_k\right)\right| < \infty \implies \sum_{k=1}^{\infty}{}' \left|\nu(E_k)\right| < \infty.$$

Exercise

If μ_1, μ_2 are positive measures & at least one is a finite measure, then $\nu = \mu_1 - \mu_2$ is a signed measure.

Exercise

Show by example that a signed measure ν need not be monotonic:

$$A \subseteq B \not\Rightarrow \nu(A) \leq \nu(B) \text{ in general}$$

Exercise

On the other hand, show that we do retain continuity from above & below for signed measures:

a. $E_k \in \mathcal{M}$ & $E_1 \subseteq E_2 \subseteq \cdots \Rightarrow \nu\left(\bigcup_k E_k\right) = \lim_{k\to\infty} \nu(E_k)$

b. $E_k \in \mathcal{M}$, $E_1 \supseteq E_2 \supseteq \cdots$, $|\nu(E_1)| < \infty$

$$\Rightarrow \nu\left(\bigcap_k E_k\right) = \lim_{k\to\infty} \nu(E_k).$$

For a positive measure, any m subset of a zero measure set must have zero measure. Therefore, for signed measures we introduce the following terminology.

<u>Definition</u>

Let ν be a signed measure on $(X, \mathcal{M})$, & let $E \in \mathcal{M}$ be given.

a. E is <u>positive</u> for ν if $\nu(F) \geq 0 \ \forall F \subseteq E$ with $F \in \mathcal{M}$

b. E is <u>negative</u> " $\nu(F) \leq 0$ "

c. E is <u>null</u> " $\nu(F) = 0$ "

<u>Exercise</u>

a. m subsets of positive sets are positive

b. Countable unions of positive sets are positive

Same for negative & null sets.

We will say that two signed measures are "mutually singular" if they "live" on complementary pieces of X.

<u>Definition</u>

Two signed measures μ, ν on $(X, \mathcal{M})$ are <u>mutually singular</u>, denoted $\mu \perp \nu$, if

$\exists A, B \in \mathcal{M}$ s.t.

a. $A \cup B = X$ & $A \cap B = \emptyset$ (ie, $A = B^c$)

b. μ is null on B, and

c. ν is null on A.

<u>Exercise</u>

Show that Lebesgue measure & the δ measure on $\mathbb{R}^d$ are mutually singular, ie,

$$dx \perp \delta.$$

Motivation

Let f be an extended integrable function w.r.t. $(X, \mathcal{M}, \mu)$, where μ is a positive measure. Then, as noted before,

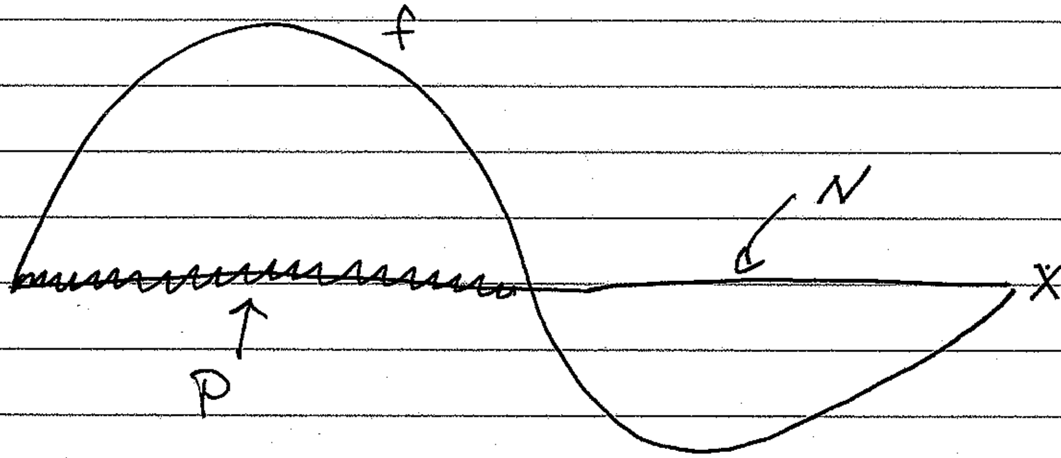
$$\nu(E) = \int_E d\mu$$

defines a signed measure on $(X, \mathcal{M})$. The next exercise obtains a fundamental decomposition of ν.

Exercise

Let f^+, f^- be the positive & negative parts of f. Define

$$P = \mathrm{supp}(f^+), \quad N = P^C.$$



Then define

$$\nu^+(E) = \nu(E \cap P), \quad \nu^-(E) = -\nu(E \cap N), \quad E \in \mathcal{M}.$$

Prove the following.

a. $\nu^+(E) = \int_E f^+ \, d\mu, \quad \nu^-(E) = \int_E f^- \, d\mu.$

b. ν^+, ν^- are positive measures.

c. $\nu = \nu^+ - \nu^-$

d. $\nu^+ \perp \nu^-$

Our next goal is to prove the Jordan Decomposition Theorem, which states that a similar decomposition holds for arbitrary signed measures, i.e., we can always write $\nu = \nu^+ - \nu^-$ as a difference of positive & mutually singular measures.

We begin with the Hahn Decomposition Theorem, which finds sets P, N like those in the example above, but for arbitrary signed measures.

Hahn Decomposition Theorem

Let ν be a signed measure on $(X, \mathcal{M})$.

a. $\exists$ positive set P & negative set N for ν such that

$$P \cup N = X \quad \& \quad P \cap N = \emptyset$$

b. If P', N' are any other sets with this property, then

$$P \Delta P' \ \& \ N \Delta N' \text{ are null sets.}$$

The symmetric difference of A & B is $A \Delta B = (A \cup B) \setminus (A \cap B)$.

Proof:

Since ν cannot assume both of the values $\pm\infty$, let us assume that is does not assume $+\infty$. Let

$$m = \sup\{\nu(E) : E \text{ is positive}\}.$$

Then we can find positive sets P_k s.t.

$$\nu(P_k) \nearrow m.$$

Set

$$P = \bigcup_k P_k$$

By an earlier exercise, P is a positive set.

Exercise: Show $\nu(P) = m$. Note that since

ν does not assume the value $+\infty$, this implies that $m < \infty$.

Set $N = X \setminus P$. Our first goal is to show that N is a negative set.

a. Claim: N contains no strictly positive subsets, i.e., any positive subset of N must be a null set.

To see this, suppose that $E \subseteq N$ is a positive set with $\nu(E) \neq 0$. Then, since N & P are disjoint,

$$\begin{aligned} \nu(E \cup P) &= \nu(E) + \nu(P) && \text{(disjoint)} \\ &> \nu(P) && \left(\begin{array}{c}E \text{ positive} \\ \nu(E) \neq 0\end{array}\right) \\ &= m \end{aligned}$$

But $E \cup P$ is positive, so this contradicts the definition of m.

Hence any positive subset E of N must satisfy

$\nu(E) = 0$. This implies (why?) that E is a null set, & proves the claim.

b. Claim: $\forall$ m $A \subseteq N$, $\nu(A) > 0 \Rightarrow \exists$ m $B \subseteq A$ with $\nu(B) > \nu(A)$.

To see this, suppose $A \subseteq N$ is m & $\nu(A) > 0$. Since $\nu(A) > 0$, A cannot be a null set. Step a therefore implies that A is not a positive set. Hence $\exists$ m $C \subseteq A$ with $\nu(C) < 0$. Set $B = A \setminus C$, then $A = B \cup C$ disjointly, so

$$\nu(A) = \nu(B) + \nu(C)$$

$0 < \nu(A) < \infty$

$-\infty \leq \nu(B) < \infty$

$-\infty \leq \nu(C) < 0$

Rearrangement is permitted, and so we obtain

$$\nu(B) = \nu(A) - \nu(C) > \nu(A),$$

which proves the claim.

c. Now we show that N is negative. For a contradiction, suppose N was not negative. Define

$$s_1 = \sup\{\nu(B) : (m)\, B \subseteq N\}$$

Since N is not negative, $s_1 > 0$.

Let $n_1 \in \mathbb{N}$ be the smallest integer with $\frac{1}{n_1} < s_1$.

Then $\exists$ (m) $A_1 \subseteq N$ with

$$0 < \frac{1}{n_1} < \nu(A_1) \leq s_1.$$

By part b,

$$s_2 = \sup\{\nu(B) : (m)\, B \subseteq A_1\} > \nu(A_1)$$

Let $n_2 \in \mathbb{N}$ be the smallest integer such that

$$\nu(A_1) + \frac{1}{n_2} < S_2.$$

Then $\exists$ ⓜ $A_2 \subseteq A_1$ with

$$\nu(A_1) + \frac{1}{n_2} < \nu(A_2) \leq S_2.$$

Continue applying part b repeatedly in the same manner. At each stage $j \geq 2$ we have

$$\nu(A_{j-1}) + \frac{1}{n_j} < \nu(A_j).$$

In particular,

$$\nu(A_1) > \frac{1}{n_1}$$

$$\nu(A_2) > \nu(A_1) + \frac{1}{n_2} > \frac{1}{n_1} + \frac{1}{n_2}$$

$$\nu(A_3) > \nu(A_2) + \frac{1}{n_3} > \frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3}$$

etc.

Thus $\nu(A_j) > \sum_{k=1}^{j} \frac{1}{n_k}$.

Now set $A = \bigcap_{j=1}^{\infty} A_j$. We have $A_1 \supseteq A_2 \supseteq \cdots$

and $0 < \nu(A_1) < \infty$, so we can apply continuity from above to conclude that

$$\infty > \nu(A) = \lim_{j \to \infty} \nu(A_j)$$

$$\geq \lim_{j \to \infty} \sum_{k=1}^{j} \frac{1}{n_k}$$

$$= \sum_{k=1}^{\infty} \frac{1}{n_k} > 0.$$

In particular, we must have $n_k \to \infty$.

Since $A \subseteq N$ & $\nu(A)$, we can again apply claim b to the set A to obtain a (m) set $B \subseteq A$ such that $\nu(A) < \nu(B)$.

Let $n \in \mathbb{N}$ be such that

$$\nu(A) + \frac{1}{n} < \nu(B),$$

and let $\varepsilon > 0$ be such that

$$\nu(A) + \frac{1}{n} + 2\varepsilon < \nu(B).$$

Now, $\nu(A_j) \longrightarrow \nu(A)$, so $\exists\, J > 0$ s.t.

$$j \geqslant J \implies |\nu(A) - \nu(A_j)| < \varepsilon.$$

In other words, $\nu(A_j) - \varepsilon < \nu(A) < \nu(A_j) + \varepsilon$

for all $j \geqslant J$. Hence for those j,

$$\begin{aligned} \nu(B) &> \nu(A) + \tfrac{1}{n} + 2\varepsilon \\ &> \nu(A_j) - \varepsilon + \tfrac{1}{n} + 2\varepsilon \\ &= \nu(A_j) + \tfrac{1}{n} + \varepsilon. \end{aligned}$$

Let $j \geqslant J$ be large enough that $n_j > n$. Then

$$\begin{aligned} \nu(A_j) + \tfrac{1}{n} &< \nu(B) \\ &\leq \sup\{\nu(D) : \text{m } D \subseteq A_j\} \\ &\qquad (\text{since } B \subseteq A \subseteq A_j) \\ &= s_j. \end{aligned}$$

But n_j is the smallest integer with

$$\nu(A_j) + \frac{1}{n_j} < s_j$$

so we have a contradiction. Therefore N must be a negative set.

It remains only to establish the uniqueness claim for P & N. Suppose that P', N' are another choice of sets with the required properties. Then

$$P \setminus P' \subseteq P \quad \text{so } P\setminus P' \text{ is positive}$$

Likewise

$$P \setminus P' \subseteq (P')^c = N' \quad \text{so } P \setminus P' \text{ is negative.}$$

Therefore $P \setminus P'$ is null, & similarly $P' \setminus P$ is null. Hence $P \Delta P' = (P \setminus P') \cup (P' \setminus P)$ is null, & similarly $N \Delta N'$ is null. $\blacksquare$

Notation

We refer to $X = P \cup N$ as a Hahn decomposition of X with respect to ν. It is not unique in general, as we can transfer any ν-null set from P to N or vice versa.

Now we come to the Jordan Decomposition of a signed measure.

Jordan Decomposition Theorem

If ν is a signed measure on $(X, \mathcal{M})$, then there exist unique positive measures ν^+, ν^- such that

$$\nu = \nu^+ - \nu^- \quad \text{and} \quad \nu^+ \perp \nu^-$$

Proof:

Let $X = P \cup N$ be a Hahn decomposition w.r.t. ν.

Define

$$\nu^+(E) = \nu(E \cap P)$$

$$E \in \mathcal{M}.$$

$$\nu^-(E) = -\nu(E \cap N)$$

Exercise: Show that ν^+, ν^- are positive measures.

Since P, N are disjoint with union X, we have $\nu^+ - \nu^- = \nu$. Also,

$$\nu^+(N) = \nu(N \cap P) = \nu(\varnothing) = 0.$$

Since $\nu^+ \geq 0$, this implies that N is a null set for ν^+, & similarly P is a null set for ν^-. Hence $\nu^+ \perp \nu^-$.

To show uniqueness, suppose we also had $\nu = \mu^+ - \mu^-$ for some positive measures with $\mu^+ \perp \mu^-$. Then, by definition of mutually singular, $\exists\, E = F^c$ such that F is null for μ^+, & E is null for E^-. We claim that $X = E \cup F$ is another Hahn decomposition of X w.r.t. ν, specifically, E is a positive set for ν & F is a negative set.

Suppose that $G \subseteq F$, G (m). Then by monotonicity,

$$0 \leq \mu^+(G) \leq \mu^+(F) = 0,$$

so

$$\nu(G) = \mu^+(G) - \mu^-(G) = -\mu^-(G) \leq 0.$$

Thus ν is indeed negative on F, & similarly positive on E.

By the Hahn Decomposition Theorem, $P \triangle E$ & $N \triangle F$ are therefore null sets for ν. Given any $A \in \mathcal{M}$, we have

$$\mu^+(A \cap F) = 0 \qquad \text{since } F \text{ is null for } \mu^+$$

$$\mu^-(A \cap E) = 0 \qquad \text{since } E \text{ is null for } \mu^-$$

and therefore

$$\nu^+(A) = \nu(A \cap P) \qquad \text{def. of } \nu^+$$

$$= \nu(A \cap E) \qquad \text{Exercise - follow from } P \triangle E \text{ is } \nu\text{-null}$$

$$= \mu^+(A \cap E) - \cancel{\mu^-(A \cap E)}^{\,0} \qquad \nu = \mu^+ - \mu^-$$

$$= \mu^+(A \cap E) + \cancel{\mu^+(A \cap F)}^{\,0}$$

$$= \mu^+(A) \qquad X = E \cup F \text{ disjointly}$$

Hence $\nu^+ = \mu^+$ & similarly $\nu^- = \mu^-$. ∎

Notation

We call $\nu = \nu^+ - \nu^-$ the Jordan decomposition of ν.

ν^+ is the positive variation of ν.

ν^- is the negative variation of ν

The total variation of ν is the positive measure

$$|\nu| = \nu^+ + \nu^-.$$

That is, $|\nu|(E) = \nu^+(E) + \nu^-(E)$, $E \in \mathcal{M}$.

Remark

Note that

$$|\nu(E)| = |\nu^+(E) - \nu^-(E)|$$

$$\leq \nu^+(E) + \nu^-(E) = |\nu|(E).$$

Moreover, $|\nu|$ is a positive measure and hence is monotonic. Therefore

$$|\nu(E)| \leq |\nu|(E) \leq |\nu|(X) \quad \forall E \in \mathcal{M}.$$

Of course, $|\nu|(X)$ could be infinite, but if it is finite then $|\nu(E)| \leq |\nu|(X) < \infty$ $\forall E \in \mathcal{M}$, so ν is a bounded measure. In fact, the converse is also true, which explains the terminology "bounded measure" instead of just "finite measure".

Exercise

Show that

$$\nu \text{ is bounded} \iff |\nu|(X) < \infty,$$

and in this case,

$$\sup_{E \in \mathcal{M}} |\nu(E)| \leq |\nu|(X)$$

Exercise

Prove the following useful equivalent characterizations of the variation measures.

$$\nu^+(E) = \sup\{\nu(A) : \text{(m) } A \subseteq E\}$$

$$\nu^-(E) = -\inf\{\nu(A) : \text{(m) } A \subseteq E\}$$

$$|\nu|(E) = \sup\left\{\sum_{k=1}^{n} |\nu(E_k)| : E = \bigcup_{k=1}^{n} E_k \text{ disjointly},\ E_k \text{ (m)}\right\}$$

Exercises.

1. Let ν be a signed measure on $(X, \mathcal{M})$. Show that
$$E \in \mathcal{M} \text{ is } \nu\text{-null} \iff |\nu|(E) = 0$$

2. Show that
$$|\nu|(E) = \sup\left\{ \left| \int_E f \, d\nu \right| : |f| \leq 1, \ f \ \text{(m)} \right\}$$

3. Let $M_r(X) = \{\nu : \nu \text{ is a bounded signed measure on } X\}$.

Show $M_r(X)$ is a (real) vector space, and $\|\nu\| = |\nu|(X)$ is a norm on X.

4. Let $X = \mathbb{N}$ & $\mathcal{M} = \mathcal{P}(\mathbb{N})$. Show that a signed measure ν on $\mathbb{N}$ is completely determined by the values $\{\nu\{k\}\}_{k \in \mathbb{N}}$, i.e.,
$$\nu \longmapsto \{\nu\{k\}\}_{k \in \mathbb{N}}$$
is an injective map of the signed measures on $\mathbb{N}$ into the space of all sequences of extended real numbers. Identify the range of this map, & identify the sequences that correspond to the bounded measures on $\mathbb{N}$.

Signed Measures and Integration

Definition

Let ν be a signed measure on (X, μ) and let $\nu = \nu^+ - \nu^-$ be its Jordan decomposition.

If $f : X \to \overline{\mathbb{R}}$ or $\mathbb{C}$ is m and $\int |f| \, d\nu^+, \int |f| \, d\nu^- < \infty$, then we set

$$\int f \, d\nu = \int f \, d\nu^+ - \int f \, d\nu^-.$$

Exercise

Show that

$$\int |f| d\nu^+, \int |f| d\nu^- < \infty \iff \int |f| \, d|\nu| < \infty,$$

and, in this case,

$$\left| \int f \, d\nu \right| \leq \int |f| \, d|\nu|$$

Definition

$$L^1(\nu) = L^1(|\nu|) = L^1(\nu^+) \cap L^1(\nu^-)$$

$$\|f\|_1 = \int |f| \, d|\nu|$$

Exercises

1. Let μ be a positive measure on $(X, \mathcal{M})$ & let g be an extended integrable function on X. Let $\nu(E) = \int_E g\, d\mu$, so ν is a signed measure.

Show that if $f \in L^1(\nu)$, then

$$\int f\, d\nu = \int fg\, d\mu.$$

2. Let $\omega : \mathbb{N} \to (0, \infty)$ be fixed. Define the <u>weighted ℓ^1 space</u>

$$\ell^1_\omega = \left\{ x = (x_k)_{k\in\mathbb{N}} : x_k \in \mathbb{R} \text{ \& } \sum_{k=1}^{\infty} |x_k|\, \omega_k < \infty \right\}$$

Set

$$\|x\|_{1,\omega} = \sum_{k=1}^{\infty} |x_k|\, \omega_k, \quad x \in \ell^1_\omega.$$

Show that ℓ^1_ω is a Banach space, i.e., a complete normed vector space w.r.t. $\|\cdot\|_{1,\omega}$.

3. Let μ be a positive measure on $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$.

Set $\omega_k = \mu\{k\}$. Show that if $f: \mathbb{N} \to [0,\infty)$, then

$\int f\,d\mu = \sum_{k=1}^{\infty} f(k)\,\omega_k$. Show that

$L^1(\mu) = \ell^1_\omega$.

4. Extend #3 to signed measures.

5. Let μ be counting measure on $(X, \mathcal{P}(X))$.
Show that if $f: X \to [0,\infty]$ then

$$\int f\,d\mu = \sup\left\{ \sum_{j=1}^{N} f(x_j) : N \in \mathbb{N},\ x_j \in X \right\}$$

6. Let μ, ν be finite signed measures on $(X, \mathcal{M})$.

Show that if $f \in L^1(\mu) \cap L^1(\nu)$, then

$$\int f\,d(\mu+\nu) = \int f\,d\mu + \int f\,d\nu.$$

3.2 The Lebesgue-Radon-Nikodym Theorem

Notation

Recall that if μ is a positive measure on $(X, \mathcal{M})$ and f is an extended μ-integrable function, then

$$\nu(E) = \int_E f \, d\mu \qquad (*)$$

defines a signed measure on $(X, \mathcal{M})$. Further,

$$\nu^+(E) = \int f^+ \, d\mu, \quad \nu^-(E) = \int f^- \, d\mu,$$

$$|\nu|(E) = \int |f| \, d\mu,$$

and ν is bounded if $f \in L^1(\nu)$. We will write

$$d\nu = f \, d\mu$$

to mean that ν is the measure given by $(*)$. Sometimes it will be convenient to abuse notation a bit and write $\nu = f \, d\mu$ to mean ν is given by $(*)$.

Definition

A signed measure ν on $(X, \mathcal{M})$ is absolutely continuous w.r.t. a positive measure μ on $(X, \mathcal{M})$, denoted $\nu \ll \mu$, if

$$\forall E \in \mathcal{M}, \quad \mu(E) = 0 \implies \nu(E) = 0.$$

Exercise

Show that if $d\nu = f\, d\mu$, then $\nu \ll \mu$.

We will see in the Radon-Nikodym Theorem that all measures absolutely continuous w.r.t. μ have the form $d\nu = f\, d\mu$ for some f.

Exercise

Show that

$$\nu \ll \mu \ \& \ \nu \perp \mu \implies \nu = 0.$$

The Lebesgue-Radon-Nikodym Theorem will write an arbitrary signed measure ν as a sum of an absolutely continuous part (w.r.t. μ) & a singular part (w.r.t. μ).

What does "absolute continuity" have to do with continuity?

<u>Theorem</u>

Suppose ν is a finite signed measure & μ a positive measure on $(X, \mathcal{M})$. Then TFAE.

a. $\nu \ll \mu$

b. $\forall \varepsilon > 0 \ \exists \delta > 0$ s.t. $\forall E \in \mathcal{M}$, $\mu(E) < \delta \Rightarrow |\nu(E)| < \varepsilon$.

Proof:

$\Rightarrow$ Suppose $\nu \geq 0$ & Suppose that statement b failed. Then $\exists \varepsilon > 0$ such that for each $\delta = 2^{-n}$ there exists an $E_n \in \mathcal{M}$ s.t.

$$\mu(E_n) < 2^{-n} \quad \text{but} \quad \nu(E_n) \geq \varepsilon.$$

Let

$$F_k = \bigcup_{n=k}^{\infty} E_n \quad \& \quad F = \bigcap_{k=1}^{\infty} F_k = \limsup_{n \to \infty} E_n.$$

Exercise: $\mu(F) = 0$ & $\nu(F) \geq \varepsilon$

because $\nu \geq 0$ & ν is finite

Hence $\nu \not\ll \mu$.

Exercise: Extend to signed ν.

$\Longleftarrow$ Exercise. $\blacksquare$

Exercise

a. Use the preceding theorem to show that if $\mu \geq 0$ and $f \in L^1(\mu)$, then $\forall \varepsilon > 0$ $\exists \delta > 0$ s.t.

$$\mu(E) < \delta \implies \left| \int_E f \, d\mu \right| < \varepsilon.$$

b. Also give a direct proof of part a by considering

$$f_n(x) = \begin{cases} |f(x)|, & \text{if } |f(x)| \leq n \\ n, & \text{if } |f(x)| > n \end{cases}$$

Note that $f_n \nearrow |f|$.

Lemma

Let μ, ν be finite positive measures on $(X, \mathcal{M})$.
Then either

a. $\nu \perp \mu$, or

b. $\exists \varepsilon > 0$ $\exists E \in \mathcal{M}$ s.t. $\mu(E) > 0$ and

E is a positive set for $\nu - \varepsilon\mu$, ie.,

$$\nu(A) \geq \varepsilon\mu(A) \quad \forall \text{ m } A \subseteq E.$$

Proof:

For each $n \in \mathbb{N}$, let $X = P_n \cup N_n$ be a Hahn decomposition for the signed measure $\nu - \frac{1}{n}\mu$. Define

$$P = \cup P_n, \qquad N = P^c = \cap N_n.$$

Then we have that N is a negative set for $\nu - \frac{1}{n}\mu$ for every n. Therefore, if $E \subseteq N$ is m, then

$$0 \leq \nu(E) \leq \frac{1}{n}\mu(E) \quad \forall n \in \mathbb{N},$$

so $\nu(E) = 0$ since μ is bounded.

Hence N is a null set for ν, which is equivalent to $\nu(N) = 0$ since $\nu \geq 0$.

If $\mu(P) = 0$ then P is a null set for μ, & therefore $\mu \perp \nu$.

On the other hand, if $\mu(P) > 0$ then we must have $\mu(P_n) > 0$ for some n. Since P_n is a positive set for $\nu - \frac{1}{n}\mu$, statement b therefore holds with $E = P_n$ & $\varepsilon = \frac{1}{n}$. ∎

<u>Definition</u>
We say that a signed measure ν is <u>σ-finite</u> if its total variation $|\nu|$ is σ-finite.

Lebesgue-Radon-Nikodym Theorem

Let ν be a σ-finite signed measure on $(X, \mathcal{M})$ and μ a σ-finite positive measure on $(X, \mathcal{M})$.

a. There exist unique σ-finite signed measures ρ, λ on $(X, \mathcal{M})$ such that

$$\nu = \rho + \lambda, \quad \rho \ll \mu, \quad \lambda \perp \mu.$$

b. There exists an extended μ-integrable function f such that $d\rho = f\,d\mu$, i.e.,

$$\nu = f\,d\mu + \lambda.$$

c. If we also have $\nu = \tilde{f}\,d\mu + \lambda$ where $\tilde{f}$ is an extended μ-integrable function, then

$$\tilde{f} = f \quad \mu\text{-a.e.}$$

Proof:

Case 1: Suppose first that μ & ν are both finite, positive measures. Let

$$\mathcal{F} = \left\{ f : X \to [0,\infty] : f \text{ ⓜ } \& \int_E f\,d\mu \le \nu(E) \quad \forall E \in \mathcal{M} \right\}$$

Since $0 \in \mathcal{F}$ we know $\mathcal{F}$ is nonempty.

Suppose $f, g \in \mathcal{F}$ and set $h = \max\{f, g\}$.

If $E \in \mathcal{M}$, then

$$\int_E h\, d\mu = \int_{E \cap \{f > g\}} f\, d\mu + \int_{E \cap \{g \geq f\}} g\, d\mu$$

$$\leq \nu(E \cap \{f > g\}) + \nu(E \cap \{g \geq f\})$$

$$= \nu(E),$$

so $h \in \mathcal{F}$.

If $f \in \mathcal{F}$, then $0 \leq \int_X f\, d\mu \leq \nu(X) < \infty$, so

$$0 \leq a = \sup\left\{\int f\, d\mu : f \in \mathcal{F}\right\} \leq \nu(X) < \infty.$$

By definition, we can find $f_n \in \mathcal{F}$ such that $\int f_n\, d\mu \longrightarrow a$. Consider

$$g_n = \max\{f_1, \ldots, f_n\} \in \mathcal{F}$$

Define

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

Then f is (m), and if $E \in \mathcal{M}$ then since $g_n \uparrow f$ we have

$$\int_E f\,d\mu = \lim_{n\to\infty} \int_E g_n\,d\mu \qquad \text{MCT}$$
$$\leq \nu(E) \qquad \text{since } g_n \in \mathcal{F}$$

and therefore $f \in \mathcal{F}$. Also,

$$a = \lim_{n\to\infty} \int f_n\,d\mu \leq \lim_{n\to\infty} \int g_n\,d\mu = \int f\,d\mu \leq a,$$

so $\int f\,d\mu = a$. Since $f \geq 0$ & $a < \infty$, this implies $0 \leq f < \infty$ μ-a.e. (so by redefining f on a set of measure zero, we can take f to be real-valued everywhere).

We claim now that

$$d\rho = f\,d\mu \quad \text{and} \quad \lambda = \nu - f\,d\mu$$

are the required measures. These certainly are signed measures, and since $f \in \mathcal{F}$ we have

$\rho(E) = \int_E f\,d\mu \leq \nu(E) \quad \forall E \in \mathcal{M}$, so

$\lambda = \nu - \rho$ is a positive measure. Further, we know $\rho = f\,d\mu \ll \mu$, and $\nu = \rho + \lambda$, so it remains only to show that $\lambda \perp \mu$.

By the lemma, if $\lambda \not\perp \mu$, then $\exists\, \varepsilon > 0$ and $E \in \mathcal{M}$ with $\mu(E) > 0$ such that E is a positive set for $\lambda - \varepsilon\mu$. We claim that $f + \varepsilon\chi_E \in \mathcal{F}$.

To see this, suppose $A \in \mathcal{M}$. Then

$$\begin{aligned}
\int_A f + \varepsilon\chi_E \, d\mu &= \int_A f\,d\mu + \varepsilon \int_{A\cap E} d\mu \\
&= \rho(A) + \varepsilon\mu(A \cap E) \\
&\leq \rho(A) + \lambda(A \cap E) \\
&\leq \rho(A) + \lambda(A) \qquad \text{since } \lambda \geq 0 \\
&= \nu(A).
\end{aligned}$$

Thus we have $f+\varepsilon\chi_E \in \mathcal{F}$. But then

$$\begin{aligned} a &= \int f\,d\mu \\ &< \int f\,d\mu + \varepsilon\mu(E) \qquad \text{since } \mu(E)>0 \\ &= \int f+\varepsilon\chi_E\,d\mu \\ &\leq a \qquad \text{since } f+\varepsilon\chi_E \in \mathcal{F}, \end{aligned}$$

which is a contradiction. Hence we must indeed have $\lambda \perp \mu$.

Now we show uniqueness. Suppose that we also had $\nu = \rho' + \lambda'$ with $\rho' \ll \mu$ & $\lambda' \perp \mu$. Then since $\nu = \rho + \lambda$, we have $\lambda - \lambda' = \rho' - \rho$. But

$$\lambda - \lambda' \perp \mu \quad \& \quad \rho' - \rho \ll \mu,$$

which implies $\lambda - \lambda' = 0 = \rho' - \rho$.

Also, if $d\rho = \tilde{f}\,d\mu = f\,d\mu$, then

$$\int_E (f - \tilde{f})\,d\mu = 0 \quad \forall E \in \mathcal{M},$$

which implies by an earlier result that $f = \tilde{f}$ μ-a.e.

<u>Case 2</u>: Suppose that μ, ν are both σ-finite <u>positive</u> measures.

By applying the disjointization trick, we can write $X = \bigcup E_j$ and $X = \bigcup F_k$ as disjoint unions with $\mu(E_j) < \infty$, $\nu(F_k) < \infty$. Then

$$X = \bigcup_{j,k} (E_j \cap F_k) = \bigcup_{\ell} A_\ell \text{ disjointly}$$

with $\mu(A_\ell), \nu(A_\ell) < \infty$ $\forall \ell$. Define

$$\mu_k(E) = \mu(E \cap A_k), \quad \nu_k(E) = \nu(E \cap A_k).$$

Then each μ_k, ν_k is a finite positive measure, so by Case 1 we can write $\nu_k = \rho_k + \lambda_k$ for some unique measures with $\rho_k \ll \mu_k$ & $\lambda_k \perp \mu_k$.

Note that

$$\mu_k(A_k^c) = \mu(A_k^c \cap A_k) = 0,$$

so A_k^c is a μ_k-null set. Therefore

$$f_k' = f_k \cdot \chi_{A_k}$$

equals f_k μ_k-a.e., so we can replace f_k with f_k' without changing λ_k or ρ_k. In other words, we can assume that $f_k(x) = 0$ $\forall x \notin A_k$.

Since the A_k are disjoint, we can therefore define

$$f = \sum_{k=1}^{\infty} f_k .$$

Since $f \geq 0$,

$$d\rho = f \, d\mu$$

defines a positive measure. Also,

$$\lambda = \sum_{k=1}^{\infty} \lambda_k$$

is a positive measure, since each $\lambda_k \geq 0$.

Exercises: λ, ρ are σ-finite, $\nu = \rho + \lambda$, $\rho \ll \mu$, $\lambda \perp \mu$, & the uniqueness statements hold.

Case 3: Exercise: Extend to an arbitrary signed σ-finite measure ν by considering ν^+ & ν^-. ∎

Notation
We refer to $\nu = \rho + \lambda$ as the Lebesgue decomposition of ν w.r.t. μ.

The special case where $\nu \ll \mu$ (i.e., $\lambda = 0$) is important.

Radon-Nikodym Theorem
If ν is a σ-finite signed measure on $(X, \mathcal{M})$ and μ is a σ-finite positive measure on $(X, \mathcal{M})$ such that $\nu \ll \mu$, then $\exists$ extended μ-integrable function f such that $d\nu = f\, d\mu$. Any two functions with this property are equal μ-a.e.

Notation

If $\nu \ll \mu$, then the function f s.t. $d\nu = f\,d\mu$ is called the Radon-Nikodym derivative of ν w.r.t. μ, and is often denoted

$$f = \frac{d\nu}{d\mu},$$

i.e.,

$$d\nu = \frac{d\nu}{d\mu}\,d\mu.$$

It is unique up to sets of μ-measure zero.

Note

By an earlier exercise, if $d\nu = f\,d\mu$, then $d|\nu| = |f|\,d\mu$, i.e., $d|\nu| = \left|\frac{d\nu}{d\mu}\right| d\mu$.

Exercise

Let ν be a σ-finite signed measure & μ, λ σ-finite positive measures. Show that

$$\nu \ll \mu \;\&\; \mu \ll \lambda \Rightarrow \nu \ll \lambda$$

and, in this case, if $d\nu = f\,d\mu$ & $d\mu = g\,d\lambda$, Then $d\nu = fg\, d\lambda$. In other words,

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}.$$

<u>Remark</u>

We know that $\delta \not\ll dx$, ie., δ is not absolutely continuous w.r.t. Lebesgue measure. Pretend that we did have $\delta \ll dx$. Then there would exist a Radon-Nikodym derivative, a function that we will call $\delta(x)$ that satisfies $d\delta = \delta(x)\, dx$. Then

$$f(0) \underset{\text{earlier exercise}}{=} \int f\, d\delta = \int f(x)\delta(x)\, dx$$

There is <u>no such function $\delta(x)$</u>, but it is common to abuse notation & write $\int f(x)\,\delta(x)\, dx = f(0)$. However, what is really meant is not an integral w.r.t. Lebesgue measure dx but rather the integral

$\int f\, d\delta = \int f(x)\, d\delta(x)$ w.r.t. the δ-measure.

Exercise

Let μ denote counting measure on $\mathbb{R}$, which is not σ-finite.

a. Show that $dx \ll \mu$, but $dx \neq f\,d\mu$ for any function f.

b. Prove that μ has no Lebesgue decomposition w.r.t. dx, i.e., $\nexists$ signed measures ρ, λ with $\mu = \rho + \lambda$, $\rho \ll dx$, & $\lambda \perp dx$.

3.3 Complex Measures

Definition

A function $\nu: \mathcal{M} \to \mathbb{C}$ is a complex measure on a measurable space $(X, \mathcal{M})$ if

a. $\nu(\emptyset) = 0$

b. If $E_1, E_2, \ldots \in \mathcal{M}$ are disjoint, then

$$\nu\left(\bigcup E_k\right) = \sum_k{}' \nu(E_k).$$

Remark

As was the case for signed measures, since the union $\bigcup E_k$ is independent of order, the series $\sum' \nu(E_k)$ must converge unconditionally, & therefore must converge absolutely, i.e., $\sum' |\nu(E_k)|$.

Note

A complex measure can only take complex values, & hence all complex measures are finite measures by definition.

Furthermore, they are bounded measures in the following sense.

Exercise/Notation

Let ν be a complex measure. Define

$$\nu_r(E) = \operatorname{Re} \nu(E) \quad \& \quad \nu_i(E) = \operatorname{Im} \nu(E), \qquad E \in \mathcal{M}.$$

Show that ν_r, ν_i are bounded signed measures, & for $E \in \mathcal{M}$ we have

$$|\nu(E)| \leq |\nu_r(E)| + |\nu_i(E)| \leq |\nu_r|(X) + |\nu_i|(X)$$

Therefore ν is bounded in the sense that

$$\sup_{E \in \mathcal{M}} |\nu(E)| < \infty.$$

We refer to ν_r as the <u>real part</u> of ν, and ν_i as the <u>imaginary part</u>.

Definition: Integration

If ν is a complex measure, then we set

$$L^1(\nu) = L^1(\nu_r) \cap L^1(\nu_i) = L^1(|\nu_r|) \cap L^1(|\nu_i|).$$

If $f \in L^1(\nu)$, then

$$\int f \, d\nu = \int f \, d\nu_r + i \int f \, d\nu_i$$

Note that $\int f \, d\nu$ is a well-defined complex scalar.

Most definitions likewise carry over by appling them to the real & imaginary parts.

Definition

A complex measure ν on $(X, \mathcal{M})$ is absolutely continuous w.r.t. a positive measure μ on $(X, \mathcal{M})$, denoted $\nu \ll \mu$, if

$$\forall E \in \mathcal{M}, \quad \mu(E) = 0 \Rightarrow \nu(E) = 0$$

Equivalently, $\nu \ll \mu$ if $\nu_r \ll \mu$ & $\nu_i \ll \mu$.

Exercise

If μ is a positive measure & $g \in L^1(\mu)$ (complex version), then $d\nu = g\,d\mu$ defines a complex measure & $\nu \ll \mu$.

Definition

A complex measure is *singular* w.r.t. another complex measure μ, denoted $\nu \perp \mu$, if

$$\nu_r \perp \mu_r, \quad \nu_r \perp \mu_i, \quad \nu_i \perp \mu_r, \quad \nu_i \perp \mu_i.$$

Lebesgue-Radon-Nikodym Theorem

Let ν be a complex measure on $(X, \mathcal{M})$ and μ a σ-finite positive measure on $(X, \mathcal{M})$.
Then $\exists f \in L^1(\mu)$ & a complex measure λ s.t.

$$\nu = f\,d\mu + \lambda, \qquad \lambda \perp \mu.$$

If we also have $\nu = \tilde{f}\,d\mu + \tilde{\lambda}$ where $\tilde{f} \in L^1(\mu)$ and $\tilde{\lambda} \perp \mu$, then

$$\tilde{\lambda} = \lambda \quad \& \quad \tilde{f} = f \;\; \mu\text{-a.e.}$$

We need the following exercise in order to define the total variation of a complex measure.

<u>Exercise</u>

Let ν be a complex measure on $(X, \mathcal{M})$. Let

$$\mu = |\nu_r| + |\nu_i|,$$

and show that μ is a bounded positive measure with

$\nu \ll \mu$. Conclude that $d\nu = f\, d\mu$ for some $f \in L^1(\mu)$.

<u>Remark</u>

The total variation of a complex measure is a little more awkward to define than it is for a signed measure. By the preceding exercise, there is at least one positive measure μ & a function $f \in L^1(\mu)$ such that $d\nu = f\, d\mu$, and we define the total variation $|\nu|$ of ν to be the measure

$$d|\nu| = |f|\, d\mu.$$

However, we want this definition to be independent of the choice of μ & f, which is the content of the following theorem.

<u>Theorem</u>

Let ν be a complex measure on $(X, \mathcal{M})$.

If μ_1, μ_2 are bounded positive measures & $f_1 \in L^1(\mu_1)$, $f_2 \in L^2(\mu_2)$ are such that

$$f_1 \, d\mu_1 = d\nu = f_2 \, d\mu_2,$$

then $|f_1| \, d\mu_1 = |f_2| \, d\mu_2$.

<u>Proof:</u>

Since $\mu_1 \ll \mu_1 + \mu_2$, $\exists \, g_1 \in L^1(\mu_1 + \mu_2)$ s.t. $d\mu_1 = g_1 \, d(\mu_1 + \mu_2)$. Likewise $\exists \, g_2 \in L^1(\mu_1 + \mu_2)$ s.t. $d\mu_2 = g_2 \, d(\mu_1 + \mu_2)$. Because μ_1, μ_2, & $\mu_1 + \mu_2$ are all positive measures, we have $g_1, g_2 \geq 0$ $(\mu_1 + \mu_2)$-a.e.

Now, by a previous exercise, since

$d\nu = f_1 \, d\mu_1$ & $d\mu_1 = g_1 \, d(\mu_1 + \mu_2)$, we have

$$d\nu = f_1 g_1 \, d(\mu_1 + \mu_2).$$

Likewise

$$d\nu = f_2 g_2 \, d(\mu_1 + \mu_2).$$

But then the uniqueness clause of the Lebesgue-Radon-Nikodym Theorem implies that

$$f_1 g_1 = f_2 g_2 \quad (\mu_1+\mu_2)\text{-a.e.}$$

Consequently,

$$|f_1|\, g_1 = |f_1 g_1| = |f_2 g_2| = |f_2|\, g_2 \quad (\mu_1+\mu_2)\text{-a.e.},$$

and hence

$$\begin{aligned} |f_1|\, d\mu_1 &= |f_1|\, g_1\, d(\mu_1+\mu_2) \\ &= |f_2|\, g_2\, d(\mu_1+\mu_2) \\ &= |f_2|\, d\mu_2 . \quad \blacksquare \end{aligned}$$

<u>Definition</u>
Let ν be a complex measure on $(X, \mathcal{M})$. Then the <u>total variation</u> $|\nu|$ of ν is the positive measure $d|\nu| = |f|\, d\mu$ where μ is any positive bounded measure and f any function in $L^1(\mu)$ such that $d\nu = f\, d\mu$.

<u>Exercise</u>

Let ν be a complex measure on $(X, \mathcal{M})$.

a. Show $|\nu(E)| \leq |\nu|(E) \quad \forall E \in \mathcal{M}$.

b. Show $\nu \ll |\nu|$, & $\exists\, g$ with $|g| = 1$ $|\nu|$-a.e.

such that $d\nu = g\, d|\nu|$. (This is called the <u>polar decomposition</u> of ν.)

c. If $f \in L^1(\nu)$, then $\left|\int f\, d\nu\right| \leq \int |f|\, d|\nu|$.

The following exercise gives some equivalent reformulations of the total variation that are sometimes easier to apply than our definition. In many texts, one of these forms is taken as the definition of $|\nu|$, and the other forms are derived from it.

<u>Hints</u>

a. By definition of total variation, $d|\nu| = |f|\, d\mu$ where $\mu \geq 0$, $f \in L^1(\mu)$, & $d\nu = f\, d\mu$. Hence $|\nu(E)| = \left|\int_E f\, d\mu\right|$.

b. Part a implies $\nu \ll |\nu|$. So $d\nu = g\, d|\nu|$ for some $g \in L^1(|\nu|)$. Then $d|\nu| = |g|\, d|\nu|$. Show $\int_E (|g| - 1)\, d|\nu| = 0 \quad \forall E \in \mathcal{M}$. Use this to show $|g| = 1$ $|\nu|$-a.e.

Exercise

If ν is a complex measure on $(X, \mathcal{M})$, then:

a. $$|\nu|(E) = \sup\left\{ \sum_{k=1}^{n}{}' |\nu(E_k)| : n \in \mathbb{N},\ E_k \in \mathcal{M},\ E = \bigcup_{k=1}^{n} E_k \text{ disjointly} \right\}$$

b. $$|\nu|(E) = \sup\left\{ \sum_{k=1}^{\infty}{}' |\nu(E_k)| : E_k \in \mathcal{M},\ E = \bigcup_{k=1}^{\infty} E_k \text{ disjointly} \right\}$$

c. $$|\nu|(E) = \sup\left\{ \left| \int_E f\, d\nu \right| : |f| \le 1 \ \ |\nu|\text{-a.e.} \right\}$$

Hints:

Let $\mu_1(E)$, $\mu_2(E)$; $\mu_3(E)$ denote the quantities defined on the right-hand side of a, b, c.
Show that $\mu_1 \le \mu_2 \le \mu_3 = |\nu|$ & $\mu_3 \le \mu_1$.

More hints.

$\mu_2 \le \mu_3$: Suppose $E = \bigcup_{k=1}^{\infty} E_k$ disjointly. Let c_k be the scalars of unit modulus such that
$$\left| \int_{E_k} d\nu \right| = c_k \int_{E_k} d\nu, \quad \text{and set } f = \sum{}' c_k \chi_{E_k}.$$

$|\nu| \le \mu_3$: $\exists f$ with $|f| = 1$ ν-a.e. such that $d\nu = f\, d|\nu|$.
Then $\bar{f}\, d\nu = \bar{f} f\, d|\nu| = |f|^2 d|\nu| = d|\nu|$.

$\mu_3 \leq \mu_1$: Suppose $\|f\| \leq 1$. Since f is bounded, $\exists$ simple functions that converge to f uniformly. Hence if $\varepsilon > 0$, then $\exists$ $\phi = \sum_{j=1}^{N} c_j \chi_{E_j}$ such that

$$\sup_{x \in X} |f(x) - \phi(x)| \leq \varepsilon. \quad \text{Consequently}$$

$|c_j| \leq 1 + \varepsilon$ since $c_j = \phi(x)$ for some x,

and $\left| \int f \, d\nu \right| \leq \left| \int \phi \, d\nu \right| + \varepsilon \|\nu\|$.

Exercise

Let ν be a complex measure. Show

$$E \in \mathcal{M} \text{ is a } \nu\text{-null set} \iff |\nu|(E) = 0.$$

Notation

For a complex measure ν, we say a property holds ν-a.e. if it holds except on a null set for ν. Hence ν-a.e. is the same as $|\nu|$-a.e.

Exercise

Let

$$M_b(X) = \{\nu : \nu \text{ is a complex measure on } (X, \mathcal{M})\}.$$

a. Show that $M_b(X)$ is a (complex) vector space, and that

$$\|\nu\| = |\nu|(X)$$

is a norm on $M_b(X)$.

b. Show that $M_b(X)$ is a Banach space, i.e., it is complete w.r.t. the norm $\|\cdot\|$ (this is not trivial, don't worry if you don't get it completely).

Hint: Suppose $\{\nu_n\}_{n\in\mathbb{N}}$ is Cauchy in $M_b(X)$.

Fix $E \in \mathcal{M}$. Show $\{\nu_n(E)\}_{n\in\mathbb{N}}$ is a Cauchy sequence of complex scalars. Define $\nu(E) = \lim_{n\to\infty} \nu_n(E)$.

Show ν is a complex measure. Hard part:

Show $\|\nu - \nu_n\| \to 0$.

<u>Exercise</u>

Let μ be a positive measure on $(X, \mathcal{M})$. Show

that if $f \in L^1(\mu)$ & $d\nu = f\,d\mu$, then

$\|\nu\| = \|f\|_1 = \int |f|\,d\mu$.

Consequently, $f \longmapsto f\,d\mu$ is an

isometric embedding of $L^1(\mu)$ into $M_b(X)$.

Exercises

1. Show that if $f: X \to \mathbb{C}$ is (m), then $\exists$ simple functions ϕ_k such that $\phi_k(x) \to f(x)$ pointwise, and $|\phi_k(x)| \le |f(x)|$ $\forall x \in X$, $k \in \mathbb{N}$.

2. Let μ be a positive measure on $(X, \mathcal{M})$, & fix $g \in L^1(\mu)$. Define $d\nu = g\, d\mu$. Show that if $f \in L^1(\nu)$, then $\int f\, d\nu = \int fg\, d\mu$.

3. Show that if ν is a complex measure on $(X, \mathcal{M})$ and $\nu(X) = |\nu|(X)$, then $\nu = |\nu|$.
   Hint: Consider the polar form of ν.

4. Let ν be a complex measure & μ a positive measure on $(X, \mathcal{M})$. Show that

$$\nu \ll \mu \iff |\nu| \ll \mu.$$

5. Define the complex conjugate of a complex measure ν by

$$\bar{\nu}(E) = \overline{\nu(E)}, \qquad E \in \mathcal{M}.$$

Show that $\bar{\nu}$ is a complex measure, and

$$\int f\, d\bar{\nu} = \overline{\int \bar{f}\, d\nu} \quad \text{for } f \in L^1(\nu).$$

6. Given a complex measure ν on $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, find an explicit description of $|\nu|$.

7. Show that
$$\nu \longmapsto (\nu\{k\})_{k\in\mathbb{N}}$$
is an isometric isomorphism of $M_b(\mathbb{N})$ onto $\ell^1(\mathbb{N})$. Thus $M_b(\mathbb{N}) \cong \ell^1(\mathbb{N})$.

Chapter 5
Banach and Hilbert Spaces

We assume that the reader is familiar with vector spaces (which are also called *linear spaces*). The scalar field associated with the vector spaces in this volume will always be either the real line $\mathbb{R}$ or the complex plane $\mathbb{C}$.

For this chapter, we will use the symbol $\mathbb{F}$ to denote a choice of either $\mathbb{R}$ or $\mathbb{C}$.

5.1 Definition and Examples of Banach Spaces

A norm on a vector space quantifies the idea of the "size" of a vector.

Definition 5.1.1. A vector space X is called a *normed linear space* if for each $x \in X$ there is a (finite) real number $\|x\|$, called the *norm* of x, such that:

(a) $\|x\| \ge 0$ for all $x \in X$,

(b) $\|x\| = 0$ if and only if $x = 0$,

(c) $\|cx\| = |c|\,\|x\|$ for all $x \in X$ and scalars c, and

(d) the *Triangle Inequality*: $\|x+y\| \le \|x\| + \|y\|$ holds for all $x, y \in X$.

Given a norm $\|\cdot\|$, we refer to the number $\|x-y\|$ as the *distance* between the vectors x and y. We call

$$B_r(x) \;=\; \bigl\{y \in X : \|x-y\| < r\bigr\}$$

the *open ball in X centered at x with radius r*. $\diamondsuit$

On occasion we will also deal with *seminorms* which, by definition, must satisfy properties (a), (c), and (d) of Definition 5.1.1, but need not satisfy property (b). For example, if we define $\|x\| = |x_1|$ for $x = (x_1, x_2) \in \mathbb{F}^2$, then $\|\cdot\|$ is a seminorm on $\mathbb{F}^2$, but it is not a norm on $\mathbb{F}^2$.

Given a normed space X, it is usually clear from context what norm we mean to use on X. Therefore, we usually just write $\|\cdot\|$ to denote the norm on X. However, when there is a possibility of confusion we may write $\|\cdot\|_X$ to specify that this norm is the norm on X, or we may write "the space $(X, \|\cdot\|)$" to emphasize that $\|\cdot\|$ represents the norm on X.

In addition to $\|\cdot\|$, we sometimes use symbols such as $|\cdot|$, $|||\cdot|||$, or $\rho(\cdot)$ to denote a norm or seminorm.

Definition 5.1.2. Let X be a normed linear space.

(a) A sequence of vectors $\{x_n\}_{n\in\mathbb{N}}$ in X *converges* to $x \in X$ if we have $\lim_{n\to\infty} \|x - x_n\| = 0$, i.e., if

$$\forall\, \varepsilon > 0, \quad \exists\, N > 0, \quad \forall\, n \geq N, \quad \|x - x_n\| < \varepsilon.$$

In this case, we write either $x_n \to x$ or $\lim_{n\to\infty} x_n = x$.

(b) A sequence of vectors $\{x_n\}_{n\in\mathbb{N}}$ in X is a *Cauchy sequence* in X if we have $\lim_{m,n\to\infty} \|x_m - x_n\| = 0$. More precisely, this means that

$$\forall\, \varepsilon > 0, \quad \exists\, N > 0, \quad \forall\, m, n \geq N, \quad \|x_m - x_n\| < \varepsilon. \qquad \diamondsuit$$

Every convergent sequence in a normed space is a Cauchy sequence (see Exercise 5.1.12). However, the converse is not true in general (consider Exercise 5.3.4).

Definition 5.1.3. We say that a normed space X is *complete* if it is the case that every Cauchy sequence in X is a convergent sequence. A complete normed linear space is called a *Banach space*. $\diamondsuit$

Sometimes we need to be explicit about which scalar field is associated with a Banach space X. We say that X is a *real Banach space* if it is a Banach space over the real field (i.e., $\mathbb{F} = \mathbb{R}$), and similarly it is a *complex Banach space* if $\mathbb{F} = \mathbb{C}$.

All Cauchy and convergent sequences in a normed space are bounded above in the following sense (see Exercise 5.1.12).

Definition 5.1.4. A sequence $\{x_n\}_{n\in\mathbb{N}}$ in a Banach space X is:

(a) *bounded below* if $\inf \|x_n\| > 0$,

(b) *bounded above* if $\sup \|x_n\| < \infty$,

(c) *normalized* if $\|x_n\| = 1$ for all n. $\diamondsuit$

The simplest example of a Banach space is the scalar field $\mathbb{F}$, where the norm on $\mathbb{F}$ is the absolute value. We will take as given the fact that $\mathbb{F}$ is complete with respect to absolute value. There are infinitely many norms on $\mathbb{F}$, but they are all positive scalar multiples of absolute value (Exercise 5.1.11), so we always assume that the norm on $\mathbb{F}$ is the absolute value.

Example 5.1.5. The next simplest example of a Banach space is $\mathbb{F}^d$, the set of all d-tuples of scalars, where d is a positive integer. There are many choices of norms for $\mathbb{F}^d$. Writing a generic vector $v \in \mathbb{F}^d$ as $v = (v_1, \dots, v_d)$, each of the following defines a norm on $\mathbb{F}^d$, and $\mathbb{F}^d$ is complete with respect to each of these norms:

$$|v|_p = \begin{cases} \bigl(|v_1|^p + \cdots + |v_d|^p\bigr)^{1/p}, & 1 \le p < \infty, \\ \max\{|v_1|, \dots, |v_d|\}, & p = \infty. \end{cases} \tag{5.1}$$

The *Euclidean norm* $|v|$ of a vector $v \in \mathbb{F}^d$ is the norm corresponding to the choice $p = 2$, i.e.,

$$|v| = |v|_2 = \sqrt{|v_1|^2 + \cdots + |v_d|^2}.$$

This particular norm has some extra algebraic properties that we will discuss further in Section 5.5. The fact that $|\cdot|_p$ is a norm on $\mathbb{F}^d$ is immediate for the cases $p = 1$ and $p = \infty$. For $1 < p < \infty$ the only norm property that is not obvious is the Triangle Inequality, and this can be shown by using exactly the same argument that we use later to prove Theorem 5.2.3. The proof that $\mathbb{F}^d$ is complete with respect to these norms is assigned as Exercise 5.1.13. $\diamondsuit$

The following example shows that we can construct norms on any finite-dimensional vector space.

Example 5.1.6. Let V be a finite-dimensional vector space. Then there exists a finite set of vectors $\mathcal{B} = \{x_1, \dots, x_d\}$ that is a basis for V, i.e., $\mathcal{B}$ spans V and $\mathcal{B}$ is linearly independent. Each $x \in V$ can be written as $x = \sum_{k=1}^{d} c_k(x)\, x_k$ for a unique choice of scalars $c_k(x)$. Given $1 \le p \le \infty$, if we set

$$\|x\|_p = \begin{cases} \bigl(|c_1(x)|^p + \cdots + |c_d(x)|\bigr)^{1/p}, & 1 \le p < \infty, \\ \max\bigl\{|c_1(x)|, \dots, |c_d(x)|\bigr\}, & p = \infty, \end{cases}$$

then $\|\cdot\|_p$ is a norm on V and V is complete with respect to this norm (see Exercise 5.1.14 and Theorem 5.2.4). $\diamondsuit$

The preceding examples illustrate the fact that there can be many norms on any given space.

Definition 5.1.7. Suppose that X is a normed linear space with respect to a norm $\|\cdot\|$ and also with respect to another norm $|||\cdot|||$. These norms are *equivalent* if there exist constants C_1, $C_2 > 0$ such that

$$\forall\, x \in X, \quad C_1\,\|x\| \le |||x||| \le C_2\,\|x\|. \qquad \diamondsuit$$

Note that if $\|\cdot\|$ and $|||\cdot|||$ are equivalent norms on X, then they define the same convergence criterion in the sense that

$$\lim_{n\to\infty} \|x - x_n\| = 0 \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} |||x - x_n||| = 0.$$

Any two of the norms $|\cdot|_p$ on $\mathbb{F}^d$ are equivalent (see Exercise 5.1.13). This is a special case of the following theorem, whose proof can be found in [Con90, Thm. 3.1].

Theorem 5.1.8. *If V is a finite-dimensional vector space, then any two norms on V are equivalent.* $\diamondsuit$

Now we give some examples of infinite-dimensional Banach spaces. We have not yet presented the tools that are needed to prove that these are Banach spaces, but will do so in Section 5.2 (see Theorem 5.2.3).

Example 5.1.9. In this example we consider some vector spaces whose elements are infinite sequences of scalars $x = (x_k) = (x_k)_{k\in\mathbb{N}}$ indexed by the natural numbers.

(a) Given $1 \le p < \infty$, we define ℓ^p to be the space of all infinite sequences of scalars that are p-summable, i.e.,

$$\ell^p \;=\; \ell^p(\mathbb{N}) \;=\; \Bigl\{x = (x_k) : \sum_k |x_k|^p < \infty\Bigr\}. \tag{5.2}$$

We will see in Theorem 5.2.4 that this is a Banach space with respect to the norm

$$\|x\|_{\ell^p} \;=\; \|(x_k)\|_{\ell^p} \;=\; \biggl(\sum_k |x_k|^p\biggr)^{1/p}. \tag{5.3}$$

(b) We define ℓ^∞ to be the space of all bounded infinite sequences of scalars:

$$\ell^\infty \;=\; \ell^\infty(\mathbb{N}) \;=\; \bigl\{x = (x_k) : (x_k) \text{ is a bounded sequence}\bigr\}.$$

This is a Banach space with respect to the *sup-norm*

$$\|x\|_{\ell^\infty} \;=\; \|(x_k)\|_{\ell^\infty} \;=\; \sup_k |x_k|. \qquad \diamondsuit$$

We can obtain analogous Banach spaces $\ell^p(I)$ by replacing the index set $\mathbb{N}$ by another countable index set I. For example, $\ell^p\bigl(\{1,\dots,d\}\bigr)$ is simply the vector space $\mathbb{F}^d$ with the norm defined in equation (5.3). Another example is $\ell^p(\mathbb{Z})$, whose elements are bi-infinite sequences of scalars that are p-summable (or bounded if $p = \infty$). Exercise 5.2.10 shows how to define $\ell^p(I)$ for uncountable index sets I.

Example 5.1.10. Now we consider some vector spaces whose elements are scalar-valued functions on a measurable domain $E \subseteq \mathbb{R}$.

(a) Given $1 \le p < \infty$, we define $L^p(E)$ to be the space of all measurable functions on E that are p-integrable, i.e.,

$$L^p(E) \;=\; \biggl\{ f\colon E \to \mathbb{F} \;:\; \int_E |f(t)|^p \, dt < \infty \biggr\}.$$

Then

$$\|f\|_{L^p} \;=\; \biggl(\int_E |f(t)|^p \, dt \biggr)^{1/p}$$

defines a seminorm on $L^p(E)$. It is not a norm because any function f such that $f = 0$ a.e. will satisfy $\|f\|_{L^p} = 0$ even though f need not be identically zero. Therefore, we usually "identify" any two functions that differ only on a set of measure zero. Regarding any two such functions as defining the same element of $L^p(E)$, we have that $\|\cdot\|_{L^p}$ is a norm. In a more technical language, the elements of $L^p(E)$ are actually equivalence classes of functions that are equal a.e., and the norm of such an equivalence class is the norm of any representative of the class.

(b) In the opening section on General Notation, we declared that a measurable function f is *essentially bounded* if there exists a real number M such that $|f(t)| \le M$ almost everywhere. We define $L^\infty(E)$ to be the space of all essentially bounded functions on E:

$$L^\infty(E) \;=\; \bigl\{ f\colon E \to \mathbb{F} \;:\; f \text{ is essentially bounded on } E \bigr\}.$$

Again identifying functions that are equal a.e., $L^\infty(E)$ is a Banach space with respect to the norm

$$\|f\|_{L^\infty} \;=\; \operatorname*{ess\,sup}_{t \in E} |f(t)| \;=\; \inf\bigl\{ M \ge 0 : |f(t)| \le M \text{ a.e.} \bigr\}. \qquad \diamondsuit \tag{5.4}$$

We will see several more examples of normed spaces and Banach spaces in Section 5.3.

Exercises

5.1.11. Show that if $\|\cdot\|$ is a norm on the scalar field $\mathbb{F}$, then there exists a positive number $\lambda > 0$ such that $\|x\| = \lambda\,|x|$, where $|x|$ is the absolute value of x.

5.1.12. Given a normed linear space X, prove the following facts.

(a) Every convergent sequence in X is Cauchy, and the limit of a convergent sequence is unique.

(b) Every Cauchy sequence in X is bounded.

(c) Reverse Triangle Inequality: $\bigl|\|x\| - \|y\|\bigr| \le \|x - y\|$ for all $x, y \in X$.

(d) Continuity of the norm: $x_n \to x \implies \|x_n\| \to \|x\|$.

(e) Continuity of vector addition:

$$x_n \to x \text{ and } y_n \to y \implies x_n + y_n \to x + y.$$

(f) Continuity of scalar multiplication:

$$x_n \to x \text{ and } c_n \to c \implies c_n x_n \to cx.$$

(g) Convexity of open balls: If $x, y \in B_r(z)$ then

$$\theta x + (1-\theta)y \in B_r(z) \quad \text{for all } 0 \le \theta \le 1.$$

5.1.13. (a) Assuming that the functions $|\cdot|_p$ on $\mathbb{F}^d$ given in equation (5.1) are norms on $\mathbb{F}^d$, show that any two of these norms are equivalent.

(b) Assuming that $\mathbb{F}$ is complete with respect to absolute value, show that $\mathbb{F}^d$ is complete with respect to any one of the norms $|\cdot|_p$.

5.1.14. Let V be a finite-dimensional vector space. Prove that the function $\|\cdot\|_1$ on V defined in Example 5.1.6 is a norm, and that V is complete with respect to this norm.

5.1.15. Show that if vectors x_n in a normed space X satisfy $\|x_{n+1} - x_n\| < 2^{-n}$ for every $n \in \mathbb{N}$, then $\{x_n\}$ is a Cauchy sequence in X.

5.1.16. Let $\{x_n\}$ be a sequence in a normed space X, and let $x \in X$ be fixed. Suppose that every subsequence $\{y_n\}$ of $\{x_n\}$ has a subsequence $\{z_n\}$ of $\{y_n\}$ such that $z_n \to x$. Show that $x_n \to x$.

5.1.17. Let X be a complex Banach space. Let $X_{\mathbb{R}} = X$ as a set, but consider $X_{\mathbb{R}}$ as a vector space over the real field. That is, vector addition in $X_{\mathbb{R}}$ is defined just as in X, but scalar multiplication in $X_{\mathbb{R}}$ is restricted to multiplication by real scalars. Let $\|\cdot\|_{\mathbb{R}} = \|\cdot\|$, and show that $(X_{\mathbb{R}}, \|\cdot\|_{\mathbb{R}})$ is a real Banach space.

5.1.18. We say that a set X (not necessarily a vector space) is a *metric space* if for each $x, y \in X$ there exists a real number $\mathrm{d}(x,y)$ such that for all x, y, $z \in X$ we have:

i. i. $\mathrm{d}(x,y) \ge 0$,
ii. $\mathrm{d}(x,y) = 0$ if and only if $x = y$,
iii. $\mathrm{d}(x,y) = \mathrm{d}(y,x)$, and
iv. the *Triangle Inequality*: $\mathrm{d}(x,y) \le \mathrm{d}(x,z) + \mathrm{d}(y,z)$.

In this case we call d a *metric* on X, and we refer to $\mathrm{d}(x,y)$ as the *distance* between x and y.

(a) Show that if X is a normed space, then $\mathrm{d}(x,y) = \|x-y\|$ is a metric on X.

(b) Make a definition of convergent and Cauchy sequences in a metric space, and show that every convergent sequence is Cauchy.

(c) Define a metric space to be *complete* if every Cauchy sequence in X converges to an element of X. Let $X = \mathbb{Q}$ and set $\mathrm{d}(x,y) = |x-y|$, the ordinary absolute value of the difference of x and y. Show that d is a metric on $\mathbb{Q}$, but $\mathbb{Q}$ is incomplete with respect to this metric.

5.1.19. Fix $0 < p < 1$, and define ℓ^p by equation (5.2) and $\|\cdot\|_{\ell^p}$ by equation (5.3).

(a) Show that $\|\cdot\|_{\ell^p}$ fails the Triangle Inequality and hence is not a norm on ℓ^p.

(b) Show that $\|x+y\|_{\ell^p}^p \le \|x\|_{\ell^p}^p + \|y\|_{\ell^p}^p$, and use this to show that ℓ^p is a vector space and $\mathrm{d}(x,y) = \|x-y\|_{\ell^p}^p$ is a metric on ℓ^p.

(c) Let $B = \{x \in \ell^p : \mathrm{d}(x,0) < 1\}$ be the "open unit ball" with respect to the metric d, and show that B is not convex. Use this to show that there is no norm $\|\cdot\|$ on ℓ^p such that $\mathrm{d}(x,y) = \|x-y\|$.

5.2 Hölder's and Minkowski's Inequalities

In this section we will prove that ℓ^p is a Banach space. The proof of the completeness of $L^p(E)$ is similar in spirit, but is somewhat more technical as various notions from measure theory are required, and will be omitted.

For $p = 1$ and $p = \infty$ it is easy to see that the function $\|\cdot\|_{\ell^p}$ defined in Example 5.1.9 satisfies the Triangle Inequality. It is not nearly so obvious that the Triangle Inequality holds when $1 < p < \infty$. The next theorem gives a fundamental inequality on the norm of a product of two functions or sequences, and we will use this inequality to prove the Triangle Inequality on ℓ^p for $1 < p < \infty$. For this result, the following notion of the dual index will be useful.

Notation 5.2.1. Given $1 \le p \le \infty$, its *dual index* is the number $1 \le p' \le \infty$ satisfying

$$\frac{1}{p} + \frac{1}{p'} \;=\; 1,$$

where we use the conventions that $1/0 = \infty$ and $1/\infty = 0$. ♢

For example, $1' = \infty,\ 2' = 2,\ 3' = 4/3$, and $\infty' = 1$. Explicitly,

$$p' = \frac{p}{p-1},$$

and we have $(p')' = p$.

Theorem 5.2.2 (Hölder's Inequality). *Fix* $1 \le p \le \infty$.

(a) *If* $x = (x_k) \in \ell^p$ *and* $y = (y_k) \in \ell^{p'}$, *then* $(x_k y_k) \in \ell^1$ *and*

$$\|(x_k y_k)\|_{\ell^1} \ \le\ \|(x_k)\|_{\ell^p}\ \|(y_k)\|_{\ell^{p'}}.$$

For $1 < p < \infty$ *this is equivalent to the statement*

$$\sum_k |x_k y_k| \ \le\ \biggl(\sum_k |x_k|^p\biggr)^{1/p} \biggl(\sum_k |y_k|^{p'}\biggr)^{1/p'}.$$

(b) *If* $f \in L^p(E)$ *and* $g \in L^{p'}(E)$, *then* $fg \in L^1(E)$ *and*

$$\|fg\|_{L^1} \ \le\ \|f\|_{L^p}\ \|g\|_{L^{p'}}.$$

For $1 < p < \infty$ *this is equivalent to the statement*

$$\int_E |f(t)\,g(t)|\,dt \ \le\ \biggl(\int_E |f(t)|^p\,dt\biggr)^{1/p} \biggl(\int_E |g(t)|^{p'}\,dt\biggr)^{1/p'}.$$

Proof. We will concentrate on the ℓ^p spaces, as the proof for $L^p(E)$ is similar. The cases $p = 1$ and $p = \infty$ are straightforward, so assume that $1 < p < \infty$.

Suppose that $x = (x_k) \in \ell^p$ and $y = (y_k) \in \ell^{p'}$ satisfy $\|x\|_{\ell^p} = 1 = \|y\|_{\ell^{p'}}$. By Exercise 5.2.5, we have the inequality

$$|x_k y_k| \ \le\ \frac{|x_k|^p}{p} + \frac{|y_k|^{p'}}{p'}, \qquad k \in \mathbb{N}.$$

Consequently,

$$\begin{aligned}
\|(x_k y_k)\|_{\ell^1} \ =\ \sum_k |x_k y_k| \ &\le\ \sum_k \biggl(\frac{|x_k|^p}{p} + \frac{|y_k|^{p'}}{p'}\biggr) \\
&=\ \frac{\|(x_k)\|_{\ell^p}^p}{p} + \frac{\|(y_k)\|_{\ell^{p'}}^{p'}}{p'} \\
&=\ \frac{1}{p} + \frac{1}{p'} \ =\ 1. \qquad (5.5)
\end{aligned}$$

Given arbitrary $x \in \ell^p$ and $y \in \ell^{p'}$, let $s = \|x\|_{\ell^p}$ and $t = \|y\|_{\ell^{p'}}$. Then $x/s = (x_k/s) \in \ell^p$ and $y/t = (y_k/t) \in \ell^{p'}$ are each unit vectors, i.e., $\|x/s\|_{\ell^p} = 1 = \|y/t\|_{\ell^{p'}}$. The result then follows by applying equation (5.5) to x/s and y/t. $\square$

Note that if $p = 2$ then the dual index is $p' = 2$ as well. Therefore, we have the following special cases of Hölder's inequality, usually referred to as the *Cauchy–Schwarz* or *Cauchy–Bunyakovski–Schwarz* inequalities:

$$\|(x_k y_k)\|_{\ell^1} \ \le\ \|(x_k)\|_{\ell^2}\ \|(y_k)\|_{\ell^2} \quad \text{and} \quad \|fg\|_{L^1} \ \le\ \|f\|_{L^2}\ \|g\|_{L^2}. \qquad (5.6)$$

ℓ^2 and $L^2(E)$ are specific examples of *Hilbert spaces*, which are discussed in more detail in Section 5.5. In particular, Theorem 5.5.5 will present a generalization of the Cauchy–Bunyakovski–Schwarz inequalities that is valid in any Hilbert space.

Now we show that $\|\cdot\|_{\ell^p}$ is a norm on ℓ^p. Although we will not prove it, a similar argument shows that $\|\cdot\|_{L^p}$ is a norm on $L^p(E)$. The Triangle Inequality on ℓ^p or L^p is often called *Minkowski's Inequality.*

Theorem 5.2.3 (Minkowski's Inequality). *If $1 \le p \le \infty$, then $\|\cdot\|_{\ell^p}$ is a norm on ℓ^p and $\|\cdot\|_{L^p}$ is a norm on $L^p(E)$.*

Proof. The cases $p = 1$ and $p = \infty$ are straightforward, so consider $1 < p < \infty$. All of the properties of a norm are clear except for the Triangle Inequality. To prove this, fix $x = (x_k)$ and $y = (y_k)$ in ℓ^p. Then we have

$$\begin{aligned}
\|x+y\|_{\ell^p}^p &= \sum_k |x_k + y_k|^{p-1}\,|x_k + y_k| \\
&\le \sum_k |x_k + y_k|^{p-1}\,|x_k| \;+\; \sum_k |x_k + y_k|^{p-1}\,|y_k| \\
&\le \biggl(\sum_k \bigl(|x_k + y_k|^{p-1}\bigr)^{p'}\biggr)^{1/p'} \biggl(\sum_k |x_k|^p\biggr)^{1/p} \\
&\qquad + \biggl(\sum_k \bigl(|x_k + y_k|^{p-1}\bigr)^{p'}\biggr)^{1/p'} \biggl(\sum_k |y_k|^p\biggr)^{1/p} \\
&= \biggl(\sum_k |x_k + y_k|^p\biggr)^{(p-1)/p} \|x\|_{\ell^p} \;+\; \biggl(\sum_k |x_k + y_k|^p\biggr)^{(p-1)/p} \|y\|_{\ell^p} \\
&= \|x+y\|_{\ell^p}^{p-1}\,\|x\|_{\ell^p} + \|x+y\|_{\ell^p}^{p-1}\,\|y\|_{\ell^p},
\end{aligned}$$

where we have applied Hölder's Inequality with exponents p' and p, and used the fact that $p' = p/(p-1)$. Dividing both sides by $\|x+y\|_{\ell^p}^{p-1}$, we obtain $\|x+y\|_{\ell^p} \le \|x\|_{\ell^p} + \|y\|_{\ell^p}$. $\square$

Finally, we show that ℓ^p is complete and therefore is a Banach space. The argument for $L^p(E)$ is similar in spirit but is technically more complicated, and will be omitted (see [Fol99] or [WZ77]).

Theorem 5.2.4. *If $1 \le p \le \infty$, then ℓ^p is a Banach space with respect to the norm $\|\cdot\|_{\ell^p}$, and $L^p(E)$ is a Banach space with respect to the norm $\|\cdot\|_{L^p}$.*

Proof. Fix $1 \le p < \infty$ (the case $p = \infty$ is similar), and suppose that $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in ℓ^p. Each x_n is a vector in ℓ^p, so let us write the components of x_n as

$$x_n \;=\; \bigl(x_n(1), x_n(2), \dots\bigr).$$

Then for each fixed index $k \in \mathbb{N}$ we have $|x_m(k) - x_n(k)| \le \|x_m - x_n\|_{\ell^p}$. Hence $\big(x_n(k)\big)_{n\in\mathbb{N}}$ is a Cauchy sequence of scalars, and therefore must converge since $\mathbb{F}$ is complete. Define $x(k) = \lim_{n\to\infty} x_n(k)$. Then x_n *converges componentwise* to $x = \big(x(1), x(2), \dots\big)$, i.e.,

$$\forall\, k \in \mathbb{N}, \quad x(k) = \lim_{n\to\infty} x_n(k).$$

We need to show that x_n converges to x *in the norm of* ℓ^p.

Choose any $\varepsilon > 0$. Then, by the definition of a Cauchy sequence, there exists an N such that $\|x_m - x_n\|_{\ell^p} < \varepsilon$ for all $m, n > N$. Fix any particular $n > N$. For each $M > 0$ we have

$$\sum_{k=1}^{M} |x(k) - x_n(k)|^p \;=\; \lim_{m\to\infty} \sum_{k=1}^{M} |x_m(k) - x_n(k)|^p \;\le\; \lim_{m\to\infty} \|x_m - x_n\|_{\ell^p}^p \;\le\; \varepsilon^p.$$

Since this is true for every M, we conclude that

$$\|x - x_n\|_{\ell^p}^p \;=\; \sum_{k=1}^{\infty} |x(k) - x_n(k)|^p \;=\; \lim_{M\to\infty} \sum_{k=1}^{M} |x(k) - x_n(k)|^p \;\le\; \varepsilon^p. \tag{5.7}$$

Consequently,

$$\|x\|_{\ell^p} \;=\; \|x - x_n + x_n\|_{\ell^p} \;\le\; \|x - x_n\|_{\ell^p} + \|x_n\|_{\ell^p} \;<\; \infty,$$

so $x \in \ell^p$. Further, since equation (5.7) holds for all $n > N$, we have that $\lim_{n\to\infty} \|x - x_n\|_{\ell^p} = 0$, i.e., $x_n \to x$ in ℓ^p. Therefore ℓ^p is complete. $\square$

Theorems 5.2.3 and 5.2.4 carry over with minimal changes to show that $\mathbb{F}^d$ is a Banach space with respect to any of the norms $|\cdot|_p$ defined in Example 5.1.5.

Exercises

5.2.5. (a) Show that if $0 < \theta < 1$, then $t^\theta \le \theta t + (1-\theta)$ for $t \ge 0$, with equality if and only if $t = 1$.

(b) Suppose that $1 < p < \infty$ and $a, b \ge 0$. Apply part (a) with $t = a^p b^{-p'}$ and $\theta = 1/p$ to show that $ab \le a^p/p + b^{p'}/p'$, with equality if and only if $b = a^{p-1}$.

5.2.6. Show that equality holds in Hölder's Inequality for sequences (part (a) of Theorem 5.2.2) if and only if there exist scalars α, β, not both zero, such that $\alpha\,|x_k|^p = \beta\,|y_k|^{p'}$ for each $k \in I$.

5.2.7. Show that if $1 \le p < q \le \infty$, then $\ell^p \subsetneq \ell^q$ and $\|x\|_{\ell^q} \le \|x\|_{\ell^p}$ for all $x \in \ell^p$.

5.2.8. Let $E \subseteq \mathbb{R}$ be measurable with $|E| < \infty$. Show that if $1 \le p < q \le \infty$, then $L^q(E) \subsetneq L^p(E)$ and $\|f\|_{L^p} \le |E|^{\frac{1}{p}-\frac{1}{q}}\, \|f\|_{L^q}$ for all $f \in L^p(E)$.

5.2.9. Show that if $x \in \ell^q$ for some finite q, then $\|x\|_{\ell^p} \to \|x\|_{\ell^\infty}$ as $p \to \infty$, but this can fail if $x \notin \ell^q$ for any finite q.

5.2.10. Given an arbitrary index set I, define $\ell^\infty(I)$ to be the space of all bounded sequences $x = (x_i)_{i\in I}$ indexed by I, with $\|x\|_\infty = \sup_{i\in I} |x_i|$. For $1 \le p < \infty$ let $\ell^p(I)$ consist of all sequences $x = (x_i)_{i\in I}$ *with at most countably many nonzero components* such that $\|x\|_{\ell^p}^p = \sum |x_i|^p < \infty$. Show that each of these spaces $\ell^p(I)$ is a Banach space with respect to $\|\cdot\|_{\ell^p}$. What is $\ell^p(I)$ if $I = \{1, \dots, d\}$?

5.2.11. Let X, Y be normed linear spaces. Given $1 \le p < \infty$, $x \in X$, and $y \in Y$, define $\|(x,y)\|_p = \bigl(\|x\|_X^p + \|y\|_Y^p\bigr)^{1/p}$ and $\|(x,y)\|_\infty = \max\{\|x\|_X, \|y\|_Y\}$.

(a) Prove that $\|\cdot\|_p$ is a norm on the Cartesian product $X \times Y$, and $\|\cdot\|_p$ and $\|\cdot\|_q$ are equivalent norms on $X \times Y$ for any $1 \le p, q \le \infty$.

(b) Show that if X and Y are Banach spaces, then $X \times Y$ is a Banach space with respect to $\|\cdot\|_p$.

5.3 Basic Properties of Banach Spaces

In this section we will give some definitions and facts that hold for normed spaces and Banach spaces.

Definition 5.3.1. Let X be a normed linear space.

(a) Recall from Definition 5.1.1 that if $x \in X$ and $r > 0$, then the *open ball in X centered at x with radius r* is $B_r(x) = \{y \in X : \|x-y\| < r\}$.

(b) A subset $U \subseteq X$ is *open* if for each $x \in U$ there exists an $r > 0$ such that $B_r(x) \subseteq U$.

(c) A subset $E \subseteq X$ is *closed* if $X \backslash E$ is open.

(d) Let $E \subseteq X$. Then $x \in X$ is a *limit point* of E if there exist $x_n \in E$ with all $x_n \ne x$ such that $x_n \to x$.

(e) The *closure* of a subset $E \subseteq X$ is the smallest closed set $\overline{E}$ that contains E, i.e., $\overline{E} = \cap \bigl\{F : F \text{ is closed and } E \subseteq F\bigr\}$.

(f) A subset $E \subseteq X$ is *dense* in X if $\overline{E} = X$. $\diamondsuit$

The following lemma gives useful equivalent reformulations of some of the notions defined above. Exercise 5.3.11 asks for proof of Lemma 5.3.2.

Lemma 5.3.2. *Given a Banach space X and given $E \subseteq X$, the following statements hold.*

(a) *E is closed if and only if it contains all of its limit points.*

(b) $\overline{E} = E \cup \{x \in X : x$ *is a limit point of* $E\}$. *Consequently, E is closed if and only if* $\overline{E} = E$.

(c) *E is dense in X if and only if every $x \in X$ is a limit point of E.* $\diamondsuit$

Once we have a space X in hand that we know is a Banach space with respect to a norm $\|\cdot\|$, we often need to know if a given subspace S of X is also a Banach space with respect to this same norm. The next lemma (whose proof is Exercise 5.3.12) gives a convenient characterization of those subspaces that are complete with respect to the norm on X.

Lemma 5.3.3. *Let S be a subspace of a Banach space X. Then S is a Banach space with respect to the norm on X if and only if S is a closed subset of X.* $\diamondsuit$

We can use Lemma 5.3.3 to give some additional examples of normed spaces that are or are not Banach spaces.

Example 5.3.4. (a) Define

$$c \;=\; c(\mathbb{N}) \;=\; \bigl\{a = (a_k) : \lim_{k\to\infty} a_k \text{ exists}\bigr\},$$
$$c_0 \;=\; c_0(\mathbb{N}) \;=\; \bigl\{a = (a_k) : \lim_{k\to\infty} a_k = 0\bigr\}.$$

Exercise 5.3.14 asks for a proof that c and c_0 are closed subspaces of ℓ^∞. Consequently, Lemma 5.3.3 implies that c and c_0 are Banach spaces with respect to the norm $\|\cdot\|_{\ell^\infty}$.

(b) Now consider

$$c_{00} \;=\; c_{00}(\mathbb{N}) \;=\; \bigl\{a = (a_k) : \text{only finitely many } a_k \text{ are nonzero}\bigr\}.$$

Even though the elements of c_{00} are infinite sequences, since only finitely many components are nonzero, they are often called "finite sequences." Note that c_{00} is a subspace of c_0, c, and ℓ^∞. By Exercise 5.3.14, c_{00} is a proper, dense subset of c_0. Consequently, $\overline{c_{00}} = c_0 \neq c_{00}$, so c_{00} is not a closed subspace of c_0 (or of c or ℓ^∞), and therefore it is not a Banach space with respect to $\|\cdot\|_{\ell^\infty}$.

A specific example of a Cauchy sequence in c_{00} that does not have a limit in c_{00} is the sequence $\{x_n\}_{n\in\mathbb{N}}$, where x_n is the vector in ℓ^p given by

$$x_n \;=\; \bigl(1, \tfrac{1}{2}, \dots, \tfrac{1}{n}, 0, 0, \dots\bigr).$$

The vectors x_n do converge in ℓ^∞ norm to the vector $x = (1, \frac{1}{2}, \frac{1}{3}, \dots)$, but *$x$ does not belong to* c_{00}. While $\{x_n\}_{n\in\mathbb{N}}$ converges in c_0 and in ℓ^∞, it does not converge in c_{00}. $\diamondsuit$

Although c_{00} is not a Banach space with respect to the sup-norm, it is a proper, dense subspace of c_0, which is a Banach space. More generally, given any normed linear space X that is not complete, there exists a unique Banach space $\widetilde{X}$ such that X is a proper, dense subspace of $\widetilde{X}$ and the norm on $\widetilde{X}$ extends the norm on X. This space $\widetilde{X}$ is called the *completion* of X (see Exercise 5.3.18).

Following are some examples of normed spaces whose elements are continuous functions. More examples appear in the Exercises.

Example 5.3.5. (a) Define

$$C(\mathbb{R}) \;=\; \bigl\{f\colon \mathbb{R}\to\mathbb{F} \;:\; f \text{ is continuous on } \mathbb{R}\bigr\}.$$

While there is no convenient norm on $C(\mathbb{R})$, Exercise 5.3.15 shows that the subspace

$$C_b(\mathbb{R}) \;=\; \bigl\{f\in C(\mathbb{R}) : f \text{ is bounded}\bigr\}$$

is a Banach space with respect to the *sup-norm* or *uniform norm*

$$\|f\|_\infty \;=\; \sup_{t\in\mathbb{R}} |f(t)|.$$

(b) By Exercise 5.3.15,

$$C_0(\mathbb{R}) \;=\; \bigl\{f\in C_b(\mathbb{R}) : \lim_{|t|\to\infty} f(t)=0\bigr\}$$

is a closed subspace of $C_b(\mathbb{R})$ with respect to the uniform norm. Therefore $C_0(\mathbb{R})$ is a Banach space with respect to $\|\cdot\|_\infty$.

(c) Recall that a continuous function $f\colon\mathbb{R}\to\mathbb{F}$ has *compact support* if $f(t)=0$ for all t outside of some finite interval. The space

$$C_c(\mathbb{R}) \;=\; \bigl\{f\in C(\mathbb{R}) : f \text{ has compact support}\bigr\}$$

is a subspace of $C_0(\mathbb{R})$ and $C_b(\mathbb{R})$. However, $C_c(\mathbb{R})$ is a dense but proper subset of $C_0(\mathbb{R})$ with respect to the uniform norm (Exercise 5.3.15). Hence $C_c(\mathbb{R})$ is not closed and therefore is not complete with respect to $\|\cdot\|_\infty$.

(d) We can similarly define spaces of functions that are continuous on domains other than $\mathbb{R}$. The most important example for us is the space of functions that are continuous on a closed finite interval $[a,b]$, which we write as follows:

$$C[a,b] \;=\; \bigl\{f\colon [a,b]\to\mathbb{F} \;:\; f \text{ is continuous on } [a,b]\bigr\}.$$

Every continuous function on $[a,b]$ is bounded, and $C[a,b]$ is a Banach space with respect to the uniform norm. $\diamondsuit$

Remark 5.3.6. For a continuous function f, the supremum of $|f(x)|$ coincides with its essential supremum, i.e., $\|f\|_\infty = \|f\|_{L^\infty}$ (see Exercise 5.3.13). Therefore, if we identify a continuous function f with the equivalence class of all functions that equal f almost everywhere, then we can regard $C_b(E)$ as being a subspace of $L^\infty(E)$. In this sense of identification, $C_b(\mathbb{R})$ and $C_0(\mathbb{R})$ are closed subspaces of $L^\infty(E)$, but $C_c(\mathbb{R})$ is not a closed subspace with respect to $\|\cdot\|_{L^\infty}$. $\diamondsuit$

When considering a space of bounded continuous functions by itself, we usually assume that the norm on the space is the uniform norm. For example, Example 5.4.5 below will present the Weierstrass Approximation Theorem, which deals with $C[a,b]$ under the uniform norm. However, if we are thinking of a space of continuous functions as being a subspace of a larger space X, then we use the norm on X unless specifically stated otherwise. For example, the next lemma considers $C[a,b]$ as a subspace of $L^p[a,b]$, and we therefore implicitly assume in this result that the norm on $C[a,b]$ is $\|\cdot\|_{L^p}$. For proof of Lemma 5.3.7, we refer to [Fol99].

Lemma 5.3.7. (a) *$C[a,b]$ is dense in $L^p[a,b]$ for each $1 \le p < \infty$.*
(b) *$C_c(\mathbb{R})$ is dense in $L^p(\mathbb{R})$ for each $1 \le p < \infty$.* $\diamondsuit$

Next we prove directly that all finite-dimensional subspaces of a normed space are closed.

Theorem 5.3.8. *If M is a finite-dimensional subspace of a normed linear space X, then M is closed.*

Proof. Suppose that $x_n \in M$ and $x_n \to y \in X$. If $y \notin M$, define

$$M_1 \;=\; M \;+\; \operatorname{span}\{y\} \;=\; \{m + cy : m \in M,\, c \in \mathbb{F}\}.$$

Since $y \notin M$, every vector $x \in M_1$ has a unique representation of the form $x = m_x + c_x y$ with $m_x \in M$ and $c_x \in \mathbb{F}$. Therefore we can define

$$\|x\|_{M_1} \;=\; \|m_x\| + |c_x|,$$

where $\|\cdot\|$ is the norm on X. This forms a norm on M_1.

Since M_1 is finite dimensional, Theorem 5.1.8 implies that all norms on M_1 are equivalent. Hence there exist constants $A,\, B > 0$ such that

$$A\,\|x\| \;\le\; \|x\|_{M_1} \;\le\; B\,\|x\|, \qquad x \in M_1.$$

Now, since $x_n \in M$, the representation of $y - x_n$ as a vector in M_1 takes $m_z = -x_n$ and $c_z = 1$. Therefore

$$1 \;\le\; \|x_n\| + 1 \;=\; \|y - x_n\|_{M_1} \;\le\; B\,\|y - x_n\| \;\to\; 0 \quad \text{as } n \to \infty.$$

This is a contradiction, so we must have $y \in M$. Hence M is closed. $\square$

The next definition provides one way to distinguish “large” Banach spaces from “small” ones. Specifically, a “small” Banach space is one that contains a *countable* dense subset; we call such a space separable. For example, $\mathbb{Q}$ is a countable dense subset of $\mathbb{R}$, so $\mathbb{R}$ is separable, and likewise $\mathbb{F}^d$ is separable for each $d \in \mathbb{N}$.

Definition 5.3.9. A normed linear space X is *separable* if it contains a countable dense subset. $\diamondsuit$

Example 5.3.10. (a) If $1 \le p < \infty$ then ℓ^p is separable, but ℓ^∞ is not separable (see Example 5.4.7 and Exercise 5.4.11). It is likewise true that $L^p(E)$ is separable for $1 \le p < \infty$ and $L^\infty(E)$ is not separable (unless $|E| = 0$), although the proof of these facts requires some knowledge of measure theory and will be omitted.

(b) In Example 5.4.5 we will see that the Weierstrass Approximation Theorem implies that the Banach space $C[a,b]$ is separable (with respect to the uniform norm). This can then be used to show that $C_0(\mathbb{R})$ is separable as well (Exercise 5.4.10). $\diamondsuit$

Exercises

5.3.11. Prove Lemma 5.3.2.

5.3.12. Prove Lemma 5.3.3.

5.3.13. Show that if $f \in C_b(\mathbb{R})$, then $\operatorname{ess\,sup}_{t\in\mathbb{R}} |f(t)| = \sup_{t\in\mathbb{R}} |f(t)|$, and consequently $\|f\|_{L^\infty} = \|f\|_\infty$.

5.3.14. (a) Show that c and c_0 are closed subspaces of ℓ^∞.

(b) Show that c_{00} is a proper, dense subspace of c_0, and hence is not closed with respect to the norm $\|\cdot\|_{\ell^\infty}$.

(c) Let $\{\delta_n\}$ denote the sequence of standard basis vectors (see Example 5.4.6). Given $x = (x_n) \in c_0$, show that $x = \sum x_n \delta_n$, where the series converges with respect to the norm $\|\cdot\|_{\ell^\infty}$. Show further that the scalars x_n in this representation are unique.

5.3.15. (a) Show that $C_b(\mathbb{R})$ is a Banach space with respect to the uniform norm $\|\cdot\|_\infty$.

(b) Show that $C_0(\mathbb{R})$ is a closed subspace of $C_b(\mathbb{R})$.

(c) Show that $C_c(\mathbb{R})$ is a proper, dense subspace of $C_0(\mathbb{R})$, and hence is not closed with respect to $\|\cdot\|_\infty$.

(d) Let $C(\mathbb{T})$ be the set of all continuous functions $f \in C(\mathbb{R})$ that are 1-periodic, i.e., $f(t+1) = f(t)$ for every $t \in \mathbb{R}$. Show that $C(\mathbb{T})$ is a closed subspace of $C_b(\mathbb{R})$.

5.3.16. Let $C_b^m(\mathbb{R})$ be the space of all m-times differentiable functions on $\mathbb{R}$ each of whose derivatives is bounded and continuous, i.e.,

$$C_b^m(\mathbb{R}) \;=\; \bigl\{f \in C_b(\mathbb{R}) : f, f', \dots, f^{(m)} \in C_b(\mathbb{R})\bigr\}.$$

(a) Show that $C_b^m(\mathbb{R})$ is a Banach space with respect to the norm

$$\|f\|_{C_b^m} \;=\; \|f\|_\infty + \|f'\|_\infty + \cdots + \|f^{(m)}\|_\infty,$$

and

$$C_0^m(\mathbb{R}) \;=\; \bigl\{f \in C_0(\mathbb{R}) \;:\; f, f', \dots, f^{(m)} \in C_0(\mathbb{R})\bigr\}$$

is a subspace of $C_b^m(\mathbb{R})$ that is also a Banach space with respect to the same norm.

(b) Now we change the norm on $C_b^1(\mathbb{R})$. Show that $(C_b^1(\mathbb{R}), \|\cdot\|_\infty)$ is a normed space, but is not complete.

5.3.17. We say that a function $f\colon \mathbb{R} \to \mathbb{F}$ is *Hölder continuous* with exponent $\alpha > 0$ if there exists a constant $K > 0$ such that

$$\forall\, x, y \in \mathbb{R}, \quad |f(x) - f(y)| \;\le\; K\,|x-y|^\alpha.$$

A function that is Hölder continuous with exponent $\alpha = 1$ is said to be *Lipschitz*.

(a) Show that if f is Hölder continuous for some $\alpha > 1$, then f is constant.

(b) Show that if f is differentiable on $\mathbb{R}$ and f' is bounded, then f is Lipschitz. Find a function g that is Lipschitz but is not differentiable at every point.

(c) Given $0 < \alpha < 1$, define

$$C^\alpha(\mathbb{R}) \;=\; \bigl\{f \in C(\mathbb{R}) : f \text{ is Hölder continuous with exponent } \alpha\bigr\}.$$

Show that $C^\alpha(\mathbb{R})$ is a Banach space with respect to the norm

$$\|f\|_{C^\alpha} \;=\; |f(0)| \;+\; \sup_{x \ne y} \frac{f(x) - f(y)}{|x-y|^\alpha}.$$

5.3.18. Let X be a normed linear space that is not complete. Let $\mathcal{C}$ be the set of all Cauchy sequences in X, and define a relation $\sim$ on $\mathcal{C}$ by declaring that $\{x_n\}_{n\in\mathbb{N}} \sim \{y_n\}_{n\in\mathbb{N}}$ if $\lim_{n\to\infty} \|x_n - y_n\| = 0$.

(a) Show that $\sim$ is an equivalence relation on $\mathcal{C}$.

(b) Let $[x_n] = \bigl\{\{y_n\}_{n\in\mathbb{N}} : \{y_n\}_{n\in\mathbb{N}} \sim \{x_n\}_{n\in\mathbb{N}}\bigr\}$ denote the equivalence class of $\{x_n\}_{n\in\mathbb{N}}$ under the relation $\sim$. Let $\widetilde{X}$ be the set of all equivalence classes $[x_n]$. Define $\bigl\|[x_n]\bigr\|_{\widetilde{X}} = \lim_{n\to\infty} \|x_n\|$. Prove that $\|\cdot\|_{\widetilde{X}}$ is a well-defined norm on $\widetilde{X}$.

(c) Given $x \in X$, let $[x]$ denote the equivalence class of the Cauchy sequence $\{x, x, x, \dots\}$. Show that $T\colon x \mapsto [x]$ is an isometric map of X into $\widetilde{X}$, where *isometric* means that $\|Tx\|_{\widetilde{X}} = \|x\|$ for every x. Show also that $T(X)$ is a dense subspace of $\widetilde{X}$ (so, in the sense of identifying of X with $T(X)$, we can consider X to be a subspace of $\widetilde{X}$).

(d) Show that $\widetilde{X}$ is a Banach space with respect to $\|\cdot\|_{\widetilde{X}}$. We call $\widetilde{X}$ the *completion* of X.

(e) Prove that $\widetilde{X}$ is unique in the sense that if Y is a Banach space and $U\colon X \to Y$ is a linear isometry such that $U(X)$ is dense in Y, then there exists a linear isometric bijection $V\colon Y \to \widetilde{X}$.

5.4 Linear Combinations, Sequences, Series, and Complete Sets

In this section we review some definitions and concepts related to sequences, linear combinations, and infinite series in normed spaces.

Definition 5.4.1. Let S be a subset of a normed linear space X.

(a) S is *finitely linearly independent*, or simply *independent* for short, if for every choice of finitely many distinct vectors $x_1, \dots, x_N \in S$ and scalars $c_1, \dots, c_N \in \mathbb{F}$ we have

$$\sum_{n=1}^{N} c_n x_n \;=\; 0 \qquad \Longrightarrow \qquad c_1 = \cdots = c_N = 0.$$

(b) The *finite linear span*, or simply the *span*, of S is the set of all finite linear combinations of elements of S, i.e.,

$$\operatorname{span}(S) \;=\; \left\{ \sum_{n=1}^{N} c_n x_n : N \in \mathbb{N},\, x_1, \dots, x_N \in S,\, c_1, \dots, c_N \in \mathbb{F} \right\}.$$

If $S = \{x_n\}_{n\in\mathbb{N}}$ is countable, then we often write $\operatorname{span}\{x_n\}_{n\in\mathbb{N}}$ instead of $\operatorname{span}(\{x_n\}_{n\in\mathbb{N}})$:

$$\operatorname{span}\{x_n\}_{n\in\mathbb{N}} \;=\; \left\{ \sum_{n=1}^{N} c_n x_n : N \in \mathbb{N} \text{ and } c_1, \dots, c_N \in \mathbb{F} \right\}.$$

(c) The *closed linear span*, or simply the *closed span*, of S is the closure in X of $\operatorname{span}(S)$, and is denoted $\overline{\operatorname{span}}(S)$. If $S = \{x_n\}_{n\in\mathbb{N}}$ then we write $\overline{\operatorname{span}}\{x_n\}_{n\in\mathbb{N}} = \overline{\operatorname{span}}(\{x_n\}_{n\in\mathbb{N}})$.

(d) $\{x_n\}_{n\in\mathbb{N}}$ is *complete* (or *total* or *fundamental*) in X if $\overline{\operatorname{span}}(S) = X$, i.e., if $\operatorname{span}(S)$ is dense in X. $\diamondsuit$

Remark 5.4.2. Unfortunately, the term "complete" is heavily overused in mathematics, and indeed we have now introduced two distinct uses for it. First, a *normed linear space* X is complete if every Cauchy sequence in X is convergent. Second, a *sequence* $\{x_n\}_{n\in\mathbb{N}}$ in a normed linear space X is complete if $\operatorname{span}\{x_n\}_{n\in\mathbb{N}}$ is dense in X. Which of these two distinct uses is meant should be clear from context. $\diamondsuit$

Only separable normed spaces can contain a countable complete sequence.

Theorem 5.4.3. *Let X be a normed space. If there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ in X that is complete, then X is separable.*

Proof. Suppose that $\{x_n\}$ is a complete sequence in X, and let

$$S \;=\; \left\{\sum_{n=1}^N r_n x_n : N > 0,\ r_n \text{ is rational}\right\},$$

where if $\mathbb{F} = \mathbb{C}$ then "rational" means that both the real and imaginary parts are rational. Then S is countable, and we claim it is dense in X. Without loss of generality, we may assume that every x_n is nonzero.

Choose any $x \in X$. Since $\operatorname{span}(S)$ is dense in X, there exists a vector

$$y \;=\; \sum_{n=1}^N c_n x_n$$

such that $\|x - y\| < \varepsilon$. For each n, choose a rational scalar r_n such that

$$|c_n - r_n| \;<\; \frac{\varepsilon}{N\,\|x_n\|},$$

and set $z = \sum_{n=1}^N r_n x_n$. Then $z \in S$ and

$$\|y - z\| \;\le\; \sum_{n=1}^N |c_n - r_n|\,\|x_n\| \;<\; \sum_{n=1}^N \frac{\varepsilon}{N\,\|x_n\|}\,\|x_n\| \;=\; \varepsilon.$$

Hence $\|x - z\| < 2\varepsilon$, so S is dense in X. $\square$

It is very important to distinguish between elements of the closed span and vectors that can be written in the form $x = \sum_{n=1}^\infty c_n x_n$. Before elaborating on this, we give the definition of an infinite series in a normed space.

Definition 5.4.4 (Convergent Series). Let $\{x_n\}$ be a sequence in a normed linear space X. Then the series $\sum_{n=1}^\infty x_n$ *converges and equals* $x \in X$ if the *partial sums* $s_N = \sum_{n=1}^N x_n$ converge to x, i.e., if

$$\lim_{N\to\infty} \|x - s_N\| \;=\; \lim_{N\to\infty} \left\| x - \sum_{n=1}^{N} x_n \right\| \;=\; 0. \qquad \diamondsuit$$

We emphasize that the definition of the closed span does *not* say that

$$\overline{\text{span}}\{x_n\}_{n\in\mathbb{N}} \;=\; \left\{ \sum_{n=1}^{\infty} c_n x_n : c_n \in \mathbb{F} \right\} \qquad \leftarrow \text{This need not hold!}$$

In particular it is not true that an arbitrary element of $\overline{\text{span}}\{x_n\}_{n\in\mathbb{N}}$ can always be written as $x = \sum_{n=1}^{\infty} c_n x_n$ for some $c_n \in \mathbb{F}$ (see Example 5.4.5). Instead,

$$\overline{\text{span}}\{x_n\}_{n\in\mathbb{N}} = \left\{ x \in X : \exists\, c_{n,N} \in \mathbb{F} \text{ such that } \sum_{n=1}^{N} c_{n,N} x_n \to x \text{ as } N \to \infty \right\}.$$

That is, an element x lies in the closed span of $\{x_n\}_{n\in\mathbb{N}}$ if and only if there exist $c_{n,N} \in \mathbb{F}$ for $N \in \mathbb{N}$ and $n = 1, \dots, N$ such that

$$\sum_{n=1}^{N} c_{n,N} x_n \to x \quad \text{ as } N \to \infty. \tag{5.8}$$

We emphasize that the scalars $c_{n,N}$ in equation (5.8) can depend on N. In contrast, to say that $x = \sum_{n=1}^{\infty} c_n x_n$ means that

$$\sum_{n=1}^{N} c_n x_n \to x \quad \text{ as } N \to \infty. \tag{5.9}$$

In order for equation (5.9) to hold, the scalars c_n must be *independent* of N.

Example 5.4.5. Consider the Banach space $C[a,b]$ under the uniform norm. The *Weierstrass Approximation Theorem* states that if $f \in C[a,b]$ and $\varepsilon > 0$, then there exists polynomial $p(x) = \sum_{k=0}^{n} c_k x^k$ such that $\|f - p\|_\infty < \varepsilon$ (see [BBT97, Cor. 9.65]). This is equivalent to saying that the sequence of monomials $\{x^k\}_{k=0}^{\infty}$ is complete in $C[a,b]$ (which implies by Theorem 5.4.3 that $C[a,b]$ is separable). However, not every function $f \in C[a,b]$ can be written as $f(x) = \sum_{k=0}^{\infty} \alpha_k x^k$ with convergence of the series in the uniform norm. A series of this form is called a *power series*, and if it converges at some point x, then it converges absolutely for all points t with $|t| < r$ where $r = |x|$. Moreover, by Exercise 5.4.12 the function f so defined is infinitely differentiable on $(-r, r)$. Therefore, for example, the function $f(x) = |x - c|$ where $a < c < b$ cannot be written as a power series, even though it belongs to the closed span of $\{x^k\}_{k=0}^{\infty}$. In the language of Banach space theory, while $\{x^k\}_{k=0}^{\infty}$ is complete in $C[a,b]$, it does *not* form a Schauder basis for $C[a,b]$.

Moreover, $\{x^k\}_{k=0}^{\infty}$ is not a basis even though it is both complete *and* finitely linearly independent.

Although we will not prove it, given $f \in C[a,b]$ it is possible to explicitly construct polynomials p_n such that $\|f - p_n\|_\infty \to 0$ as $n \to \infty$. The *Bernstein polynomials for f* are one example of such a construction. $\diamondsuit$

On the other hand, the next example shows that it is possible for the closed span of a sequence to consist of "infinite linear combinations" of the sequence elements.

Example 5.4.6 (Standard Basis for c_0). For each $n \in \mathbb{N}$, we let δ_n denote the sequence $\delta_n = (\delta_{nk})_{k\in\mathbb{N}} = (0, \dots, 0, 1, 0, \dots)$, where the 1 is in the nth component. The finite span of $\{\delta_n\}$ is $\operatorname{span}\{\delta_n\} = c_{00}$. Given $x = (x_n) \in c_0$, set $s_N = \sum_{n=1}^{N} x_n \delta_n$. Then since $x_n \to 0$ we have

$$\lim_{N\to\infty} \|x - s_N\|_{\ell^\infty} \;=\; \lim_{N\to\infty} \sup_{n>N} |x_n| \;=\; \limsup_{n\to\infty} |x_n| \;=\; 0.$$

Hence

$$x \;=\; \sum_{n=1}^{\infty} x_n \delta_n, \tag{5.10}$$

where the series converges with respect to $\|\cdot\|_{\ell^\infty}$, which is the norm of c_0. Further, the scalars x_n in equation (5.10) are unique. Thus every x is a limit of elements of $\operatorname{span}\{\delta_n\}$ so $\{\delta_n\}$ is complete in c_0. However, even more is true. In the language of Banach space theory, the fact that every $x \in c_0$ has a unique representation of the form given in equation (5.10) says that $\{\delta_n\}$ forms a *Schauder basis* for c_0. Not only do we know that finite linear combinations of the vectors δ_n are dense in c_0, but we can actually write every element of c_0 as a unique "infinite linear combination" of the δ_n. Hence

$$c_0 \;=\; \overline{\operatorname{span}}\{\delta_n\} \;=\; \left\{\sum_{n=1}^{\infty} c_n\delta_n : c_n \in \mathbb{F} \text{ and } \lim_{n\to\infty} c_n = 0\right\}.$$

We call $\{\delta_n\}$ the *standard basis for c_0*. $\diamondsuit$

Note that if we use a norm other than $\|\cdot\|_{\ell^\infty}$, then the closed span of $\{\delta_n\}$ might be different.

Example 5.4.7 (Standard Basis for ℓ^p). If $x = (x_n) \in \ell^p$ where $1 \le p < \infty$ and we set $s_N = \sum_{n=1}^{N} x_n\delta_n$, then

$$\lim_{N\to\infty} \|x - s_N\|_{\ell^p} \;=\; \lim_{N\to\infty} \sum_{n=N+1}^{\infty} |x_n|^p \;=\; 0,$$

so we have $x = \sum x_n\delta_n$ with convergence of this series in ℓ^p-norm. Further, if $x = \sum c_n\delta_n$ for some scalars c_n, then we must have $c_n = x_n$ (why?). In the terminology of Banach space theory, the sequence $\{\delta_n\}$ is a Schauder basis for ℓ^p, which we call the *standard basis* for ℓ^p. Consequently ℓ^p is separable when p is finite, and an explicit countable dense subset is

$$S = \bigl\{x = (x_1, \dots, x_n, 0, 0, \dots) : n \in \mathbb{N},\, x_n \text{ rational}\bigr\}.$$

In contrast, ℓ^∞ is not separable; see Exercise 5.4.11. ◇

We introduce the following terminology to distinguish between a sequence that is merely complete in a Banach space X and one that has properties similar to those possessed by the standard basis in c_0 and ℓ^p.

Definition 5.4.8 (Schauder Basis). A sequence $\{x_n\}_{n\in\mathbb{N}}$ of vectors in a Banach space X is a *Schauder basis* for X if every $x \in X$ can be written

$$x = \sum_{n=1}^{\infty} c_n x_n \tag{5.11}$$

for a unique choice of scalars $c_n \in \mathbb{F}$. ◇

In particular, if $\{x_n\}_{n\in\mathbb{N}}$ is a Schauder basis and we write x as in equation (5.11), then the partial sums $s_N = \sum_{n=1}^{N} c_n x_n$ belong to the finite span of $\{x_n\}_{n\in\mathbb{N}}$ and converge to x as $N \to \infty$. Hence $\mathrm{span}\{x_n\}_{n\in\mathbb{N}}$ is dense in X. Therefore every Schauder basis is a complete sequence, but Example 5.4.5 shows us that a complete sequence need not be a Schauder basis.

We warn the reader not to confuse the meaning of "basis" in the sense of Definition 5.4.8 with the familiar notion of a "basis" in finite-dimensional linear algebra. A vector space basis (which we call a *Hamel basis* in this volume) is one that spans and is linearly independent using *finite linear combinations* only, whereas Definition 5.4.8 is worded in terms of "infinite linear combinations."

Exercises

5.4.9. Suppose that $\sum x_n$ and $\sum y_n$ are convergent series in a normed space X. Show that $\sum (x_n + y_n)$ is convergent and equals $\sum x_n + \sum y_n$.

5.4.10. Show that $C_0(\mathbb{R})$ is separable.

5.4.11. (a) Show that if X is a normed linear space and there exists an uncountable set $S \subseteq X$ such that $\|x - y\| = 1$ for every $x \neq y \in S$, then X is not separable.

(b) Let $D = \{x = (x_1, x_2, \dots) \in \ell^\infty : x_k = 0, 1 \text{ for each } k\}$. Show that D is uncountable and that if x, y are two distinct vectors in D, then $\|x-y\|_{\ell^\infty} = 1$. Conclude that ℓ^∞ is not separable.

(c) Show that $L^\infty(\mathbb{R})$ is not separable.

5.4.12. Let $(c_k)_{k\ge 0}$ be a fixed sequence of real numbers. Show that if the series $\sum_{k=0}^\infty c_k y^k$ converges for some $y \in \mathbb{R}$, then the series $f(x) = \sum_{k=0}^\infty c_k x^k$ converges absolutely for all $|x| < |y|$, and show that this function f is infinitely differentiable for all x with $|x| < |y|$.

5.5 Hilbert Spaces

A *Hilbert space* is a Banach space with additional geometric properties. In particular, the norm of a Hilbert space is obtained from an *inner product* that mimics the properties of the dot product of vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$. Recall that the dot product of u, $v \in \mathbb{C}^n$ is defined by

$$u \cdot v \;=\; u_1\overline{v_1} + \cdots + u_n\overline{v_n}. \tag{5.12}$$

The Euclidean norm $|v| = (|v_1|^2 + \cdots + |v_n|^2)^{1/2}$ is related to the dot product by the equation $|v| = (v\cdot v)^{1/2}$. For $p \ne 2$ the norm $|v|_p = (|v_1|^p + \cdots + |v_n|^p)^{1/p}$ has no such relation to the dot product. In fact, when $p \ne 2$ there is *no* way to define a "generalized dot product" $u \cdot v$ that has the same essential algebraic properties as the usual dot product and which also satisfies $|v|_p = (v \cdot v)^{1/2}$. These "essential algebraic properties" of the dot product are the properties (a)–(d) that appear in the following definition. Note that if $\mathbb{F} = \mathbb{R}$, then the complex conjugate appearing in this definition is superfluous.

Definition 5.5.1. A vector space H is an *inner product space* if for each x, $y \in H$ there exists a scalar $\langle x, y\rangle \in \mathbb{F}$, called the *inner product* of x and y, so that the following statements hold:

(a) $\langle x, x\rangle$ is real and $\langle x, x\rangle \ge 0$ for each $x \in H$,

(b) $\langle x, x\rangle = 0$ if and only if $x = 0$,

(c) $\langle y, x\rangle = \overline{\langle x, y\rangle}$ for all x, $y \in H$, and

(d) $\langle ax + by, z\rangle = a\langle x, z\rangle + b\langle y, z\rangle$ for all x, y, $z \in H$ and all a, $b \in \mathbb{F}$. ◇

Alternative symbols for inner products include $[\cdot,\cdot]$, $(\cdot,\cdot)$, etc.

Combining parts (c) and (d) of Definition 5.5.1, it follows by induction that

$$\Biggl\langle \sum_{n=1}^N c_n x_n,\, y \Biggr\rangle \;=\; \sum_{n=1}^N c_n\, \langle x_n, y\rangle \quad\text{and}\quad \Biggl\langle x,\, \sum_{n=1}^N c_n y_n \Biggr\rangle \;=\; \sum_{n=1}^N \overline{c_n}\, \langle x, y_n\rangle.$$

If H is an inner product space, then we will see in Theorem 5.5.5 that $\|x\| = \langle x, x\rangle^{1/2}$ defines a norm for H, called the *induced norm*. Hence all inner product spaces are normed linear spaces. If H is complete with respect to this induced norm, then H is called a *Hilbert space*. Thus Hilbert spaces are those Banach spaces whose norms can be derived from an inner product.

A given vector space may have many inner products.

Definition 5.5.2. We say that two inner products $\langle\cdot,\cdot\rangle$ and $(\cdot,\cdot)$ for a Hilbert space H are equivalent if the corresponding induced norms $\|x\|^2 = \langle x,x\rangle$ and $|||x|||^2 = (x,x)$ are equivalent in the sense of Definition 5.1.7. $\diamondsuit$

Now we give some examples of Hilbert spaces.

Example 5.5.3. (a) $\mathbb{F}^d$ is a Hilbert space with respect to the dot product given in equation (5.12). Also, if A is a positive definite $n \times n$ matrix with entries in $\mathbb{F}$, then

$$\langle x,y\rangle_A \;=\; Ax\cdot y, \qquad x,y\in\mathbb{F}^d, \tag{5.13}$$

defines another inner product on $\mathbb{F}^d$, and $\mathbb{F}^d$ is a Hilbert space with respect to this inner product. Moreover, every inner product on $\mathbb{F}^d$ has the form given in equation (5.13) for some positive definite matrix A, and all such inner products on $\mathbb{F}^d$ are equivalent (see Exercise 5.5.15).

(b) ℓ^2 is a Hilbert space with respect to the inner product

$$\big\langle (x_k),\,(y_k)\big\rangle \;=\; \sum_{k=1}^{\infty} x_k\,\overline{y_k}, \qquad (x_k),(y_k)\in\ell^2.$$

Note that the Cauchy–Bunyakovski–Schwarz Inequality implies that the series above converges absolutely for each choice of sequences (x_k), $(y_k)\in\ell^2$. Since c_{00} is a subset of ℓ^2, we can use the same rule to define an inner product on c_{00}. Thus c_{00} is an inner product space, but it is not complete with respect to this inner product.

(c) $L^2(E)$ is a Hilbert space with respect to the inner product

$$\langle f,g\rangle \;=\; \int_E f(t)\,\overline{g(t)}\,dt, \qquad f,g\in L^2(E).$$

The fact that the integral above exists is again a consequence of the Cauchy–Bunyakovski–Schwarz Inequality. The same rule defines an inner product on any subspace X of $L^2(E)$, but X will only be complete with respect to that inner product if it is a closed subspace of $L^2(E)$. For example,

$$X \;=\; \big\{f\in L^2(\mathbb{R}) : f=0 \text{ a.e. on } (-\infty,0)\big\}$$

is a closed subspace of $L^2(\mathbb{R})$ and hence is a Hilbert space with respect to the inner product of $L^2(\mathbb{R})$. On the other hand, $C_c(\mathbb{R})$ is a proper dense subspace of $L^2(\mathbb{R})$ and hence is not closed in $L^2(\mathbb{R})$. If we place the inner product from $L^2(\mathbb{R})$ on $C_c(\mathbb{R})$, then $C_c(\mathbb{R})$ is an inner product space, but it is not complete and therefore is not a Hilbert space with respect to that inner product. $\diamondsuit$

Here are a few basic properties of an inner product (see Exercise 5.5.16). We say that vectors x, y in a Hilbert space are *orthogonal* if $\langle x, y\rangle = 0$, and in this case we often write $x \perp y$.

Lemma 5.5.4. *Let H be a Hilbert space, and let $x, y \in H$ be given.*

(a) Polar Identity: $\|x+y\|^2 = \|x\|^2 + 2\,\mathrm{Re}(\langle x,y\rangle) + \|y\|^2$.

(b) Pythagorean Theorem: *If $\langle x,y\rangle = 0$ then $\|x \pm y\|^2 = \|x\|^2 + \|y\|^2$.*

(c) Parallelogram Law: $\|x+y\|^2 + \|x-y\|^2 = 2\big(\|x\|^2 + \|y\|^2\big)$. $\diamondsuit$

The following result generalizes the Cauchy–Bunyakovski–Schwarz inequality to any Hilbert space H, and shows that the induced norm is indeed a norm on H.

Theorem 5.5.5. *Let H be an inner product space.*

(a) Cauchy–Bunyakovski–Schwarz Inequality:

$$|\langle x,y\rangle| \le \|x\|\,\|y\| \qquad \text{for all } x, y \in H.$$

(b) $\|x\| = \langle x,x\rangle^{1/2}$ *is a norm on H.*

(c) $\|x\| = \sup\limits_{\|y\|=1} |\langle x,y\rangle|$.

Proof. (a) If $x = 0$ or $y = 0$ then there is nothing to prove, so suppose that both are nonzero. Write $\langle x,y\rangle = \alpha\,|\langle x,y\rangle|$ where $\alpha \in \mathbb{F}$ and $|\alpha| = 1$. Then for $t \in \mathbb{R}$ we have by the Polar Identity that

$$\begin{aligned} 0 \;\le\; \|x - \alpha t y\|^2 \;&=\; \|x\|^2 - 2\,\mathrm{Re}(\bar\alpha t\,\langle x,y\rangle) + t^2\,\|y\|^2 \\ &=\; \|x\|^2 - 2t\,|\langle x,y\rangle| + t^2\,\|y\|^2. \end{aligned}$$

This is a real-valued quadratic polynomial in the variable t. In order for it to be nonnegative, it can have at most one real root. This requires that the discriminant be at most zero, so $\big(-2\,|\langle x,y\rangle|\big)^2 - 4\,\|x\|^2\,\|y\|^2 \le 0$. The desired inequality then follows upon rearranging.

(b) The only property that is not obvious is the Triangle Inequality. From the Polar Identity and the Cauchy–Bunyakovski–Schwarz Inequality,

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

(c) We have $\sup_{\|y\|=1} |\langle x,y\rangle| \le \|x\|$ by the Cauchy–Bunyakovski–Schwarz Inequality, and the opposite inequality follows by considering $y = x/\|x\|$. $\square$

We obtain the following useful facts as corollaries of Cauchy–Bunyakovski–Schwarz.

Corollary 5.5.6. *Let H be a Hilbert space.*

(a) Continuity of the inner product: *If $x_n \to x$ and $y_n \to y$ in H, then $\langle x_n, y_n\rangle \to \langle x, y\rangle$.*

(b) *If the series $x = \sum_{n=1}^{\infty} x_n$ converges in H, then for any $y \in H$ we have*

$$\langle x, y\rangle \;=\; \left\langle \sum_{n=1}^{\infty} x_n,\, y \right\rangle \;=\; \sum_{n=1}^{\infty} \langle x_n, y\rangle.$$

Proof. (a) Suppose $x_n \to x$ and $y_n \to y$. Since convergent sequences are bounded, $C = \sup \|x_n\| < \infty$. Therefore

$$\begin{aligned}
|\langle x, y\rangle - \langle x_n, y_n\rangle| \;&\le\; |\langle x - x_n, y\rangle| + |\langle x_n, y - y_n\rangle| \\
&\le\; \|x - x_n\|\,\|y\| + \|x_n\|\,\|y - y_n\| \\
&\le\; \|x - x_n\|\,\|y\| + C\,\|y - y_n\| \;\to\; 0 \quad \text{as } n \to \infty.
\end{aligned}$$

(b) Suppose that the series $x = \sum_{n=1}^{\infty} x_n$ converges in H, and let $s_N = \sum_{n=1}^{N} x_n$ denote the partial sums of this series. Then, by definition, $s_N \to x$ in H. Hence, given $y \in H$ we have

$$\begin{aligned}
\sum_{n=1}^{\infty} \langle x_n, y\rangle \;&=\; \lim_{N\to\infty} \left(\sum_{n=1}^{N} \langle x_n, y\rangle \right) \\
&=\; \lim_{N\to\infty} \left\langle \sum_{n=1}^{N} x_n,\, y \right\rangle \;=\; \lim_{N\to\infty} \langle s_N, y\rangle \;=\; \langle x, y\rangle,
\end{aligned}$$

where at the last step we have used the continuity of the inner product. □

Now we give the definition and basic properties of orthogonal projections in a Hilbert space.

Theorem 5.5.7. *Let H be a Hilbert space, and let M be a closed subspace of H. Given $x \in H$, there exists a unique element $p \in M$ that is closest to x, i.e., $\|x - p\| = \mathrm{dist}(x, M) = \inf\{\|x - m\| : m \in M\}$.*

Proof. Fix $x \in H$, and let $d = \mathrm{dist}(x, M) = \inf\{\|x - m\| : m \in M\}$. Then, by definition, there exist $y_n \in M$ such that $d \le \|x - y_n\| \to d$ as $n \to \infty$. Therefore, if we fix any $\varepsilon > 0$ then we can find an N such that

$$n > N \qquad \Longrightarrow \qquad d^2 \le \|x - y_n\|^2 \le d^2 + \varepsilon^2.$$

By the Parallelogram Law,

$$\|(x-y_n)-(x-y_m)\|^2+\|(x-y_n)+(x-y_m)\|^2 \;=\; 2\left(\|x-y_n\|^2+\|x-y_m\|^2\right).$$

Hence,

$$\begin{aligned}\left\|\frac{y_m-y_n}{2}\right\|^2 \;&=\; \frac{1}{4}\,\|(x-y_n)-(x-y_m)\|^2\\ &=\; \frac{\|x-y_n\|^2}{2}+\frac{\|x-y_m\|^2}{2}-\left\|x-\frac{y_m+y_n}{2}\right\|^2.\end{aligned}$$

However, $\frac{y_m+y_n}{2} \in M$ since M is a subspace, so $\|x-\frac{y_m+y_n}{2}\| \ge d$. Also, if $m, n > N$ then $\|x-y_n\|^2$, $\|x-y_m\|^2 \le d^2+\varepsilon^2$. Therefore, for $m, n > N$ we have

$$\left\|\frac{y_m-y_n}{2}\right\|^2 \;\le\; \frac{d^2+\varepsilon^2}{2}+\frac{d^2+\varepsilon^2}{2}-d^2 \;=\; \varepsilon^2.$$

Thus, $\|y_m-y_n\| \le 2\varepsilon$ for all $m, n > N$, which says that the sequence $\{y_n\}$ is Cauchy. Since H is complete, this sequence must converge, so $y_n \to p$ for some $p \in H$. But $y_n \in M$ for all n and M is closed, so we must have $p \in M$. Since $x-y_n \to x-p$, it follows from the continuity of the norm (Exercise 5.1.12) that

$$\|x-p\| \;=\; \lim_{n\to\infty}\|x-y_n\| \;=\; d,$$

and hence $\|x-p\| \le \|x-y\|$ for every $y \in M$. Thus p is a closest point in M to x, and we leave as an exercise the task of proving that p is unique (Exercise 5.5.20). □

The proof of Theorem 5.5.7 carries over without change to show that if K is a closed, convex subset of a Hilbert space, then given any $x \in H$ there is a unique point p in K that is closest to x.

Definition 5.5.8. Let M be a closed subspace of a Hilbert space H.

(a) Given $x \in H$, the unique vector $p \in M$ that is closest to x is called the *orthogonal projection* of x onto M.

(b) For $x \in H$ let Px denote the vector that is the orthogonal projection of x onto M. Then the mapping $P\colon x \mapsto Px$ is called the *orthogonal projection* of H onto M. ◇

Now we define orthogonal complements, which play an important role in the analysis of Hilbert spaces. The lack of orthogonal projections and orthogonal complements in non-Hilbert spaces is often what makes the analysis of generic Banach spaces so much more difficult than it is for Hilbert spaces.

Definition 5.5.9 (Orthogonal Complement). Let A be a subset (not necessarily closed or a subspace) of a Hilbert space H. The *orthogonal complement* of A is

$$A^\perp \;=\; \bigl\{x \in H : x \perp A\bigr\} \;=\; \bigl\{x \in H : \langle x, y\rangle = 0 \text{ for all } y \in A\bigr\}. \qquad \diamondsuit$$

The orthogonal complement of A is always a closed subspace of H, even if A is not (Exercise 5.5.21).

Lemma 5.5.10. *If A is a subset of a Hilbert space H, then $A^\perp$ is a closed subspace of H.* $\diamondsuit$

Orthogonal projections can be characterized in terms of orthogonal complements as follows (see Exercise 5.5.22).

Theorem 5.5.11. *If M is a closed subspace of a Hilbert space H and $x \in H$, then the following statements are equivalent.*

(a) *$x = p + e$ where p is the orthogonal projection of x onto M.*

(b) *$x = p + e$ where $p \in M$ and $e \in M^\perp$.*

(c) *$x = p + e$ where e is the orthogonal projection of x onto $M^\perp$.* $\diamondsuit$

Consequently, if P is the orthogonal projection of H onto M, then the orthogonal projection of H onto $M^\perp$ is $I - P$.

Here are some properties of orthogonal complements. (Exercise 5.5.23).

Lemma 5.5.12. *Let H be a Hilbert space.*

(a) *If M is a closed subspace of H, then $(M^\perp)^\perp = M$.*

(b) *If A is any subset of H, then*

$$A^\perp \;=\; \mathrm{span}(A)^\perp \;=\; \overline{\mathrm{span}}(A)^\perp \qquad \textit{and} \qquad (A^\perp)^\perp \;=\; \overline{\mathrm{span}}(A).$$

(c) *A sequence $\{x_n\}_{n\in\mathbb{N}}$ in H is complete if and only if the following statement holds:*

$$x \in H \text{ and } \langle x, x_n\rangle = 0 \text{ for every } n \quad \Longrightarrow \quad x = 0. \qquad \diamondsuit$$

Definition 5.5.13 (Orthogonal Direct Sum). Let M, N be closed subspaces of a Hilbert space H.

(a) The *direct sum* of M and N is $M + N = \{x + y : x \in M, y \in N\}$.

(b) We say that M and N are *orthogonal subspaces*, denoted $M \perp N$, if $x \perp y$ for every $x \in M$ and $y \in N$.

(c) If M, N are orthogonal subspaces in H, then we call their direct sum the *orthogonal direct sum* of M and N, and denote it by $M \oplus N$. $\diamondsuit$

The proof of the next lemma is Exercise 5.5.24.

Lemma 5.5.14. *Let M, N be closed, orthogonal subspaces of H.*

(a) *$M \oplus N$ is a closed subspace of H.*

(b) *$M \oplus M^\perp = H$.* $\diamondsuit$

Exercises

5.5.15. An $n \times n$ matrix A is said to be *positive definite* if $Ax \cdot x > 0$ for all $x \in \mathbb{F}^n$.

(a) Show that if A is a positive definite $n \times n$ matrix then equation (5.13) defines an inner product on $\mathbb{F}^d$.

(b) Show that if $\langle \cdot, \cdot \rangle$ is an inner product on $\mathbb{F}^d$ then there exists some positive definite matrix A such that equation (5.13) holds.

(c) Prove that the inner products on $\mathbb{F}^d$ defined by equation (5.13) are all equivalent.

5.5.16. Prove Lemma 5.5.4.

5.5.17. Show that if $p \neq 2$ then the norm $\|\cdot\|_{\ell^p}$ on ℓ^p is not induced from any inner product on ℓ^p.

5.5.18. Show that equality holds in the Cauchy–Bunyakovski–Schwarz Inequality if and only if $x = cy$ or $y = cx$ for some scalar $c \in \mathbb{F}$.

5.5.19. We say that $\langle \cdot, \cdot \rangle$ is a *semi-inner product* on a vector space H if properties (a), (c), and (d) of Definition 5.5.1 are satisfied. Prove that if $\langle \cdot, \cdot \rangle$ is a semi-inner product then $\|x\| = \langle x, x \rangle^{1/2}$ defines a seminorm on H, and $|\langle x, y \rangle| \le \|x\|\,\|y\|$ for all $x, y \in H$.

5.5.20. Given a closed subspace M of a Hilbert space H and given $x \in H$, prove that the point in M closest to x is unique.

5.5.21. Prove Lemma 5.5.10.

5.5.22. Prove Theorem 5.5.11.

5.5.23. Prove Lemma 5.5.12.

5.5.24. Prove Lemma 5.5.14.

5.5.25. Let H, K be Hilbert spaces. Show that $H \times K$ is a Hilbert space with respect to the inner product $\big\langle (h_1, k_1), (h_2, k_2) \big\rangle = \langle h_1, h_2 \rangle_H + \langle k_1, k_2 \rangle_K$.

5.5.26. Prove that the completion $\widetilde{H}$ (Exercise 5.3.18) of an inner product space H is a Hilbert space with respect to an inner product that extends the inner product on H.

5.6 Orthogonal Sequences in Hilbert Spaces

Two vectors x, y in a Hilbert space are *orthogonal* if $\langle x, y\rangle = 0$. Sequences in a Hilbert space which possess the property that any two distinct elements are orthogonal have a number of useful features, which we consider in this section.

Definition 5.6.1. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in a Hilbert space H.

(a) $\{x_n\}_{n\in\mathbb{N}}$ is an *orthogonal sequence* if $\langle x_m, x_n\rangle = 0$ whenever $m \neq n$.

(b) $\{x_n\}_{n\in\mathbb{N}}$ is an *orthonormal sequence* if $\langle x_m, x_n\rangle = \delta_{mn}$, i.e., $\{x_n\}_{n\in\mathbb{N}}$ is orthogonal and $\|x_n\| = 1$ for every n.

(c) We recall from Definition 5.4.8 that $\{x_n\}_{n\in\mathbb{N}}$ is a *Schauder basis* for H if every $x \in H$ can be written $x = \sum_{n=1}^{\infty} c_n x_n$ for a unique choice of scalars c_n.

(d) An orthonormal sequence $\{x_n\}_{n\in\mathbb{N}}$ is an *orthonormal basis* if it is both orthonormal and a Schauder basis. $\diamondsuit$

The definition of orthogonal and orthonormal sequences in a Hilbert space H can be extended to arbitrary subsets of H. In particular, we say that $S \subseteq H$ is orthogonal if given any $x \neq y \in S$ we have $\langle x, y\rangle = 0$.

While orthogonality only makes sense in an inner product space, the definition of a Schauder basis extends without change to Banach spaces. Here in this section we will concentrate on the specific issue of basis properties of orthonormal sequences in Hilbert spaces. We note that in the Banach space literature and in this volume the word *basis* is reserved for *countable* sequences that satisfy statement (c) of Definition 5.6.1.

We first recall the Pythagorean Theorem (Lemma 5.5.4), extended by induction to finite collections of orthogonal vectors.

Lemma 5.6.2 (Pythagorean Theorem). *If $\{x_1, \dots, x_N\}$ are orthogonal vectors in an inner product space H, then $\left\|\sum_{n=1}^{N} x_n\right\|^2 = \sum_{n=1}^{N} \|x_n\|^2$.* $\diamondsuit$

The following result summarizes some basic results connected to convergence of infinite series of orthonormal vectors.

Theorem 5.6.3. *If $\{x_n\}$ is an orthonormal sequence in a Hilbert space H, then the following statements hold.*

(a) Bessel's Inequality: $\sum |\langle x, x_n\rangle|^2 \le \|x\|^2$ *for every $x \in H$.*

(b) *If $x = \sum c_n x_n$ converges, then $c_n = \langle x, x_n\rangle$.*

(c) $\sum c_n x_n$ *converges* $\iff$ $\sum |c_n|^2 < \infty$.

(d) $x \in \overline{\operatorname{span}}\{x_n\}_{n\in\mathbb{N}} \iff x = \sum \langle x, x_n\rangle\, x_n$.

(e) *If $x \in H$, then $p = \sum \langle x, x_n\rangle\, x_n$ is the orthogonal projection of x onto $\overline{\operatorname{span}}\{x_n\}_{n\in\mathbb{N}}$.*

Proof. We will prove some statements, and the rest are assigned as Exercise 5.6.12.

(a) Choose $x \in H$. For each $N \in \mathbb{N}$ define $y_N = x - \sum_{n=1}^{N} \langle x, x_n \rangle \, x_n$. If $1 \le m \le N$, then

$$\langle y_N, x_m \rangle \;=\; \langle x, x_m \rangle - \sum_{n=1}^{N} \langle x, x_n \rangle \, \langle x_n, x_m \rangle \;=\; \langle x, x_m \rangle - \langle x, x_m \rangle \;=\; 0.$$

Thus $y_N \perp x_1, \dots, x_N$. Therefore, by the Pythagorean Theorem,

$$\begin{aligned} \|x\|^2 \;&=\; \Bigl\| y_N + \sum_{n=1}^{N} \langle x, x_n \rangle \, x_n \Bigr\|^2 \\ &=\; \|y_N\|^2 \;+\; \sum_{n=1}^{N} \|\langle x, x_n \rangle \, x_n\|^2 \\ &=\; \|y_N\|^2 \;+\; \sum_{n=1}^{N} |\langle x, x_n \rangle|^2 \;\ge\; \sum_{n=1}^{N} |\langle x, x_n \rangle|^2. \end{aligned}$$

Letting $N \to \infty$, we obtain Bessel's Inequality.

(c) Suppose that $\sum_{n=1}^{\infty} |c_n|^2 < \infty$. Set

$$s_N \;=\; \sum_{n=1}^{N} c_n x_n \qquad \text{and} \qquad t_N \;=\; \sum_{n=1}^{N} |c_n|^2.$$

We know that $\{t_N\}_{N \in \mathbb{N}}$ is a convergent (hence Cauchy) sequence of scalars, and we must show that $\{s_N\}_{N \in \mathbb{N}}$ is a convergent sequence of vectors. We have for $N > M$ that

$$\begin{aligned} \|s_N - s_M\|^2 \;&=\; \Bigl\| \sum_{n=M+1}^{N} c_n x_n \Bigr\|^2 \\ &=\; \sum_{n=M+1}^{N} \|c_n x_n\|^2 \;=\; \sum_{n=M+1}^{N} |c_n|^2 \;=\; |t_N - t_M|. \end{aligned}$$

Since $\{t_N\}_{N \in \mathbb{N}}$ is Cauchy, we conclude that $\{s_N\}_{N \in \mathbb{N}}$ is a Cauchy sequence in H and hence converges.

(d) Choose $x \in \overline{\operatorname{span}}\{x_n\}_{n \in \mathbb{N}}$. By Bessel's Inequality, $\sum |\langle x, x_n \rangle|^2 < \infty$, and therefore by part (c) we know that the series $y = \sum \langle x, x_n \rangle \, x_n$ converges. Given any particular $m \in \mathbb{N}$, by applying Corollary 5.5.6(b) we have

$$\begin{aligned}
\langle x-y, x_m\rangle &= \langle x, x_m\rangle - \left\langle \sum_n \langle x, x_n\rangle\, x_n,\, x_m \right\rangle \\
&= \langle x, x_m\rangle - \sum_n \langle x, x_n\rangle \langle x_n, x_m\rangle \\
&= \langle x, x_m\rangle - \langle x, x_m\rangle = 0.
\end{aligned}$$

Thus $x - y \in \{x_n\}_{n\in\mathbb{N}}^{\perp} = \overline{\operatorname{span}}\{x_n\}_{n\in\mathbb{N}}^{\perp}$. However, we also have $x - y \in \overline{\operatorname{span}}\{x_n\}_{n\in\mathbb{N}}$, so $x - y = 0$. $\square$

It is tempting to conclude from Theorem 5.6.3 that if $\{x_n\}_{n\in\mathbb{N}}$ is an orthonormal sequence in a Hilbert space H, then *every* $x \in H$ can be written $x = \sum \langle x, x_n\rangle\, x_n$. This, however, is not always the case, for there may not be "enough" vectors in the sequence to span all of H. In particular, if $\{x_n\}_{n\in\mathbb{N}}$ is not complete, then its closed span is only a proper closed subspace of H and not all of H. For example, a finite sequence of orthonormal vectors $\{x_1, \dots, x_N\}$ can only span a finite-dimensional subspace of an infinite-dimensional Hilbert space, and therefore cannot be complete in an infinite-dimensional space. As another example, if $\{x_n\}_{n\in\mathbb{N}}$ is an orthonormal sequence in H, then $\{x_{2n}\}$ is also an orthonormal sequence in H. However, x_1 is orthogonal to every x_{2n}, so it follows from Lemma 5.5.12 that $\{x_{2n}\}$ is incomplete.

The next theorem presents several equivalent conditions which imply that an orthonormal sequence is complete in H.

Theorem 5.6.4. *If $\{x_n\}$ is an orthonormal sequence in a Hilbert space H, then the following statements are equivalent.*

(a) *$\{x_n\}$ is complete in H.*

(b) *$\{x_n\}$ is a Schauder basis for H, i.e., for each $x \in H$ there exists a unique sequence of scalars (c_n) such that $x = \sum c_n x_n$.*

(c) *$x = \sum \langle x, x_n\rangle\, x_n$ for each $x \in H$.*

(d) Plancherel's Equality: *$\|x\|^2 = \sum |\langle x, x_n\rangle|^2$ for all $x \in H$.*

(e) Parseval's Equality: *$\langle x, y\rangle = \sum \langle x, x_n\rangle \langle x_n, y\rangle$ for all $x, y \in H$.*

Proof. We will prove some implications, and assign the rest as Exercise 5.6.12.

(a) $\Rightarrow$ (c). If $\{x_n\}$ is complete, then its closed span is all of H by definition, so by Theorem 5.6.3(d) we have that $x = \sum \langle x, x_n\rangle\, x_n$ for every $x \in H$.

(d) $\Rightarrow$ (c). Suppose that $\|x\|^2 = \sum |\langle x, x_n\rangle|^2$ for all $x \in H$. Fix $x \in H$, and define $s_N = \sum_{n=1}^{N} \langle x, x_n\rangle\, x_n$. Then, by the Polar Identity and the Pythagorean Theorem,

$$\begin{aligned}\|x - s_N\|^2 &= \|x\|^2 - 2\,\mathrm{Re}(\langle x, s_N\rangle) + \|s_N\|^2 \\ &= \|x\|^2 - 2\sum_{n=1}^{N} |\langle x, x_n\rangle|^2 + \sum_{n=1}^{N} |\langle x, x_n\rangle|^2 \\ &= \|x\|^2 - \sum_{n=1}^{N} |\langle x, x_n\rangle|^2 \;\to\; 0 \quad \text{as } N \to \infty.\end{aligned}$$

Hence $x = \sum \langle x, x_n\rangle\, x_n$. $\square$

As the Plancherel and Parseval Equalities are equivalent, these terms are often used interchangeably.

Note that Theorem 5.6.4 implies that every *complete orthonormal sequence* in a Hilbert space is actually a *Schauder basis* for H. This need not be true for nonorthogonal sequences. An example in ℓ^2 is constructed in Exercise 5.6.16. In fact, that example is complete and finitely linearly independent yet is not a Schauder basis for ℓ^2.

Now we give some examples of orthonormal bases.

Example 5.6.5 (Standard Basis for ℓ^2). By Example 5.4.7, the standard basis $\{\delta_n\}$ is complete in ℓ^2, and it is clearly orthonormal. Hence it is an orthonormal basis for ℓ^2. $\diamondsuit$

Example 5.6.6 (The Trigonometric System). Let $H = L^2(\mathbb{T})$ denote the space of complex-valued functions that are 1-periodic on $\mathbb{R}$ and are square integrable on the interval $[0,1]$. The norm and inner product are defined by integrating on the interval $[0,1]$, i.e.,

$$\langle f, g\rangle \;=\; \int_0^1 f(t)\,\overline{g(t)}\,dt \qquad \text{and} \qquad \|f\|_{L^2}^2 \;=\; \int_0^1 |f(t)|^2\,dt.$$

For each $n \in \mathbb{Z}$, define a function e_n by

$$e_n(t) \;=\; e^{2\pi i n t}, \qquad t \in \mathbb{R}.$$

We call $\{e_n\}_{n\in\mathbb{Z}}$ the *trigonometric system*. Given integers $m \neq n$,

$$\begin{aligned}\langle e_m, e_n\rangle &= \int_0^1 e_m(t)\,\overline{e_n(t)}\,dt \\ &= \int_0^1 e^{-2\pi i(m-n)t}\,dt \;=\; \frac{e^{-2\pi i(m-n)} - 1}{-2\pi i(m-n)} \;=\; 0.\end{aligned}$$

Hence $\{e_n\}_{n\in\mathbb{Z}}$ is an orthogonal sequence in $L^2(\mathbb{T})$, and it is easy to check that it is orthonormal.

It is a more subtle fact that $\{e_n\}_{n\in\mathbb{Z}}$ is complete in $L^2(\mathbb{T})$. One approach to proving this relies on techniques from harmonic analysis. We will simply assume that the trigonometric system is complete in $L^2(\mathbb{T})$. Assuming this completeness, Theorem 5.6.4 implies that $\{e_n\}_{n\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{T})$.

If $f \in L^2(\mathbb{T})$ then the expansion $f = \sum_{n\in\mathbb{Z}} \langle f, e_n\rangle\, e_n$ is called the *Fourier series representation* of f, and $\big(\langle f, e_n\rangle\big)_{n\in\mathbb{Z}}$ is the sequence of *Fourier coefficients* of f. The Fourier coefficients are often denoted by

$$\widehat{f}(n) \;=\; \langle f, e_n\rangle = \int_0^1 f(t)\, e^{-2\pi i n t}\, dt, \qquad n \in \mathbb{Z}.$$

The elements of the space $L^2(\mathbb{T})$ are 1-periodic functions on the real line. Sometimes it is more convenient to work with the space $L^2[0,1]$ consisting of complex-valued square integrable functions whose domain is the interval $[0,1]$. All of the statements above apply equally to $L^2[0,1]$, i.e., $\{e_n\}_{n\in\mathbb{Z}}$ is an orthonormal basis for $L^2[0,1]$. In fact, by the periodicity of the exponentials, $\{e_n\}_{n\in\mathbb{Z}}$ is an orthonormal basis for $L^2(I)$ where I is any interval in $\mathbb{R}$ of length 1. $\diamondsuit$

Note that we are only guaranteed that the Fourier series of $f \in L^2(\mathbb{T})$ will converge in L^2-norm. Establishing the convergence of Fourier series in other senses can be extremely difficult. In Chapter 14 of [Heil11]; we prove there that the symmetric partial sums $s_N(x) = \sum_{n=-N}^{n} \widehat{f}(n)\, e^{2\pi i n x}$ of the Fourier series of $f \in L^p(\mathbb{T})$ converge in L^p-norm when $1 < p < \infty$. One of the deepest results in harmonic analysis, which we will not prove, is the *Carleson–Hunt Theorem,* which states that the symmetric partial sums of the Fourier series of a function $f \in L^p(\mathbb{T})$ converge pointwise almost everywhere to f when $1 < p < \infty$.

Notation 5.6.7. If we prefer to work with real-valued functions,

$$\{1\} \cup \{\sqrt{2}\, \sin 2\pi n t\}_{n\in\mathbb{N}} \cup \{\sqrt{2}\, \cos 2\pi n t\}_{n\in\mathbb{N}}$$

forms an orthonormal basis for $L^2(\mathbb{T})$ or $L^2[0,1]$ when we take $\mathbb{F} = \mathbb{R}$ (see Exercise 5.6.19). $\diamondsuit$

In light of Example 5.6.6, if $\{e_n\}$ is an orthonormal basis for an arbitrary Hilbert space H, then the basis representation $x = \sum \langle x, e_n\rangle\, e_n$ is sometimes called the *generalized Fourier series* of $x \in H$, and $\big(\langle x, e_n\rangle\big)$ is called the sequence of *generalized Fourier coefficients* of x.

The next example gives another important orthonormal basis.

Example 5.6.8 (The Haar System). Let $\chi = \chi_{[0,1)}$ be the box function. The function

$$\psi \;=\; \chi_{[0,1/2)} - \chi_{[1/2,1)}$$

is called the *Haar wavelet.* For integer n, $k \in \mathbb{Z}$, define

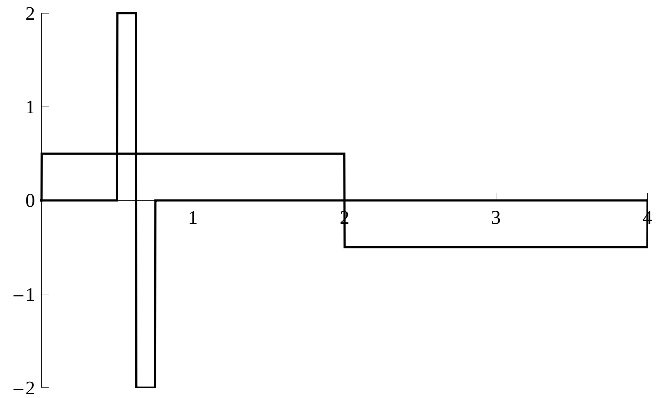


Fig. 5.1 Graphs of $\psi_{-2,0}$ and $\psi_{2,2}$.

$$\psi_{n,k}(t) \;=\; 2^{n/2}\psi(2^n t - k).$$

The *Haar system* for $L^2(\mathbb{R})$ is

$$\big\{\chi(t-k)\big\}_{k\in\mathbb{Z}} \;\cup\; \big\{\psi_{n,k}\big\}_{n\geq 0,\, k\in\mathbb{Z}}.$$

Direct calculations show that the Haar system is an orthonormal sequence in $L^2(\mathbb{R})$ (see Exercise 5.6.20 or the "proof by picture" in Figure 5.1), and we will prove that the Haar system is complete in $L^2(\mathbb{R})$.

Suppose that $f \in L^2(\mathbb{R})$ is orthogonal to each element of the Haar system. Considering the integer translates $\chi(t-k)$ of the box function, this implies that

$$\int_k^{k+1} f(t)\,dt \;=\; 0, \qquad k \in \mathbb{Z}.$$

Next, since $f \perp \chi$ we have

$$\int_0^{1/2} f(t)\,dt + \int_{1/2}^1 f(t)\,dt \;=\; \int_0^1 f(t)\,dt \;=\; \langle f, \chi\rangle \;=\; 0,$$

and since $f \perp \psi$ we have

$$\int_0^{1/2} f(t)\,dt - \int_{1/2}^1 f(t)\,dt \;=\; \langle f, \psi\rangle \;=\; 0.$$

Adding and subtracting,

$$\int_0^{1/2} f(t)\,dt \;=\; 0 \;=\; \int_{1/2}^1 f(t)\,dt.$$

Continuing in this way, it follows that

$$\int_{I_{n,k}} f(t)\,dt \;=\; 0 \quad \text{for every dyadic interval } I_{n,k} = \Bigl[\frac{k}{2^n}, \frac{k+1}{2^n}\Bigr].$$

Given $t \in \mathbb{R}$, for each $n \in \mathbb{N}$ there exists some dyadic interval $J_n(t) = I_{n,k_n(t)}$ such that $t \in J_n(t)$. The diameter of $J_n(t)$ shrinks rapidly with n, and we have $\cap J_n(t) = \{t\}$. We appeal now to a fundamental result from real analysis, the Lebesgue Differentiation Theorem, which implies that for almost every $t \in \mathbb{R}$ we have

$$f(t) \;=\; \lim_{n\to\infty} \frac{1}{|J_n(t)|} \int_{J_n(t)} f(u)\,du.$$

Since $\int_{J_n(t)} f(u)\,du = 0$ for every n, we conclude that $f = 0$ a.e. $\diamondsuit$

Remark 5.6.9. (a) The sequence $\{\psi_{n,k}\}_{n,k\in\mathbb{Z}}$, containing dilations of the Haar wavelet at all dyadic scales 2^n with $n \in \mathbb{Z}$, also forms an orthonormal basis for $L^2(\mathbb{R})$, and this basis is also referred to as the Haar system for $L^2(\mathbb{R})$. This system is the simplest example of a *wavelet orthonormal basis* for $L^2(\mathbb{R})$. Unfortunately, our proof that the Haar system is an orthonormal basis yields little insight into the elegant construction of general wavelet orthonormal bases, which is the topic of Chapter 12 in [Heil11].

(b) The system originally introduced by Haar is

$$\bigl\{\chi\bigr\} \;\cup\; \bigl\{\psi_{n,k}\bigr\}_{n\ge 0,\, k=0,\dots,2^n-1},$$

which forms an orthonormal basis for $L^2[0,1]$. $\diamondsuit$

Suppose that a Hilbert space H has an orthonormal basis $\{e_n\}$. Then

$$S \;=\; \biggl\{\sum_{n=1}^{N} r_n e_n : N > 0, \;\; \text{rational } r_n \in \mathbb{F}\biggr\} \tag{5.14}$$

is a countable, dense subset of H, so H is separable (see Theorem 5.4.3). The next result proves the converse, i.e., every separable Hilbert space possesses an orthonormal basis. This proof requires some familiarity with *Zorn's Lemma*, which is an equivalent form of the *Axiom of Choice*.

Theorem 5.6.10. *Let H be a Hilbert space.*

(a) *H contains a subset T that is both complete and orthonormal.*

(b) *H has an orthonormal basis $\{e_n\}_{n\in\mathbb{N}}$ if and only if H is separable.*

Proof. (a) Let $\mathcal{S}$ denote the set of all orthonormal subsets of H. Inclusion of sets forms a partial order on $\mathcal{S}$.

Suppose that $\mathcal{C} = \{S_i\}_{i\in I}$ is a chain in H, i.e., I is an arbitrary index set and for each $i,\, j \in I$ we have either $S_i \subseteq S_j$ or $S_j \subseteq S_i$. Define $S = \cup_{i\in I} S_i$. If $x,\, y$ are two distinct elements of S, then $x \in S_i$ and $y \in S_j$ for some i

and j. Since $\mathcal{C}$ is a chain, we must either have $x, y \in S_i$ or $x, y \in S_j$. In any case, $\langle x, y\rangle = 0$ since S_i and S_j are each orthonormal. Thus S is itself orthonormal. Since $S_i \subseteq S$ for every $i \in I$, this tells us that S is an *upper bound* for the chain $\mathcal{C}$.

Zorn's Lemma says that, given a partially ordered set, if every chain has an upper bound, then the set has a maximal element. Therefore, $\mathcal{S}$ must have a maximal element, i.e., there exists some orthonormal set $T \in \mathcal{S}$ which has the property that if $S \in \mathcal{S}$ and S is comparable to T (either $S \subseteq T$ or $T \subseteq S$), then we must have $S \subseteq T$.

We claim now that T is a complete orthonormal subset of H. If T is not complete, then $\overline{\operatorname{span}}(T)$ is a proper subset of H, and hence there exists some nonzero vector $x \in \overline{\operatorname{span}}(T)^{\perp}$. By rescaling, we may assume $\|x\| = 1$. But then $T' = T \cup \{x\}$ is orthonormal and $T \subsetneq T'$, contradicting the fact that T is a maximal element of $\mathcal{S}$. Hence T must be complete.

(b) We already know that if H contains an orthonormal basis then it is separable, so suppose that H is an arbitrary separable Hilbert space. By part (a), H contains a complete orthonormal subset T. By Exercise 5.6.28, any orthonormal subset of a separable Hilbert space must be countable, so we can write $T = \{e_n\}_{n\in\mathbb{N}}$. Thus T is a complete orthonormal sequence, and therefore it is an orthonormal basis by Theorem 5.6.4. □

Example 5.6.11. An example of a nonseparable Hilbert space is the space $\ell^2(\mathbb{R})$ consisting of all sequences $x = (x_i)_{i\in\mathbb{R}}$ indexed by the real line with at most countably many terms nonzero and such that $\sum |x_i|^2 < \infty$. The inner product on this space is $\langle x, y\rangle = \sum x_i\,\overline{y_i}$ (see Exercise 5.6.28). ♢

Exercises

5.6.12. Prove the remaining parts of Theorems 5.6.3 and 5.6.4.

5.6.13. (a) Let M be a proper, closed subspace of a Hilbert space H. Given $x \in H\backslash M$, let p be the orthogonal projection of x onto M, and show that the vector $y = (x-p)/\|x-p\|$ satisfies $\|y\| = 1$, $y \in M^{\perp}$, and $\operatorname{dist}(y, M) = \inf_{m\in M} \|y - m\| = 1$.

(b) Show that if H is an infinite-dimensional Hilbert space, then there exists an infinite orthonormal sequence $\{e_n\}$ in H, and no subsequence $\{e_{n_k}\}$ is Cauchy. Contrast this with the *Bolzano–Weierstrass Theorem*, which states that every bounded sequence in $\mathbb{F}^d$ has a convergent subsequence.

Remark: In another language, this problem shows that the closed unit ball $D = \{x \in H : \|x\| \le 1\}$ in H is *not compact* when H is infinite dimensional.

5.6.14. This exercise will extend Exercise 5.6.13 to an arbitrary normed linear space X.

(a) Prove *F. Riesz's Lemma*: If M is a proper, closed subspace of X and $\varepsilon > 0$, then there exists $x \in X$ with $\|x\| = 1$ such that $\operatorname{dist}(x, M) = \inf_{m \in M} \|x - m\| > 1 - \varepsilon$.

(b) Prove that if X is infinite dimensional then there exists a bounded sequence $\{x_n\}_{n\in\mathbb{N}}$ in X that has no convergent subsequences (hence the closed unit ball $D = \{x \in X : \|x\| \le 1\}$ is never compact in an infinite-dimensional Banach space).

5.6.15. Let $\{x_n\}$ be a finitely linearly independent sequence in a Hilbert space H. Show that there exists an orthogonal sequence $\{y_n\}$ in H such that $\operatorname{span}\{y_1, \dots, y_N\} = \operatorname{span}\{x_1, \dots, x_N\}$ for each $N \in \mathbb{N}$ (this is the *Gram–Schmidt orthogonalization procedure*).

5.6.16. We say that a sequence $\{x_n\}$ in a Banach space X is *ω-dependent* if there exist scalars c_n, not all zero, such that $\sum c_n x_n = 0$, where the series converges in the norm of X. A sequence is *ω-independent* if it is not ω-dependent.

(a) Show that if $\{x_n\}_{n\in\mathbb{N}}$ is a basis for a Hilbert space H then $\{x_n\}_{n\in\mathbb{N}}$ is ω-independent.

(b) Let $\alpha, \beta \in \mathbb{C}$ be fixed nonzero scalars such that $|\alpha/\beta| > 1$. Let $\{\delta_n\}_{n\in\mathbb{N}}$ be the standard basis for ℓ^2, and define $x_0 = \delta_1$ and $x_n = \alpha\delta_n + \beta\delta_{n+1}$ for $n \in \mathbb{N}$. Prove that $\{x_n\}_{n\ge 0}$ is complete and finitely independent in ℓ^2, but is not ω-independent and therefore is not a basis for ℓ^2.

5.6.17. Let $\{x_n\}$ be a sequence in a Hilbert space H. Prove that the following two statements are equivalent.

(a) For each $m \in \mathbb{N}$ we have $x_m \notin \overline{\operatorname{span}}\{x_n\}_{n\ne m}$ (such a sequence is said to be *minimal*).

(b) There exists a sequence $\{y_n\}$ in H such that $\langle x_m, y_n\rangle = \delta_{mn}$ for all $m, n \in \mathbb{N}$ (we say that sequences $\{x_n\}$ and $\{y_n\}$ satisfying this condition are *biorthogonal*).

Show further that, in case these hold, the sequence $\{y_n\}$ is unique if and only if $\{x_n\}$ is complete.

5.6.18. Let $\psi = \chi_{[0,1/2)} - \chi_{[1/2,1)}$ be the Haar wavelet. Compute the Fourier coefficients of ψ and apply the Plancherel Equality to show that $\frac{\pi^2}{8} = \sum_{n=1}^{\infty} \frac{1}{(2n-1)^2}$.

5.6.19. This exercise provides a real-valued analogue of the orthonormal basis $\{e^{2\pi int}\}_{n\in\mathbb{Z}}$ for complex $L^2(\mathbb{T})$ discussed in Example 5.6.6. Let real $L^2(\mathbb{T})$ consist of all real-valued, 1-periodic functions on $\mathbb{R}$ that are square integrable on $[0,1]$. Assuming the fact that $\{e^{2\pi int}\}_{n\in\mathbb{Z}}$ is complete in complex $L^2(\mathbb{T})$, show that $\{1\} \cup \{\sqrt{2}\,\sin 2\pi nt\}_{n\in\mathbb{N}} \cup \{\sqrt{2}\,\cos 2\pi nt\}_{n\in\mathbb{N}}$ forms an orthonormal basis for real $L^2(\mathbb{T})$ if we take $\mathbb{F} = \mathbb{R}$, and forms an orthonormal basis for complex $L^2(\mathbb{T})$ if we take $\mathbb{F} = \mathbb{C}$.

5.6.20. (a) Prove that the Haar system discussed in Remark 5.6.9(a) is an orthonormal basis for $L^2(\mathbb{R})$.

(b) Prove that the Haar system given in Remark 5.6.9(b) is an orthonormal basis for $L^2[0,1]$.

5.6.21. Let $\{x_n\}$ be an orthonormal basis for a separable Hilbert space H. Show that if $\sum \|x_n - y_n\|^2 < 1,$ then $\{y_n\}$ is complete in H. Show that this conclusion need not hold if $\sum \|x_n - y_n\|^2 = 1$.

5.6.22. Let $\{x_n\}$ be an orthonormal sequence in a Hilbert space H. Given a vector $x \in H$, show that $x \in \overline{\mathrm{span}}\{x_n\}$ if and only if $\|x\|^2 = \sum |\langle x, x_n\rangle|^2$.

5.6.23. Let M be a closed subspace of a Hilbert space H. If M is finite dimensional, let $\dim(M)$ be the dimension of M, otherwise set $\dim(M) = \infty$. Let P be the orthogonal projection of H onto M. Show that if $\{e_n\}$ is any orthonormal basis for H, then $\sum \|Pe_n\|^2 = \dim(M)$.

5.6.24. This result is due to Vitali. Let $\{f_n\}$ be an orthonormal sequence in $L^2[a,b]$. Show that $\{f_n\}$ is complete if and only if

$$\sum_{n=1}^{\infty} \left| \int_a^x f_n(t)\,dt \right|^2 = x - a, \qquad x \in [a,b].$$

5.6.25. Use the Vitali criterion (Exercise 5.6.24) to prove that the following two statements are equivalent.

(a) The trigonometric system $\{e^{2\pi i n x}\}_{n\in\mathbb{Z}}$ is complete in $L^2(\mathbb{T})$.

(b) $\displaystyle\sum_{n=1}^{\infty} \frac{1-\cos 2\pi n x}{\pi^2 n^2} = x - x^2$ for $x \in [0,1]$.

Remark: If we assume *Euler's formula*

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6},$$

then statement (b) reduces to

$$\sum_{n=1}^{\infty} \frac{\cos 2\pi n x}{\pi^2 n^2} = x^2 - x + \frac{1}{6}, \qquad x \in [0,1]. \tag{5.15}$$

Conversely, equation (5.15) implies Euler's formula by taking $x = 0$.

5.6.26. This result is due to Dalzell. Let $\{f_n\}$ be an orthonormal sequence in $L^2[a,b]$. Show that $\{f_n\}$ is complete if and only if

$$\sum_{n=1}^{\infty} \int_a^b \left| \int_a^x f_n(t)\,dt \right|^2 = \frac{(b-a)^2}{2}.$$

5.6.27. This result is due to Boas and Pollard. Suppose that $\{f_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for $L^2[a,b]$. Show that there exists a function $m \in L^\infty[a,b]$ such that $\{mf_n\}_{n\ge 2}$ is complete in $L^2[a,b]$.

5.6.28. (a) Let H be a Hilbert space. Show that if H contains an uncountable orthonormal subset, then H is not separable.

(b) Prove that the space $\ell^2(\mathbb{R})$ defined in Example 5.6.11 is a Hilbert space. Given $t \in \mathbb{R}$, define $e_t(t) = 1$ and $e_t(i) = 0$ for $i \neq t$. Show that $\{e_t\}_{t\in\mathbb{R}}$ is a complete uncountable orthonormal system for $\ell^2(\mathbb{R})$, and conclude that $\ell^2(\mathbb{R})$ is nonseparable.

5.6.29. For each $\xi \in \mathbb{R}$, define a function $e_\xi\colon \mathbb{R} \to \mathbb{C}$ by $e_\xi(t) = e^{2\pi i\xi t}$. Let $H = \mathrm{span}\{e_\xi\}_{\xi\in\mathbb{R}}$, i.e., H consists of all finite linear combinations of the functions e_ξ. Show that

$$\langle f, g\rangle \;=\; \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(t)\,\overline{g(t)}\,dt, \qquad f, g \in H,$$

defines an inner product on H, and $\{e_\xi\}_{\xi\in\mathbb{R}}$ is an uncountable orthonormal system in H.

Remark: The completion $\widetilde{H}$ of H is an important nonseparable Hilbert space. In particular, it contains the class of *almost periodic functions* and it plays an important role in Wiener's theory of *generalized harmonic analysis*. Since $\mathrm{span}\{e_\xi\}_{\xi\in\mathbb{R}} = H$ and H is dense in $\widetilde{H}$, $\{e_\xi\}_{\xi\in\mathbb{R}}$ is a complete orthonormal system in $\widetilde{H}$.

Chapter 6
Operator Theory

Throughout this chapter the symbol $\mathbb{F}$ will denote a choice of either the real line $\mathbb{R}$ or the complex plane $\mathbb{C}$.

Much of our previous focus has been on particular vector spaces and on particular vectors in those spaces. Most of those spaces were metric, normed, or inner product spaces. Now we will concentrate on classes of *operators* that transform vectors in one space into vectors in another space. We will be especially interested in operators that map one normed space into another normed space.

6.1 Linear Operators on Normed Spaces

Since a normed space is a vector space, it has a "structure" that a generic metric space need not possess. In this section we will study *linear functions*, which preserve this vector space structure. Linearity is distinct from continuity; a continuous function preserves convergent sequences, while a linear function preserves the operations of vector addition and multiplication by scalars. We will see that we can completely characterize the linear functions that are also continuous in terms of a concept called *boundedness*. First, however, we introduce some terminology and notation.

We often refer to functions on vector spaces as *operators* or *transformations*, especially if they are linear (indeed, some authors restrict the use of these terms solely to linear functions). We will use these three terms interchangeably when dealing with vector spaces, i.e., for us the words *function*, *operator*, and *transformation* all mean the same thing.

Operators on normed spaces are often denoted by capital letters, such as A or L, or by Greek letters, such as μ or ν. Also, it is traditional, and often notationally simpler, to make the use of parentheses optional when denoting the output of a transformation. That is, we often write Ax instead of $A(x)$, although we include parentheses when there is any danger of confusion. Operators are just functions, so all of the standard terminology for functions defined in the Preliminaries applies to operators on vector spaces. Here is some additional terminology that specifically applies to operators on vector spaces.

Definition 6.1.1 (Notation for Operators). Let X and Y be vector spaces, and let $A\colon X \to Y$ be an operator.

(a) A is *linear* if $A(ax+by) = aAx + bAy$ for $x, y \in X$ and $a, b \in \mathbb{F}$.

(b) A is *antilinear* if $A(ax+by) = \overline{a}Ax + \overline{b}Ay$ for $x, y \in X$ and $a, b \in \mathbb{F}$.

(c) The *kernel* or *nullspace* of A is $\ker(A) = \{x \in X : Ax = 0\}$.

(d) The *range* of A is $\mathrm{range}(A) = A(X) = \{Ax : x \in X\}$, and the *rank* of A is the vector space dimension of its range:

$$\mathrm{rank}(A) \;=\; \dim(\mathrm{range}(A)).$$

If $\mathrm{range}(A)$ is finite dimensional, then we say that A is a *finite-rank operator.*

(e) If $Y = \mathbb{F}$, then A is called a *functional.* $\diamondsuit$

The *zero operator* is the function $0\colon X \to Y$ that maps every vector in X to the zero vector in Y. The zero operator and the zero vector in a vector space are both denoted by the symbol 0, i.e., the rule for the zero operator is $0(x) = 0$ for $x \in X$. It should usually be clear from context whether the symbol 0 is meant to denote an operator or a vector.

The proof of the following lemma is assigned as Problem 6.1.4.

Lemma 6.1.2. *Let X, Y be vector spaces. If $A\colon X \to Y$ is a linear operator, then the following statements hold.*

(a) $\ker(A)$ *is a subspace of X, and* $\mathrm{range}(A)$ *is a subspace of Y.*

(b) $A(0) = 0$*, and therefore* $0 \in \ker(A)$.

(c) *A is injective if and only if* $\ker(A) = \{0\}$.

(d) *If A is a linear* bijection*, then its inverse mapping $A^{-1}\colon Y \to X$ is also a linear bijection.* $\diamondsuit$

Example 6.1.3. The space $C_b^1(\mathbb{R})$, is the set of all bounded differentiable functions $f\colon \mathbb{R} \to \mathbb{F}$ whose derivative f' is both continuous and bounded. Let $D\colon C_b^1(\mathbb{R}) \to C_b(\mathbb{R})$ be the derivative operator defined by $Df = f'$ for $f \in C_b^1(\mathbb{R})$. Since the derivative of a sum is the sum of the derivatives, we have

$$D(f+g) \;=\; (f+g)' \;=\; f'+g' \;=\; Df + Dg. \tag{6.1}$$

We also know that differentiation "respects" scalar multiplication, i.e.,

$$D(cf) \;=\; (cf)' \;=\; cf' \;=\; cDf. \tag{6.2}$$

Combining equations (6.1) and (6.2), we see that D is a linear operator.

Note that it is not the functions f and g in this example that are linear but rather the operator D that transforms a function f into its derivative f'. For example, the functions $f(x) = \sin x$ and $g(x) = e^{-x^2}$ both belong to $C_b^1(\mathbb{R})$, but they are not linear. Instead, the *transformation* D is linear because it sends the input $af + bg$ to the output $af' + bg'$.

If $f(x) = c$ is any constant function, then f belongs to $C_b^1(\mathbb{R})$ and $Df = f' = 0$. Hence the kernel of D contains more than just the zero function. That is, $\ker(D) \neq \{0\}$. Since D is linear, Lemma 6.1.2 therefore implies that D is not injective. Indeed, we can see this directly because if f and g are any two constant functions, then $Df = 0 = Dg$.

Thus $\ker(D)$ contains every constant function. On the other hand, since we know that the *only* functions whose derivatives are identically zero are constant functions, it follows that $\ker D$ is exactly the set of all constant functions. In accordance with Lemma 6.1.2, this is a subspace of $C_b^1(\mathbb{R})$; in fact, it is a one-dimensional subspace. $\diamondsuit$

If Y is a normed space, then we sometimes need to consider the closure of the range of A in Y. Since $\mathrm{range}(A)$ is a subspace of Y, the closure of the range is a closed subspace of Y. For simplicity of notation, we denote the closure of the range by

$$\overline{\mathrm{range}}(A) \;=\; \overline{\mathrm{range}(A)}.$$

If the range of A is a closed set, i.e., if $\overline{\mathrm{range}}(A) = \mathrm{range}(A)$, then we say that A *has closed range*.

The domain X and the codomain Y of most of the operators that we will consider in this chapter will be normed spaces. Consequently there will often be several norms in play at any given time. We typically will let the symbols $\|\cdot\|$ denote both the norm on X and the norm on Y, letting context determine the space that the norm is on. For example, if x is a vector in X, then its image Ax is a vector in Y, so in most circumstances it is safe to simply write $\|x\|$ and $\|Ax\|$ since it is clear that $\|x\|$ must denote the norm on X while $\|Ax\|$ must denote the norm on Y. If we should need to emphasize precisely which spaces the norm is to be taken with respect to, then we will write $\|x\|_X$ and $\|Ax\|_Y$.

Problems

6.1.4. Prove Lemma 6.1.2.

6.1.5. Although this chapter is mostly concerned with linear operators, this problem is about an interesting nonlinear operator.

(a) Let

$$M = \{f \in C[0,1] : f(0) = 0, f(1) = 1\}.$$

Prove that M is a closed subset (but not a subspace) of $C[0,1]$, and therefore M is a complete metric space with respect to the uniform metric.

(b) Given $f \in M$, define Af by

$$Af(x) = \begin{cases} \frac{1}{2} f\left(\frac{x}{1-x}\right), & 0 \le x < \frac{1}{2}, \\ 1 - \frac{1}{2} f\left(\frac{1-x}{x}\right), & \frac{1}{2} \le x \le 1. \end{cases}$$

Prove that $Af \in M$, and therefore A is an operator that maps M into M.

(c) Show that

$$\|Af - Ag\|_{\mathrm{u}} \le \frac{\|f-g\|_{\mathrm{u}}}{2}, \qquad f, g \in M.$$

Consequently A is Lipschitz on M.

(d) Use the Banach Fixed Point Theorem to prove that there exists a unique function $m \in M$ such that $Am = m$. This function is called the *Minkowski question mark function* or the *slippery Devil's staircase*.

(e) Set $f_0(x) = x$, and define $f_{n+1} = Af_n$ for $n \in \mathbb{N}$. Prove that f_n converges uniformly to m (this fact was used to generate the approximation to m that appears in Figure 6.1).

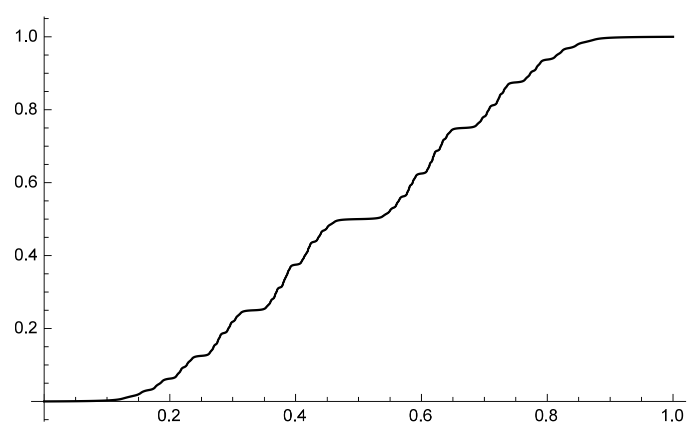


Fig. 6.1 The function f_7 from Problem 6.1.5, approximating the slippery Devil's staircase.

6.2 The Definition of a Bounded Operator

In many circumstances, it is convenient to say that a function $f\colon X \to Y$ is "bounded" on X if the range of f is a bounded subset of Y. For example, this is what we mean when we say that $f(x) = \sin x$ is a bounded function on $\mathbb{R}$, and this type of boundedness is exactly what the subscript "b" stands for in the definition of the space $C_b(X)$.

However, if X and Y are normed spaces and $A\colon X \to Y$ is a *linear* operator then the range of A is a *subspace* of Y. No subspace other than the trivial subspace $\{0\}$ is a bounded set, so if A is not the zero operator then $\text{range}(A)$ is not a bounded subset of Y. Therefore, a nonzero linear operator can never be bounded in the sense that its range is a bounded subset of Y. Instead, we will formulate the idea of "boundedness" for *linear* operators by looking at the relationship between the size of $\|x\|$ and the size of $\|Ax\|$. If there is a limit to how large $\|Ax\|$ can be *in comparison to* $\|x\|$, then we will say that A *is bounded*. Here is the precise definition.

Definition 6.2.1 (Bounded Linear Operator). Let X, Y be normed spaces. We say that a linear operator $A\colon X \to Y$ is *bounded* if there exists a finite constant $K \geq 0$ such that

$$\|Ax\| \leq K\,\|x\|, \qquad \text{all } x \in X. \qquad \diamondsuit \tag{6.3}$$

In particular, if A is bounded then $\|Ax\| \leq K$ for every unit vector x. We give a name to the supremum of $\|Ax\|$ over the unit vectors x.

Definition 6.2.2 (Operator Norm). Let X and Y be normed spaces. If X is nontrivial, then the *operator norm* of a linear operator $A\colon X \to Y$ is

$$\|A\| \;=\; \sup_{\|x\|=1} \|Ax\|. \tag{6.4}$$

If $X = \{0\}$, then we set $\|A\| = 0$. $\diamondsuit$

On those occasions where we need to carefully specify the spaces or norms that we are dealing with, we will write $\|A\|_{X\to Y}$ to denote the operator norm of $A\colon X \to Y$. Another common notation, useful when we want to emphasize the distinction between an operator norm and the norm of a vector, is to write $\|A\|_{\text{op}}$ instead of just $\|A\|$.

Remark 6.2.3. We will implicitly assume in most proofs that X is nontrivial, as the case $X = \{0\}$ is usually easy to deal with. The reader should check that the statements of theorems are valid even if $X = \{0\}$. $\diamondsuit$

We have not yet shown that the operator norm is a norm in any sense, but to illustrate its meaning let

$$S \;=\; \{x \in X : \|x\| = 1\}$$

be the unit sphere in X. The operator norm is the supremum of the norms of the vectors in

$$A(S) \;=\; \{Ax : x \in S\} \;=\; \{Ax : \|x\| = 1\}.$$

Since $A(S)$ is the direct image of the unit sphere under A, the operator norm $\|A\|$ is in a sense the "maximum distortion of the unit sphere" by A (although the supremum in the definition of $\|A\|$ need not be achieved, and we could have $\|A\| = \infty$).

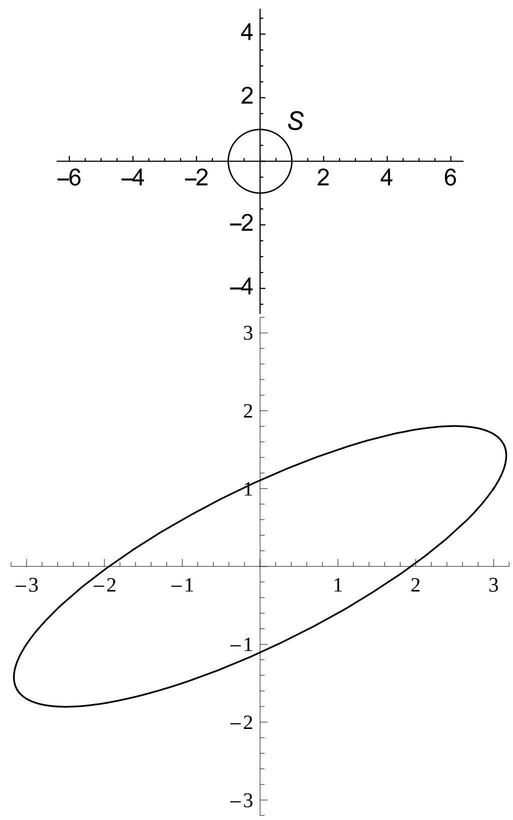


Fig. 6.2 Left: The unit circle $S = \{x \in \mathbb{R}^2 : \|x\|_2 = 1\}$. Right: The image of S under the linear operator $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ whose matrix is given in equation (6.5). The operator norm $\|A\|$ is the length of a semimajor axis of the ellipse $A(S)$.

For example, consider $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$. Every linear operator $A\colon \mathbb{R}^n \to \mathbb{R}^m$ is given by an $m \times n$ matrix, which we also call A. That is, there is an $m \times n$ matrix $A = [a_{ij}]$ such that Ax (the image of x under the operator A) is simply the product of the matrix $A = [a_{ij}]$ with the vector x. If we place the Euclidean norm on both $\mathbb{R}^n$ and $\mathbb{R}^m$ then S is the ordinary unit sphere in $\mathbb{R}^n$, and $A(S)$ is an ellipsoid in $\mathbb{R}^m$ (although it might be degenerate, i.e., one or more of its axes might have zero length). The ellipse $A(S)$ corresponding to $m = n = 2$ and the matrix

$$A = \begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix} \tag{6.5}$$

is shown in Figure 6.2. The operator norm of A is the norm of the vector on $A(S)$ farthest from the origin, i.e., it is the length of a semimajor axis of the ellipsoid.

Here are some properties of bounded linear operators.

Lemma 6.2.4. *If X and Y are normed spaces and $A\colon X \to Y$ is a linear operator, then the following statements hold.*

(a) *A is bounded if and only if $\|A\| < \infty$.*

(b) *$\|Ax\| \le \|A\|\,\|x\|$ for all $x \in X$.*

(c) *If A is bounded, then $\|A\|$ is the smallest constant $K \ge 0$ such that $\|Ax\| \le K\|x\|$ for all $x \in X$.*

(d) *If $X \ne \{0\}$, then $\|A\| = \sup_{\|x\| \le 1} \|Ax\| = \sup_{x \ne 0} \dfrac{\|Ax\|}{\|x\|}$.*

Proof. We will prove part (a), and assign the proofs of the remaining parts as Problem 6.2.5).

(a) $\Rightarrow$. If A is bounded, then there exists a finite constant $K \ge 0$ such that $\|Ax\| \le K\,\|x\|$ for all $x \in X$. Consequently

$$\|A\| \;=\; \sup_{\|x\|=1} \|Ax\| \;\le\; \sup_{\|x\|=1} K\,\|x\| \;=\; K \;<\; \infty.$$

$\Leftarrow$. Suppose that $\|A\| < \infty$. Choose any nonzero vector $x \in X$, and set $y = \frac{x}{\|x\|}$. Then $\|y\| = 1$, so

$$\|A\| \;=\; \sup_{\|z\|=1} \|Az\| \;\ge\; \|Ay\| \;=\; \left\| A\left(\frac{x}{\|x\|}\right) \right\| \;=\; \frac{1}{\|x\|}\,\|Ax\|.$$

Rearranging, we see that $\|Ax\| \le \|A\|\,\|x\|$. This is also true when $x = 0$, so taking $K = \|A\|$ we have $\|Ax\| \le K\,\|x\|$ for every x. Since $K = \|A\|$ is a finite number, this shows that A is bounded. $\square$

Problems

6.2.5. Prove parts (b)–(c) of Lemma 6.2.4.

6.2.6. Let X and Y be normed spaces, and suppose that $A\colon X \to Y$ is a bounded linear operator. Show that if the norm on X is replaced by an equivalent norm and the norm on Y is replaced by an equivalent norm, then A remains bounded with respect to these new norms.

6.2.7. Fix $A \in \mathcal{B}(X)$, where X is a normed space, and suppose that λ is an eigenvalue of A (i.e., $Ax = \lambda x$ for some nonzero $x \in X$). Prove that $|\lambda| \le \|A\|$.

6.2.8. Let $N \in \mathbb{N}$ be a *fixed* integer, and define an operator $A_N\colon \ell^1 \to \ell^1$ by

$$A_N x \;=\; (x_1, \dots, x_N, 0, 0, \dots), \qquad x = (x_k)_{k\in\mathbb{N}} \in \ell^1.$$

Prove the following statements.

(a) A_N is bounded and linear, and $\|A_N\| = 1$.

(b) A_N is not injective, not surjective, and not an isometry.

(c) $\ker(A_N)$ is a proper closed subspace of ℓ^1.

(d) $\mathrm{range}(A_N)$ is a proper closed subspace of ℓ^1.

6.2.9. Define an operator $B\colon \ell^1 \to \ell^1$ by

$$Bx \;=\; \bigl(\tfrac{x_k}{k}\bigr)_{k\in\mathbb{N}} \;=\; \bigl(x_1, \tfrac{x_2}{2}, \tfrac{x_3}{3}, \dots\bigr), \qquad x = (x_k)_{k\in\mathbb{N}} \in \ell^1.$$

Prove the following statements.

(a) B is bounded and linear, and $\|B\| = 1$.

(b) B is injective (and therefore $\ker(B) = \{0\}$), but B is not an isometry.

(c) $\mathrm{range}(B)$ is a proper dense subspace of ℓ^1, but it is *not closed*.

6.2.10. Let X be a Banach space and Y a normed vector space. Suppose that $A\colon X \to Y$ is bounded and linear, and suppose there exists a scalar $c > 0$ such that

$$\|Ax\| \;\ge\; c\,\|x\|, \qquad \text{all } x \in X.$$

Prove that A is injective and $\mathrm{range}(A)$ is a closed subspace of Y. How does this relate to Problems 6.2.8 and 6.2.9?

6.3 Examples

We will give some examples of bounded and unbounded linear operators. First, however, we prove that if the *domain* of a linear operator is finite dimensional, then the operator is automatically bounded.

Theorem 6.3.1. *Let X and Y be normed vector spaces. If X is finite dimensional and $A\colon X \to Y$ is linear, then A is bounded.*

Proof. Because X is finite dimensional, all norms on X are equivalent (see Theorem 5.1.8). As a consequence, A is bounded with respect to one norm on X if and only if it is bounded with respect to any other norm (this is Problem 6.2.6). Therefore we can work with any norm on X that we like, and for this proof we choose the analogue of the ℓ^1-norm on $\mathbb{F}^d$. That is, we fix a particular basis $\mathcal{B} = \{e_1, \dots, e_d\}$ for X, write $x \in X$ as $x = \sum_{k=1}^d c_k(x)\, e_k$, and declare the norm of x to be

$$\|x\|_1 \;=\; \sum_{k=1}^d |c_k(x)|.$$

If $x \in X$, then $Ax = \sum_{k=1}^d c_k(x)\, Ae_k$ since A is linear. Therefore

$$\begin{aligned} \|Ax\| \;&=\; \left\| \sum_{k=1}^d c_k(x)\, Ae_k \right\| \\ &\le\; \sum_{k=1}^d |c_k|\, \|Ae_k\| \qquad \text{(Triangle Inequality)} \\ &\le\; \left(\sum_{k=1}^d |c_k| \right) \left(\max_{k=1,\dots,d} \|Ae_k\| \right) \\ &=\; C\, \|x\|_1, \end{aligned}$$

where $C = \max \|Ae_k\|$. The number C is finite since there are only finitely many vectors Ae_k. Since we have shown that $\|Ax\| \le C\, \|x\|_1$ for every x, the operator A is bounded. $\square$

Thus, all linear operators on *finite-dimensional domains* are bounded. In contrast, we will show that a linear operator on an infinite-dimensional domain can be unbounded, *even if its range is finite dimensional*! We will illustrate this with some examples of operators whose range is the one-dimensional field of scalars $\mathbb{F}$. According to Definition 6.1.1, a function of the form $\mu\colon X \to \mathbb{F}$ is called a *functional* on X. Since the norm on $\mathbb{F}$ is absolute value, the operator norm of such a functional μ is

$$\|\mu\| \;=\; \sup_{\|x\|=1} |\mu(x)|. \tag{6.6}$$

We start with an example of a *bounded* linear functional δ on the infinite-dimensional domain $C_b(\mathbb{R})$. This functional δ will map each element of $C_b(\mathbb{R})$ to a scalar. The input to δ is a continuous bounded function $f \in C_b(\mathbb{R})$, and the output is a scalar $\delta(f)$. This particular functional δ will determine which scalar to output by evaluating f at the origin.

Example 6.3.2. Consider $X = C_b(\mathbb{R})$, the space of all bounded continuous functions $f\colon \mathbb{R} \to \mathbb{F}$. This is a Banach space with respect to the uniform norm

$$\|f\|_{\mathrm{u}} \;=\; \sup_{x\in\mathbb{R}} |f(x)|.$$

Define a functional $\delta\colon C_b(\mathbb{R}) \to \mathbb{F}$ by

$$\delta(f) \;=\; f(0), \qquad f \in C_b(\mathbb{R}).$$

That is, $\delta(f)$ is simply the value that f takes at the origin. For example, if $f(x) = \cos x$ then

$$\delta(f) \;=\; \cos 0 \;=\; 1,$$

if $g(x) = e^{-|x|}$ then

$$\delta(g) \;=\; e^0 \;=\; 1,$$

and if $h(x) = f(x) + g(x) = \cos x + e^{-|x|}$ then

$$\delta(h) \;=\; \cos 0 + e^0 \;=\; 2.$$

Note that δ is not injective since $\delta(f) = \delta(g)$ even though $f \neq g$. On the other hand, δ is surjective (why?). Can you explicitly describe the kernel of δ? Is it a *closed* subspace of $C_b(\mathbb{R})$?

We will show that δ is a bounded linear functional on $C_b(\mathbb{R})$. To show linearity, choose $f,\, g \in C_b(\mathbb{R})$ and scalars $a,\, b \in \mathbb{F}$. Then, by definition of addition of functions,

$$\delta(af+bg) \;=\; (af+bg)(0) \;=\; af(0)+bg(0) \;=\; a\,\delta(f) + b\,\delta(g). \tag{6.7}$$

Hence δ is linear on $C_b(\mathbb{R})$.

To see that δ is bounded, fix any $f \in C_b(\mathbb{R})$. Then

$$|\delta(f)| \;=\; |f(0)| \;\le\; \sup_{x\in\mathbb{R}} |f(x)| \;=\; \|f\|_{\mathrm{u}}.$$

This shows that δ is bounded, and its operator norm satisfies

$$\|\delta\| \;=\; \sup_{\|f\|_{\mathrm{u}}=1} |\delta(f)| \;\le\; \sup_{\|f\|_{\mathrm{u}}=1} \|f\|_{\mathrm{u}} \;=\; 1.$$

In fact, we can determine the exact value of $\|\delta\|$. Consider the *two-sided exponential function* $g(x) = e^{-|x|}$, pictured in Figure 6.3. This function satisfies $g(0) = 1$ and $|g(x)| \le 1$ for all x (any other function $g \in C_b(\mathbb{R})$ that has these two properties would do just as well for this argument). We have $\|g\|_{\mathrm{u}} = 1$, so g is a unit vector in $C_b(\mathbb{R})$. Therefore

$$\|\delta\| \;=\; \sup_{\|f\|_{\mathrm{u}}=1} |\delta(f)| \;\ge\; |\delta(g)| \;=\; |g(0)| \;=\; 1.$$

Hence the operator norm of δ is precisely $\|\delta\| = 1$. $\diamondsuit$

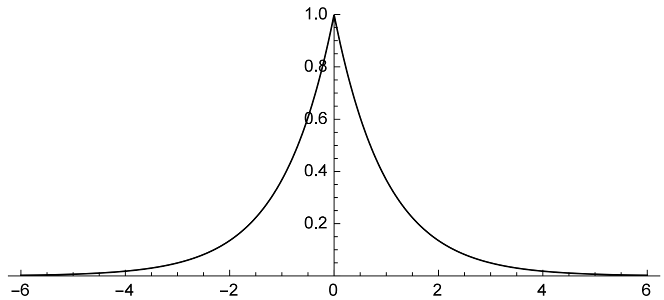


Fig. 6.3 The *two-sided exponential function* $g(x) = e^{-|x|}$.

The functional δ is just one of infinitely many different linear functionals that we can define on $C_b(\mathbb{R})$. For example, we could define $\delta_a \colon C_b(\mathbb{R}) \to \mathbb{F}$ to be evaluation at a instead of evaluation at the origin, i.e., we could set $\delta_a(f) = f(a)$. Here is a different type of bounded linear functional on $C_b(\mathbb{R})$.

Example 6.3.3. Define a functional $\mu \colon C_b(\mathbb{R}) \to \mathbb{F}$ by

$$\mu(f) \;=\; \int_{-\infty}^{\infty} f(x)\, e^{-|x|}\, dx, \qquad f \in C_b(\mathbb{R}). \tag{6.8}$$

The reader should verify that μ is well-defined (i.e., the integral in equation (6.8) actually exists) and that μ is linear. If f is any function in $C_b(\mathbb{R})$, then, by the definition of the uniform norm, $|f(x)| \le \|f\|_{\mathrm{u}}$ for each $x \in \mathbb{R}$. Consequently,

$$\begin{aligned} |\mu(f)| \;=\; \left|\int_{-\infty}^{\infty} f(x)\, e^{-|x|}\, dx\right| \;&\le\; \int_{-\infty}^{\infty} |f(x)|\, e^{-|x|}\, dx \\ &\le\; \int_{-\infty}^{\infty} \|f\|_{\mathrm{u}}\, e^{-|x|}\, dx \\ &=\; \|f\|_{\mathrm{u}} \int_{-\infty}^{\infty} e^{-|x|}\, dx \\ &=\; 2\|f\|_{\mathrm{u}}. \end{aligned}$$

This shows that μ is bounded. Indeed, by taking the supremum over the unit vectors in $C_b(\mathbb{R})$ we see that

$$\|\mu\| \;=\; \sup_{\|f\|_{\mathrm{u}}=1} |\mu(f)| \;\le\; \sup_{\|f\|_{\mathrm{u}}=1} 2\, \|f\|_{\mathrm{u}} \;=\; 2.$$

Can you determine the exact value of the operator norm of μ? $\diamondsuit$

Next we construct a linear functional that is bounded with respect to one norm on its domain, but unbounded when we place a different norm on the domain. The domain for this example is $X = C_b^1(\mathbb{R})$, which was also used earlier in Example 6.1.3.

Example 6.3.4. $C_b^1(\mathbb{R})$ is a Banach space with respect to its "standard norm"

$$\|f\|_{C_b^1} \;=\; \|f\|_{\mathrm{u}} \;+\; \|f'\|_{\mathrm{u}} \;=\; \sup_{x\in\mathbb{R}} |f(x)| \;+\; \sup_{x\in\mathbb{R}} |f'(x)|.$$

Define a functional $\delta' \colon C_b^1(\mathbb{R}) \to \mathbb{F}$ by

$$\delta'(f) \;=\; -f'(0), \qquad f \in C_b^1(\mathbb{R}). \tag{6.9}$$

(The minus sign in the definition of δ' is not important for this example, but it is traditionally included in the definition of δ'.) An argument similar to the calculation in equation (6.7) shows that δ' is linear.

We will prove that δ' is bounded when the norm on $C_b^1(\mathbb{R})$ is $\|\cdot\|_{C_b^1}$. To see this, fix any $f \in C_b^1(\mathbb{R})$. Then

$$|\delta'(f)| \;=\; |f'(0)| \;\le\; \|f'\|_{\mathrm{u}} \;\le\; \|f\|_{\mathrm{u}} + \|f'\|_{\mathrm{u}} \;=\; \|f\|_{C_b^1}.$$

Hence δ' is bounded, and by taking the supremum over all unit vectors f (i.e., unit with respect to the norm $\|\cdot\|_{C_b^1}$) we see that

$$\|\delta'\| \;=\; \sup_{\|f\|_{C_b^1}=1} |\delta'(f)| \;\le\; \sup_{\|f\|_{C_b^1}=1} \|f\|_{C_b^1} \;=\; 1.$$

In fact, we will show that the operator norm of δ' is exactly 1. To do this, consider the function $f_n(x) = \sin nx$. We have $f_n'(x) = n\cos nx$, so

$$\|f_n\|_{C_b^1} \;=\; \|\sin nx\|_{\mathrm{u}} \;+\; \|n\cos nx\|_{\mathrm{u}} \;=\; 1+n.$$

Therefore the function

$$g_n(x) \;=\; \frac{f_n(x)}{\|f_n\|_{C_b^1}} \;=\; \frac{\sin nx}{n+1}$$

satisfies $\|g_n\|_{C_b^1} = 1$, i.e., g_n is one of the unit vectors in $C_b^1(\mathbb{R})$. Since

$$\delta'(g_n) \;=\; -g_n'(0) \;=\; -\frac{n\cos 0}{n+1} \;=\; -\frac{n}{n+1},$$

we see that

$$\|\delta'\| \;=\; \sup_{\|f\|_{C_b^1}=1} |\delta'(f)| \;\ge\; \sup_{n\in\mathbb{N}} |\delta'(g_n)| \;=\; \sup_{n\in\mathbb{N}} \frac{n}{n+1} \;=\; 1.$$

Therefore $\|\delta'\| = 1$.

However, everything changes if we decide to place the uniform norm on $C_b^1(\mathbb{R})$ instead of its standard norm. To see why, let δ' be the same functional as before, i.e., δ' is defined by equation (6.9), but now assume that the norm on $C_b^1(\mathbb{R})$ is $\|\cdot\|_{\mathrm{u}}$ instead of $\|\cdot\|_{C_b^1}$. Consider again the function $f_n(x) = \sin nx$. With respect to the uniform norm we have

$$\|f_n\|_{\mathrm{u}} \;=\; 1,$$

so f_n is a unit vector with respect to $\|\cdot\|_{\mathrm{u}}$. However,

$$\delta'(f_n) \;=\; -f_n'(0) \;=\; -n\cos 0 \;=\; -n,$$

so

$$\|\delta'\| \;=\; \sup_{\|f\|_{\mathrm{u}}=1} |\delta'(f)| \;\ge\; \sup_{n\in\mathbb{N}} |\delta'(f_n)| \;=\; \sup_{n\in\mathbb{N}} n \;=\; \infty.$$

Therefore δ' is *unbounded* when the norm on $C_b^1(\mathbb{R})$ is $\|\cdot\|_{\mathrm{u}}$. $\diamondsuit$

It may seem that we cheated a bit in the preceding example, since $C_b^1(\mathbb{R})$ is not complete with respect to $\|\cdot\|_{\mathrm{u}}$. Could it be the case that all linear functionals on a *complete* normed space are bounded? Unfortunately, the answer is no. Problem 6.3.6 shows that if X is *any* infinite-dimensional normed space, then there exists a linear functional $\mu\colon X \to \mathbb{F}$ that is not bounded (although we need to appeal to the Axiom of Choice to prove the existence of such functionals on arbitrary normed spaces). Boundedness is a desirable property, but we cannot take it for granted when our domain is infinite dimensional. Although we will mostly focus on bounded operators, many important linear operators that arise in mathematics, physics, and engineering are unbounded.

Problems

6.3.5. Let $A = [a_{ij}]$ be an $m \times n$ matrix, which we identify with the linear operator $A\colon \mathbb{F}^n \to \mathbb{F}^m$ that maps x to Ax. Since $\mathbb{F}^n$ is finite dimensional, Theorem 6.3.1 implies that A is bounded with respect to any norm on $\mathbb{F}^n$. However, the exact value of the operator norm of A depends on which norms we choose to place on $\mathbb{F}^n$ and $\mathbb{F}^m$. Prove the following statements.

(a) If the norm on both $\mathbb{F}^n$ and $\mathbb{F}^m$ is $\|\cdot\|_1$ then the operator norm of A is

$$\|A\| \;=\; \max_{j=1,\dots,n} \biggl\{\sum_{i=1}^{m} |a_{ij}|\biggr\}.$$

(b) If the norm on $\mathbb{F}^n$ and $\mathbb{F}^m$ is $\|\cdot\|_\infty$ then the operator norm of A is

$$\|A\| \;=\; \max_{i=1,\dots,m} \biggl\{\sum_{j=1}^{n} |a_{ij}|\biggr\}.$$

(c) If the norm on $\mathbb{F}^n$ is $\|\cdot\|_1$ and the norm on $\mathbb{F}^m$ is $\|\cdot\|_\infty$ then the operator norm of A is

$$\|A\| \;=\; \max_{\substack{i=1,\dots,m\\ j=1,\dots,n}} |a_{ij}|.$$

(d) If the norm on $\mathbb{F}^n$ is $\|\cdot\|_\infty$ and the norm on $\mathbb{F}^m$ is $\|\cdot\|_1$ then the operator norm of A satisfies

$$\|A\| \;\le\; \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|.$$

Show by example that equality holds for some matrices but not for others.

6.3.6. (a) Let c_{00} be the space of "finite sequences" under the ℓ^1-norm. Given $N \in \mathbb{N}$ and $x = (x_1, x_2, \dots, x_N, 0, 0, \dots) \in c_{00}$, define

$$Ax \;=\; (x_1, 2x_2, \dots, Nx_N, 0, 0, \dots).$$

Prove that $A\colon c_{00} \to c_{00}$ is linear but unbounded.

(b) Let X be any infinite-dimensional space. The Axiom of Choice implies that there exists some Hamel basis for X, say $\{x_i\}_{i\in I}$. By dividing each vector by its norm, we may assume that $\|x_i\| = 1$ for each $i \in I$. Let $J_0 = \{j_1, j_2, \dots\}$ be any infinite countable subsequence of I. Define $\mu(x_{j_n}) = n$ for $n \in \mathbb{N}$, and set $\mu(x_i) = 0$ for $i \in I \setminus J_0$. By the definition of a Hamel basis, each nonzero vector $x \in X$ can be written uniquely as $x = \sum_{k=1}^{N} c_k x_{i_k}$ for some indices $i_1, \dots, i_N \in I$ and nonzero scalars $c_1, \dots, c_N$. Use this to extend μ to a linear functional whose domain is X, and show that $\mu\colon X \to \mathbb{F}$ is unbounded.

6.4 Equivalence of Bounded and Continuous Linear Operators

We are especially interested in linear operators that map one normed space into another normed space. We will prove that, for these operators, boundedness and continuity are entirely equivalent.

Theorem 6.4.1 (Boundedness Equals Continuity). *Let X and Y be normed vector spaces. If $A\colon X \to Y$ is a linear operator, then*

$$A \text{ is bounded} \quad\Longleftrightarrow\quad A \text{ is continuous.}$$

Proof. $\Rightarrow$. Suppose that A is bounded, and choose any vector $x \in X$. Suppose that $x_n \to x$ in X. Applying linearity and Lemma 6.2.4, we see that

$$\|Ax_n - Ax\| \;=\; \|A(x_n - x)\| \;\le\; \|A\|\,\|x_n - x\| \;\to\; 0.$$

Thus $Ax_n \to Ax$. As this is true for every $x \in X$, A is continuous.

$\Leftarrow$. Suppose that A is continuous but unbounded. Then

$$\|A\| \;=\; \sup_{\|x\|=1} \|Ax\| \;=\; \infty,$$

so for each $n \in \mathbb{N}$ there must exist a vector $x_n \in X$ such that $\|x_n\| = 1$ but $\|Ax_n\| \geq n$. Setting $y_n = x_n/n$, we see that

$$\|y_n - 0\| \;=\; \|y_n\| \;=\; \frac{\|x_n\|}{n} \;=\; \frac{1}{n} \;\to\; 0 \quad \text{as } n \to \infty.$$

Thus $y_n \to 0$. Since A is continuous and linear, it follows that $Ay_n \to A0 = 0$. By the continuity of the norm, this implies that

$$\lim_{n\to\infty} \|Ay_n\| \;=\; \|0\| \;=\; 0.$$

However, for each $n \in \mathbb{N}$ we have

$$\|Ay_n\| \;=\; \frac{1}{n}\,\|Ax_n\| \;\geq\; \frac{1}{n}\cdot n \;=\; 1,$$

which is a contradiction. Therefore A must be bounded. $\square$

In summary, if X and Y are normed spaces and $A\colon X \to Y$ is linear, then the terms "continuous" and "bounded" are entirely interchangeable when applied to A.

Problems

6.4.2. Given normed spaces X and Y, prove the following statements.

(a) If $A\colon X \to Y$ is continuous, then $\ker(A)$ is a closed subset of X (even if A is not linear).

(b) If $A \in \mathcal{B}(X, Y)$, then $\ker(A)$ is a closed subspace of X.

6.4.3. Let X, Y be normed spaces, and suppose that A, $B\colon X \to Y$ are bounded linear operators. Prove that if $Ax = Bx$ for all x in a dense set $S \subseteq X$, then $A = B$.

6.5 The Space $\mathcal{B}(X, Y)$

If X and Y are normed spaces, then we collect the bounded linear operators that map X into Y to form another space that we call $\mathcal{B}(X, Y)$.

Definition 6.5.1. Let X and Y be normed vector spaces. The set of all bounded linear operators from X into Y is denoted by

$$\mathcal{B}(X,Y) \;=\; \bigl\{A\colon X \to Y \;:\; A \text{ is bounded and linear}\bigr\}.$$

If $X = Y$ then we write $\mathcal{B}(X) = \mathcal{B}(X,X)$. $\diamondsuit$

Since boundedness and continuity are equivalent for linear operators on normed spaces, we can also describe $\mathcal{B}(X,Y)$ as the space of continuous linear operators from X into Y.

The following lemma (whose proof is Problem 6.5.8) shows that the operator norm is a norm on $\mathcal{B}(X,Y)$, and therefore $\mathcal{B}(X,Y)$ is a normed space with respect to the operator norm.

Lemma 6.5.2. *If X and Y are normed spaces, then the operator norm is a norm on $\mathcal{B}(X,Y)$. That is:*

(a) $0 \le \|A\| < \infty$ *for every* $A \in \mathcal{B}(X,Y)$,

(b) $\|A\| = 0$ *if and only if* $A = 0$,

(c) $\|cA\| = |c|\,\|A\|$ *for all* $A \in \mathcal{B}(X,Y)$ *and all scalars* c, *and*

(d) $\|A+B\| \le \|A\| + \|B\|$ *for all* $A,\ B \in \mathcal{B}(X,Y)$. $\diamondsuit$

Next we prove that if Y is complete, then $\mathcal{B}(X,Y)$ is complete as well. The idea of the proof is similar to previous completeness proofs such as Theorems 5.2.4. That is, given a Cauchy sequence we first find a candidate that the sequence appears to converge to, and then prove that the sequence does converge *in norm* to that candidate.

Theorem 6.5.3. *If X is a normed vector space and Y is a Banach space, then $\mathcal{B}(X,Y)$ is a Banach space with respect to the operator norm.*

Proof. Assume that $\{A_n\}_{n\in\mathbb{N}}$ is a sequence of operators in $\mathcal{B}(X,Y)$ that is Cauchy with respect to the operator norm. That is, we assume that for each $\varepsilon > 0$ there is some $N > 0$ such that

$$m, n > N \implies \|A_m - A_n\| < \varepsilon.$$

Fix any vector $x \in X$. Since

$$\|A_m x - A_n x\| \;\le\; \|A_m - A_n\|\,\|x\|,$$

it follows that $\{A_n x\}_{n\in\mathbb{N}}$ is a Cauchy sequence of vectors in Y. Since Y is complete, this sequence must converge. That is, there exists some $y \in Y$ such that $Ax_n \to y$ as $n \to \infty$. Define $Ax = y$. This gives us a candidate limit operator A, and the reader should check that this mapping A is linear.

Every Cauchy sequence in a normed space is bounded, so we must have $C = \sup \|A_n\| < \infty$. If we fix $x \in X$, then $A_n x \to Ax$ by the definition of A, and therefore

$$
\begin{aligned}
\|Ax\| &= \lim_{n\to\infty} \|A_n x\| && \text{(continuity of the norm)}\\
&\le \sup_{n\in\mathbb{N}} \|A_n x\| \\
&\le \sup_{n\in\mathbb{N}} \|A_n\|\,\|x\| && \text{(Lemma 6.2.4)}\\
&= C\,\|x\| && \text{(definition of } C\text{)}.
\end{aligned}
$$

Taking the supremum over all unit vectors x, we see that A is bounded and $\|A\| \le C$.

Finally, we must show that $A_n \to A$ in operator norm. Fix any $\varepsilon > 0$. Since $\{A_n\}_{n\in\mathbb{N}}$ is Cauchy, there exists an integer $N > 0$ such that

$$m, n > N \quad \Longrightarrow \quad \|A_m - A_n\| < \frac{\varepsilon}{2}.$$

Choose any vector $x \in X$ with $\|x\| = 1$. Since $A_m x \to Ax$ as $m \to \infty$, there exists some $m > N$ such that

$$\|Ax - A_m x\| \;<\; \frac{\varepsilon}{2}.$$

In fact, this inequality will hold for all large enough m, but for this proof we only need one m. Using that fixed m, given any $n > N$ we compute that

$$
\begin{aligned}
\|Ax - A_n x\| &\le \|Ax - A_m x\| + \|A_m x - A_n x\|\\
&\le \|Ax - A_m x\| + \|A_m - A_n\|\,\|x\|\\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2}\cdot 1 \;=\; \varepsilon.
\end{aligned}
$$

Taking the supremum over all unit vectors x, it follows that

$$\|A - A_n\| \;=\; \sup_{\|x\|=1} \|Ax - A_n x\| \;\le\; \varepsilon, \qquad \text{all } n > N.$$

Hence $A_n \to A$ in operator norm, and therefore $\mathcal{B}(X,Y)$ is complete. □

Since the elements of $\mathcal{B}(X,Y)$ are functions, we know how to add elements of $\mathcal{B}(X,Y)$ and how to multiply them by scalars. However, there is a also third natural operation that we can perform on functions—we can compose them if their domains and codomains match appropriately. Given operators $A \in \mathcal{B}(X,Y)$ and $B \in \mathcal{B}(Y,Z)$, we typically denote the composition of B with A by BA. That is, $BA\colon X \to Z$ is the function whose rule is

$$(BA)(x) \;=\; B(A(x)), \qquad x \in X.$$

The next lemma (whose proof is Problem 6.5.8) shows that the composition of bounded operators is bounded.

Lemma 6.5.4. *Let X, Y, Z be normed spaces. If $A \in \mathcal{B}(X,Y)$ and $B \in \mathcal{B}(Y,Z)$, then $BA \in \mathcal{B}(X,Z)$ and we have the following* submultiplicative property *of the operator norm:*

$$\|BA\| \;\le\; \|B\|\,\|A\|. \qquad \diamondsuit \tag{6.10}$$

In particular, taking $X = Y$ we see that

$$A, B \;\in\; \mathcal{B}(X) \;\implies\; AB, BA \in \mathcal{B}(X).$$

Therefore $\mathcal{B}(X)$ is closed under compositions of operators, and furthermore composition is submultiplicative in the sense that $\|AB\| \le \|A\|\,\|B\|$.

6.5.1 Densely Defined Operators

We often encounter operators whose domain is only a dense subset of a normed space X instead of all of X. For example, suppose that we have an application that deals with operators on $X = C_b(\mathbb{R})$, but for some reason we want to consider the operator δ' that was introduced earlier. The rule for that operator is

$$\delta'(f) \;=\; -f'(0),$$

but this is only well-defined if f is differentiable at the origin. Therefore we cannot take $X = C_b(\mathbb{R})$ to be the domain of the operator δ'. On the other hand, δ' is well-defined on many dense subspaces of $C_b(\mathbb{R})$, such as $C_b^1(\mathbb{R})$. Hence even though δ' is not defined on the entire domain $C_b(\mathbb{R})$, in some sense it is defined on "most" of $C_b(\mathbb{R})$. We introduce some terminology to deal with this type of situation.

Notation 6.5.5 (Densely Defined Operators). Let X and Y be normed linear spaces.

(a) If S is a dense subspace of X, then an operator $A\colon S \to Y$ whose domain is S is said to be *densely defined on* X.

(b) We say that A *is a densely defined mapping of* X *into* Y, or simply $A\colon X \to Y$ *is densely defined on* X, if there is some dense subspace S of X such that $A\colon S \to Y$ is a well-defined operator. $\diamondsuit$

In this terminology, the operator δ' is densely defined on $C_b(\mathbb{R})$. Here is a related example.

Example 6.5.6. Consider the differentiation operator D defined by $Df = f'$. This operator is not defined on the space $L^p(\mathbb{R})$, but there are many dense subspaces of $L^p(\mathbb{R})$ on which D is well-defined, such as $S = L^p(\mathbb{R}) \cap C^1(\mathbb{R})$, or $S = C_c^\infty(\mathbb{R})$. Hence D is densely defined on $L^p(\mathbb{R})$.

Note that if we take $S = C_c^\infty(\mathbb{R})$ to be the domain of the operator D, then D maps this subset of $L^p(\mathbb{R})$ back into $L^p(\mathbb{R})$:

$$f \in C_c^\infty(\mathbb{R}) \quad \Longrightarrow \quad Df \;=\; f' \;\in\; C_c^\infty(\mathbb{R}) \;\subseteq\; L^p(\mathbb{R}).$$

Following Definition 6.5.5(b), we therefore say that "D is a densely defined mapping of $L^p(\mathbb{R})$ into itself." Abusing notation somewhat, we may even write that "$D\colon L^p(\mathbb{R}) \to L^p(\mathbb{R})$ is densely defined," or refer to "the differentiation operator D on $L^p(\mathbb{R})$," with the understanding that D is only defined on a dense subspace of $L^p(\mathbb{R})$. $\diamondsuit$

Suppose that A is a densely defined mapping of X into Y, and let S be a dense subspace of X on which A is defined. It can be very important to know whether A has a "good" extension to an operator whose domain is the entire space X. The following exercise shows that if $A\colon S \to Y$ is bounded and linear, and if Y is complete, then A has a unique extension to a bounded linear operator whose domain is X. Unfortunately, it is not always possible to find a "natural" extension of a densely defined unbounded linear operator.

Exercise 6.5.7 (Extension of Bounded Operators). Let S be a dense subspace of a normed space X, and assume that Y is a Banach space. Suppose $A \in \mathcal{B}(S,Y)$, i.e., A is a bounded linear mapping of S into Y. Set $R = \mathrm{range}(A)$, and let $\overline{R} = \overline{\mathrm{range}}(A)$ be the closure of the range in Y. Prove the following statements.

(a) If $f \in X$ and $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of vectors in S such that $f_n \to f$, then $\{Af_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in Y. Consequently there is a vector $g \in Y$ such that $Af_n \to g$. Moreover, this vector g is independent of the choice of sequence $\{f_n\}_{n\in\mathbb{N}}$.

(b) There exists a unique operator $\widetilde{A} \in \mathcal{B}(X,Y)$ whose restriction to S is A.

(c) The operator norm of $\widetilde{A}$ equals the operator norm of A, i.e.,

$$\|\widetilde{A}\| \;=\; \|A\|. \qquad \diamondsuit$$

If $A \in \mathcal{B}(X,Y)$ satisfies the hypotheses of Exercise 6.5.7, then we often denote the extended operator by the same symbol A, instead of writing $\widetilde{A}$.

Problems

6.5.8. Prove Lemmas 6.5.2, and 6.5.4.

6.5.9. Let X, Y be normed spaces, and suppose that the series $\sum_{n=1}^\infty x_n$ converges in X. Show that if $A \in \mathcal{B}(X,Y)$, then $\sum_{n=1}^\infty Ax_n$ converges in Y.

6.5.10. Let X, Y be normed spaces. Suppose that $A_n \in \mathcal{B}(X,Y)$ and the series $A = \sum_{n=1}^{\infty} A_n$ converges in operator norm. Prove that for each $x \in X$ we have

$$Ax \;=\; \sum_{n=1}^{\infty} A_n x,$$

where this series converges in the norm of Y.

6.5.11. Let X be a Banach space, and let $A \in \mathcal{B}(X)$ be given. Let $A^0 = I$ be the identity map on X, and for $n > 0$ let $A^n = A \cdots A$ (the composition of A with itself n times). Prove the following statements.

(a) The series

$$e^A \;=\; \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

converges absolutely in operator norm, and $\|e^A\| \le e^{\|A\|}$.

(b) For each $x \in X$ we have

$$e^A(x) \;=\; \sum_{k=0}^{\infty} \frac{A^k x}{k!},$$

where the series on the right converges absolutely in X.

(c) If A, $B \in \mathcal{B}(X)$ and $AB = BA$, then $e^A e^B = e^B e^A$.

6.5.12. Let X be a normed space and let Y be a Banach space. Assume that S is a dense subspace of X, and $A\colon S \to Y$ is a bounded linear operator. Prove the following statements.

(a) If $x \in X$ and $\{x_n\}_{n\in\mathbb{N}}$ is a sequence of vectors in S such that $x_n \to x$, then $\{Ax_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in Y. Consequently there is a vector $y \in Y$ such that $Ax_n \to y$. Moreover, this vector y is independent of the choice of sequence $\{x_n\}_{n\in\mathbb{N}}$, and therefore we can define $\widetilde{A}\colon X \to Y$ by $\widetilde{A}x = y$.

(b) $\widetilde{A}$ is bounded and linear, and its operator norm equals the operator norm of A, i.e.,

$$\|A\| \;=\; \sup_{\substack{x\in S,\\ \|x\|=1}} \|Ax\| \;=\; \sup_{\substack{x\in X,\\ \|x\|=1}} \|\widetilde{A}x\| \;=\; \|\widetilde{A}\|.$$

Further, $\widetilde{A}$ is an extension of A, i.e., $\widetilde{A}x = Ax$ for all $x \in S$.

(c) $\widetilde{A}$ is the *unique* operator in $\mathcal{B}(X,Y)$ whose restriction to S is A.

6.6 Isometries and Isomorphisms

We give the following name to operators that *preserve the norm of vectors*.

Definition 6.6.1 (Isometries). Let X and Y be normed vector spaces. An operator $A\colon X \to Y$ that satisfies

$$\|Ax\| \;=\; \|x\| \quad \text{for all } x \in X$$

is called an *isometry*, or is said to be *norm-preserving*. $\diamondsuit$

An isometry need not be linear. For example, if X is a normed space and we define $A\colon X \to \mathbb{F}$ by $Ax = \|x\|$, then A is an isometry, but A is nonlinear if $X \neq \{0\}$ (why?).

However, we will mostly be interested in linear isometries. The next lemma gives some of their properties.

Lemma 6.6.2. *If X, Y are normed spaces and $A\colon X \to Y$ is a linear isometry, then the following statements hold.*

(a) *A is bounded, A is injective, and $\|A\| = 1$ (except when X is trivial, in which case $\|A\| = 0$).*

(b) *If X is a Banach space, then A has closed range.*

Proof. (a) If $Ax = 0$, then, since A is norm-preserving, $\|x\| = \|Ax\| = 0$. Therefore $x = 0$, so we have $\ker(A) = \{0\}$. Hence A is injective by Lemma 6.1.2. Further, if X is nontrivial (so unit vectors exist), then

$$\|A\| \;=\; \sup_{\|x\|=1} \|Ax\| \;=\; \sup_{\|x\|=1} \|x\| \;=\; 1.$$

(b) Suppose that vectors $y_n \in \operatorname{range}(A)$ converge to a vector $y \in Y$. By the definition of the range, for each n there is some vector $x_n \in X$ such that $Ax_n = y_n$. As A is linear and isometric, we therefore have

$$\|x_m - x_n\| \;=\; \|A(x_m - x_n)\| \;=\; \|Ax_m - Ax_n\| \;=\; \|y_m - y_n\|. \tag{6.11}$$

But $\{y_n\}_{n\in\mathbb{N}}$ is Cauchy in Y (because it converges), so equation (6.11) implies that $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy in X. Since X is complete, there is some $x \in X$ such that $x_n \to x$. As A is bounded and therefore continuous, this implies that $Ax_n \to Ax$. By assumption we also have $Ax_n = y_n \to y$, so the uniqueness of limits implies that $y = Ax$. Thus $y \in \operatorname{range}(A)$, so $\operatorname{range}(A)$ is closed. $\square$

We will exhibit an operator that satisfies $\|A\| = 1$ but is not an isometry.

Example 6.6.3 (Left Shift). Define $L\colon \ell^2 \to \ell^2$ by the rule

$$Lx \;=\; (x_2, x_3, \dots), \qquad x = (x_1, x_2, \dots) \in \ell^2.$$

For example, if $x = \bigl(1, \frac{1}{2}, \frac{1}{3}, \dots\bigr)$, then $Lx = \bigl(\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\bigr)$. According to Problem 6.6.10, this *left-shift operator* L is bounded, linear, surjective, and satisfies $\|L\| = 1$, but it is not injective and is not an isometry. $\diamondsuit$

Although an isometry must be injective, it need not be surjective. Here is an example.

Example 6.6.4 (Right Shift). Define $R\colon \ell^2 \to \ell^2$ by

$$Rx = (0, x_1, x_2, x_3, \dots), \qquad x = (x_1, x_2, \dots) \in \ell^2.$$

According to Problem 6.6.10, this *right-shift operator* is bounded, linear, injective, and is an isometry. However, it is not surjective since, for example, the first standard basis vector $\delta_1 = (1, 0, 0, \dots)$ does not belong to the range of R. $\diamondsuit$

Thus an isometry need not be surjective, but it will always be injective. A *surjective* linear isometry $A\colon X \to Y$ has a number of remarkable properties, including the following.

- A is injective because it is an isometry, and it is surjective by assumption. Consequently A is a bijection, and therefore it defines a one-to-one correspondence between the elements of X and the elements of Y. Hence *A preserves cardinality.*
- *A preserves the vector space operations* because it is linear.
- A is continuous because it is bounded. Consequently *A preserves convergent sequences*, i.e., if $x_n \to x$ in X then $Ax_n \to Ax$ in Y.
- *A preserves the norms of vectors* because it is an isometry.
- Since A is linear and norm-preserving, $\|Ax - Ay\| = \|A(x-y)\| = \|x-y\|$ for all $x, y \in X$. Thus *A preserves distances between vectors.*
- If $y \in Y$ then $y = Ax$ for some unique $x \in X$, and therefore

  $$\|A^{-1}y\| = \|x\| = \|Ax\| = \|y\|.$$

  Hence A^{-1} is also a surjective linear isometry.
- Because A is continuous, the inverse image of an open set is open, and therefore *A^{-1} maps open sets to open sets.* Likewise, because A^{-1} is continuous, *A maps open sets to open sets.* Hence *A preserves the topology.*

Thus, if A is a surjective linear isometry from X to Y then A bijectively identifies the elements of X with those of Y in a way that preserves the vector space operations, preserves the topology of the spaces, and preserves the norms of elements (and A^{-1} does the same in the reverse direction). In essence, X and Y are the "same" space with their elements renamed. We introduce some terminology for this situation.

Definition 6.6.5 (Isometric Isomorphism). Let X and Y be normed vector spaces.

(a) If $A\colon X \to Y$ is linear isometry that is surjective, then A is called an *isometric isomorphism* of X onto Y.

(e) If there exists an isometric isomorphism $A\colon X \to Y$, then we say that X and Y are *isometrically isomorphic*, and in this case we write $X \cong Y$. $\diamondsuit$

The left-shift and right-shift operators L and R discussed above are not isometric isomorphisms. Here is an example of an isometric isomorphism.

Example 6.6.6. Suppose that $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for an infinite-dimensional separable Hilbert space H. If x is a vector in H, then Bessel's Inequality (Theorem 5.6.3) implies that

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

Therefore the sequence of inner products $\bigl(\langle x, e_n\rangle\bigr)_{n\in\mathbb{N}}$ belongs to ℓ^2. Define $T\colon H \to \ell^2$ by

$$Tx \;=\; \bigl(\langle x, e_n\rangle\bigr)_{n\in\mathbb{N}}, \qquad x \in \ell^2.$$

The reader should check that the fact that the inner product is linear in the first variable implies that T is a linear operator.

By the Plancherel Equality,

$$\|Tx\|_2^2 \;=\; \sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2 \;=\; \|x\|^2.$$

Therefore T is an isometry.

To show that T is surjective, choose any sequence $c = (c_n)_{n\in\mathbb{N}} \in \ell^2$. By Theorem 5.6.3, the series

$$x \;=\; \sum_{n=1}^{\infty} c_n x_n$$

converges, and $c_n = \langle x, e_n\rangle$ for every n. Hence $Tx = c$, so T is surjective.

Thus T is an isometric isomorphism of H onto ℓ^2, so $H \cong \ell^2$. $\diamondsuit$

Example 6.6.6 shows that *every infinite-dimensional separable Hilbert space is isometrically isomorphic to* ℓ^2. As a consequence (why?), if H and K are any two infinite-dimensional separable Hilbert spaces, then $H \cong K$. If H and K are finite-dimensional Hilbert spaces, then $H \cong K$ if and only if H and K have the same dimension (see Problem 6.6.11).

The next definition introduces a class of operators that includes the isometric isomorphisms, but is more general in the sense that the operators need not be norm-preserving.

Definition 6.6.7 (Topological Isomorphism). Let X and Y be normed spaces. If $A\colon X \to Y$ is a bijective linear operator such that both A and A^{-1} are continuous, then we say that A is a *topological isomorphism*. $\diamondsuit$

For example, let A be an invertible $n \times n$ matrix, and identify A with the mapping $A\colon \mathbb{F}^n \to \mathbb{F}^n$ that takes x to Ax. This is a linear map, and it is a bijection since A is invertible. Since $\mathbb{F}^n$ is finite-dimensional, Theorem 6.3.1 implies that both A and A^{-1} are bounded, so A is a topological isomorphism.

Here is an example of a topological isomorphism on an infinite-dimensional normed space.

Example 6.6.8. Let $X = Y = C_b(\mathbb{R})$, set $\phi(x) = 1 + e^{-|x|}$ for $x \in \mathbb{R}$, and define $A\colon C_b(\mathbb{R}) \to C_b(\mathbb{R})$ by

$$Af \;=\; f\phi, \qquad f \in C_b(\mathbb{R}).$$

That is, Af is simply the pointwise product of f and ϕ:

$$(Af)(x) \;=\; f(x)\left(1+e^{-|x|}\right) \;=\; f(x) \;+\; f(x)\,e^{-|x|}, \qquad x \in \mathbb{R}.$$

Since ϕ is bounded and continuous, the product $f\phi$ is also bounded and continuous and therefore belongs to $C_b(\mathbb{R})$. Hence A does in fact map $C_b(\mathbb{R})$ into itself, and the reader should check that it is linear.

Note that ϕ is a bounded function; in fact, its uniform norm is $\|\phi\|_{\mathrm{u}} = 2$. This implies that A is bounded map, because for each $f \in C_b(\mathbb{R})$ we have

$$\|Af\|_{\mathrm{u}} \;=\; \|f\phi\|_{\mathrm{u}} \;=\; \sup_{x\in\mathbb{R}} |f(x)\phi(x)| \;\le\; 2\,\sup_{x\in\mathbb{R}} |f(x)| \;=\; 2\,\|f\|_{\mathrm{u}}.$$

Taking the supremum over all unit vectors f, we see that $\|A\| \le 2$. In fact, according to Problem 6.6.12, the operator norm of A is precisely $\|A\| = 2$.

If $Af = 0$, then $f(x)\phi(x) = 0$ for every x. Since $\phi(x)$ is always nonzero, this implies that $f(x) = 0$ for every x. Hence $f = 0$, and therefore $\ker(A) = \{0\}$, i.e., the kernel of A only contains the zero function. Since A is linear, this implies that A is injective.

Now, not only is ϕ a bounded function, but it is also "bounded away from zero," because

$$1 \;<\; \phi(x) \;\le\; 2, \qquad \text{all } x \in \mathbb{R}. \tag{6.12}$$

Hence its reciprocal $1/\phi$ is a bounded, continuous function. Therefore, if h is a function in $C_b(\mathbb{R})$ then $f = h/\phi$ is continuous and bounded. This function $f \in C_b(\mathbb{R})$ satisfies $Af = f\phi = h$, so A is surjective.

Thus A is a bijection that maps $C_b(\mathbb{R})$ onto itself. The inverse function is given by $A^{-1}f = f/\phi$. Since $\phi(x) > 1$ for every x, we see that A^{-1} is bounded, because

$$\|A^{-1}f\|_{\mathrm{u}} \;=\; \|f/\phi\|_{\mathrm{u}} \;=\; \sup_{x\in\mathbb{R}} \left|\frac{f(x)}{\phi(x)}\right| \;\le\; \sup_{x\in\mathbb{R}} \left|\frac{f(x)}{1}\right| \;=\; \|f\|_{\mathrm{u}}.$$

This shows that $\|A^{-1}\| \le 1$, and the reader should use Problem 6.6.12 to prove that $\|A^{-1}\| = 1$.

In summary, A is a bounded linear bijection and A^{-1} is bounded as well. Therefore A is a topological isomorphism (and so is A^{-1}). However, A is not isometric (why?), so it is not an isometric isomorphism. What happens in this example if we replace $\phi(x) = 1 + e^{-|x|}$ by $\phi(x) = e^{-|x|}$? Is A still bounded? Is it injective? Is it a topological isomorphism? $\square$

Although we will not prove it, the *Inverse Mapping Theorem* is an important result from the field of *functional analysis* that states that if X and Y are Banach spaces and $A\colon X \to Y$ is a bounded linear bijection, then A^{-1} is automatically bounded and therefore A is a topological isomorphism. Thus, when X and Y are *complete*, every continuous linear bijection is actually a topological isomorphism. For one proof of the Inverse Mapping Theorem (which is a corollary of the *Open Mapping Theorem*), see [Heil11, Thm. 2.29].

6.6.1 Orthogonal Projections

Orthogonal projection opertors are extremely useful tools for the analysis of Hilbert spaces. Recall from Definition 5.5.8 that if M is a closed subspace M of a Hilbert space H, then the *orthogonal projection of a vector* $f \in H$ *onto* M is the unique vector $p \in M$ that is closest to f. Equivalently, the orthogonal projection of f is the unique vector $p \in M$ that satisfies

$$f - p \;\in\; M^{\perp}.$$

We are especially interested here in the *mapping* that takes f to its orthogonal projection p. Following Definition 5.5.8, this operator is called the *orthogonal projection of* H *onto* M. The next exercises identifies some of the basic properties of orthogonal projections.

Exercise 6.6.9 (Properties of Orthogonal Projections). Let M be a closed subspace of a Hilbert space H, and let P be the orthogonal projection of H onto M. Prove that the following statements hold.

(a) Pf is the unique vector in M such that $f - Pf \in M^{\perp}$.

(b) For each vector $f \in H$,

$$\|f - Pf\| \;=\; \operatorname{dist}(f, M) \;=\; \inf\bigl\{\|h - Ph\| : h \in H\bigr\}.$$

(c) P is linear and $\|Pf\| \le \|f\|$ for every $f \in H$.

(d) If $M = \{0\}$ then $P = 0$, and if $M = H$ then $P = I$. In any other case, $\|P\| = 1$ but P is not an isometry.

(e) $P^2 = P$ (an operator that equals its own square is said to be *idempotent*).

(f) $\langle Pf, g\rangle = \langle f, Pg\rangle$ for all $f,\, g \in H$ (using terminology we will introduce later, this says that P is *self-adjoint*).

(g) $\ker(P) = M^{\perp}$ and $\operatorname{range}(P) = M$.

(h) $I - P$ is the orthogonal projection of H onto $M^{\perp}$. $\diamondsuit$

Problems

6.6.10. Prove the statements made about the left-shift and right-shift operators in Examples 6.6.3 and 6.6.4.

6.6.11. Given finite-dimensional Hilbert spaces H and K, prove that $H \cong K$ if and only if $\dim(H) = \dim(K)$.

6.6.12. Given $\phi \in C_b(\mathbb{R})$, prove that

$$M_\phi f \;=\; f\phi, \qquad f \in C_b(\mathbb{R}),$$

defines a bounded linear mapping $M_\phi \colon C_b(\mathbb{R}) \to C_b(\mathbb{R})$, and its operator norm is $\|M_\phi\| = \|\phi\|_{\mathrm{u}}$.

6.6.13. Let $\phi \colon \mathbb{R} \to \mathbb{C}$ be a measurable function and fix $1 \le p \le \infty$.

(a) Show that if $\phi \in L^\infty(\mathbb{R})$, then the rule

$$M_\phi f \;=\; f\phi, \qquad f \in L^p(\mathbb{R}),$$

defines a bounded linear mapping $M_\phi \colon L^p(\mathbb{R}) \to L^p(\mathbb{R})$, and its operator norm is precisely $\|M_\phi\| = \|\phi\|_\infty$.

(b) Conversely, show that if $f\phi \in L^p(\mathbb{R})$ for every $f \in L^p(\mathbb{R})$, then we must have $\phi \in L^\infty(\mathbb{R})$. $\diamondsuit$

Hint: If $p < \infty$ and $\phi \notin L^\infty(\mathbb{R})$, then infinitely many $E_k = \{k \le |\phi| < k+1\}$ must have positive measure.

6.6.14. Fix $1 \le p \le \infty$. Given a function $\phi \in L^\infty(\mathbb{R})$, let M_ϕ be the operator defined in Exercise 6.6.13.

(a) Determine a necessary and sufficient condition on ϕ that implies that $M_\phi \colon L^p(\mathbb{R}) \to L^p(\mathbb{R})$ is injective.

(b) Determine a necessary and sufficient condition on ϕ that implies that $M_\phi \colon L^p(\mathbb{R}) \to L^p(\mathbb{R})$ is surjective.

(c) Show directly that if M_ϕ is injective but not surjective, then the inverse mapping $M_\phi^{-1} \colon \mathrm{range}(M_\phi) \to L^p(\mathbb{R})$ is unbounded.

6.6.15. Given $a \in \mathbb{R}$, let $T_a \colon C_b(\mathbb{R}) \to C_b(\mathbb{R})$ be the operator that translates a function right by a units. Explicitly, if $f \in C_b(\mathbb{R})$, then $T_a f$ is the function given by $(T_a f)(x) = f(x-a)$. Prove that T_a is an isometric isomorphism. What is $(T_a)^{-1}$? Find all eigenvalues of T_a, i.e., find all scalars $\lambda \in \mathbb{F}$ for which there exists some nonzero function $f \in C_b(\mathbb{R})$ such that $T_a f = \lambda f$.

6.6.16. Let X, Y, and Z be normed spaces. Given topological isomorphisms $A \colon X \to Y$ and $B \colon Y \to Z$, prove that $BA \colon X \to Z$ is a topological isomorphism. Likewise show that the composition of isometries is an isometry, and the composition of isometric isomorphisms is an isometric isomorphism.

6.6.17. Let X and Y be normed spaces, and assume that $A\colon X \to Y$ is a linear operator. Prove the following statements.

(a) If A is surjective and there exist constants c, $C > 0$ such that

$$c\,\|x\| \;\le\; \|Ax\| \;\le\; C\,\|x\|, \qquad \text{all } x \in X,$$

then A is a topological isomorphism.

(b) If A is a topological isomorphism, then

$$\frac{\|x\|}{\|A^{-1}\|} \;\le\; \|Ax\| \;\le\; \|A\|\,\|x\|, \qquad \text{all } x \in X.$$

6.6.18. Let $A\colon X \to Y$ be a topological isomorphism of a normed space X onto a normed space Y. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence of vectors in X. Prove that $\{x_n\}_{n\in\mathbb{N}}$ is a complete sequence in X if and only if $\{Ax_n\}_{n\in\mathbb{N}}$ is a complete sequence in Y.

6.6.19. Let X be a Banach space, and let $A \in \mathcal{B}(X)$ be fixed. Define $A^0 = I$ (the identity operator on X).

(a) Show that if $\|A\| < 1$, then $I - A$ is a topological isomorphism and

$$(I-A)^{-1} \;=\; \sum_{n=0}^{\infty} A^n,$$

where this series (called a *Neumann series* for $(I-A)^{-1}$) converges in operator norm.

(b) Suppose that $B \in \mathcal{B}(X)$ is a topological isomorphism. Prove that if $\|A - B\| < 1/\|B^{-1}\|$, then A is also a topological isomorphism.

(c) Prove that the set of topological isomorphisms that map X onto itself is an open subset of $\mathcal{B}(X)$.

6.6.20. Let $A\colon X \to Y$ be a topological isomorphism that maps a normed space X onto a normed space Y. Prove that a sequence $\{f_n\}_{n\in\mathbb{N}}$ is complete in X if and only if $\{Af_n\}_{n\in\mathbb{N}}$ is complete in Y (recall that a sequence $\{f_n\}_{n\in\mathbb{N}}$ is *complete* in X if $\operatorname{span}\{f_n\}_{n\in\mathbb{N}}$ is dense in X).

6.6.21. This problem will show that any incomplete normed space can be naturally embedded into a larger normed space that is complete.

Let X be a normed linear space that is not complete. Let $\mathcal{C}$ be the set of all Cauchy sequences in X, and define a relation $\sim$ on $\mathcal{C}$ by declaring that $\{f_n\} \sim \{g_n\}$ if $\lim_{n\to\infty} \|f_n - g_n\| = 0$.

(a) Prove that $\sim$ is an equivalence relation on $\mathcal{C}$.

(b) Let $[f_n] = \big\{\{g_n\} : \{g_n\} \sim \{f_n\}\big\}$ denote the equivalence class of $\{f_n\}$ with respect to the relation $\sim$. Let $\widetilde{X}$ be the set of all possible equivalence classes $[f_n]$. Define $\|[f_n]\|_{\widetilde{X}} = \lim_{n\to\infty} \|f_n\|$. Prove that $\|\cdot\|_{\widetilde{X}}$ is well-defined and is a norm on $\widetilde{X}$.

(c) Given $f \in X$, let $[f]$ denote the equivalence class of the Cauchy sequence $\{f, f, f, \dots\}$. Prove that $T(f) = [f]$ is an isometric map of X into $\widetilde{X}$. Also show that $T(X)$ is a dense subspace of $\widetilde{X}$ (so, in the sense of identifying of X with $T(X)$, we can consider X to be a dense subspace of $\widetilde{X}$).

(d) Prove that $\widetilde{X}$ is a Banach space with respect to the norm $\|\cdot\|_{\widetilde{X}}$. We call $\widetilde{X}$ the *completion* of X.

(e) Prove that $\widetilde{X}$ is unique in the sense that if Y is a Banach space and $U\colon X \to Y$ is a linear isometry such that $U(X)$ is dense in Y, then there exists an isometric isomorphism $V\colon Y \to \widetilde{X}$.

(f) Prove that the completion $\widetilde{H}$ of an inner product space H is a Hilbert space with respect to an inner product that extends the inner product on H.

6.7 Infinite Matrices

Suppose that A is an infinite matrix, i.e.,

$$A = [a_{ij}]_{i,j\in\mathbb{N}} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots \\ a_{21} & a_{22} & a_{23} & \cdots \\ a_{31} & a_{32} & a_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

where each a_{ij} is a scalar. If we think of $x = (x_k)_{k\in\mathbb{N}}$ as being a "column vector," then we can *formally define* $y = Ax$ to be the product of the matrix A with the vector x by extending the familiar definition of a matrix-vector product from finite-dimensional linear algebra. That is, we let $y = Ax$ be the vector whose ith component is the *ith* row of A times the vector x:

$$y_i = (Ax)_i = \sum_{j=1}^{\infty} a_{ij}x_j. \tag{6.13}$$

However, this is only a "formal" definition because we do not know whether this equation even makes any sense—the series in equation (6.13) might not even converge! On the other hand, if we impose some appropriate restrictions on the infinite matrix A, then we may be able to show that the series in equation (6.13) does converge. If we can further show that for each $x \in \ell^p$ the vector $y = Ax = (y_i)_{i\in\mathbb{N}}$ belongs to ℓ^q, then we have a mapping that takes a vector $x \in \ell^p$ and outputs a vector $Ax \in \ell^q$. In this case we usually "identify" the matrix A with the mapping that sends x to Ax, and simply say that *A maps ℓ^p into ℓ^q*.

We will focus here on the case $q = p$. The following theorem gives a sufficient condition on an infinite matrix A that implies that A maps ℓ^p into ℓ^p for each $1 \le p \le \infty$ (in fact, this condition implies that $A\colon \ell^p \to \ell^p$ is *bounded* for each p). We call this result *Schur's Test* because it is the discrete version of a more general theorem known by that name. In the statement of Schur's Test we write $\|A\|_{\ell^p \to \ell^p}$ to emphasize that the operator norm of A depends on the value of p.

Theorem 6.7.1 (Schur's Test). *Let $A = [a_{ij}]_{i,j\in\mathbb{N}}$ be an infinite matrix such that*

$$C_1 \;=\; \sup_{i\in\mathbb{N}} \sum_{j=1}^{\infty} |a_{ij}| \;<\; \infty, \qquad C_2 \;=\; \sup_{j\in\mathbb{N}} \sum_{i=1}^{\infty} |a_{ij}| \;<\; \infty. \tag{6.14}$$

Then the following statements hold for each $1 \le p \le \infty$.

(a) *If $x \in \ell^p$, then the series in equation (6.13) converges for each $i \in \mathbb{N}$, and the vector $y = Ax = (y_i)_{i\in\mathbb{N}}$ belongs to ℓ^p.*

(b) *A is a bounded linear mapping of ℓ^p into ℓ^p.*

(c) *The operator norm of $A\colon \ell^p \to \ell^p$ satisfies*

$$\|A\|_{\ell^p \to \ell^p} \;\le\; C_1^{1/p'}\, C_2^{1/p},$$

where p' is the dual index *to p and we use the convention that $1/\infty = 0$.*

Proof. We will give the proof for $1 < p < \infty$, and assign the proof for the cases $p = 1$ and $p = \infty$ as Problem 6.7.5.

Note first that the entries of A are bounded since, for example,

$$|a_{ij}| \;\le\; \sum_{k=1}^{\infty} |a_{kj}| \;\le\; C_2, \qquad \text{all } i, j \in \mathbb{N}. \tag{6.15}$$

Fix $i \in \mathbb{N}$, and let $x = (x_k)_{k\in\mathbb{N}}$ be any vector in ℓ^p. Let $b_{ij} = |a_{ij}|^{1/p'}$ and $c_{ij} = |a_{ij}|^{1/p}$. Since $\frac{1}{p} + \frac{1}{p'} = 1$,

$$|a_{ij}| \;=\; |a_{ij}|^{1/p'}\, |a_{ij}|^{1/p} \;=\; b_{ij}\, c_{ij}.$$

Therefore, by applying Hölder's Inequality we see that

$$\begin{aligned}
\sum_{j=1}^{\infty} |a_{ij} x_j| \;&=\; \sum_{j=1}^{\infty} b_{ij}\,\bigl(c_{ij}\,|x_j|\bigr) \\
&\le\; \biggl(\sum_{j=1}^{\infty} b_{ij}^{p'}\biggr)^{1/p'} \biggl(\sum_{j=1}^{\infty} c_{ij}^{p}\,|x_j|^p\biggr)^{1/p} \qquad \text{(by Hölder)}
\end{aligned}$$

$$
\begin{aligned}
&= \left(\sum_{j=1}^{\infty} |a_{ij}|\right)^{1/p'} \left(\sum_{j=1}^{\infty} |a_{ij}|\,|x_j|^p\right)^{1/p} \\
&\le C_1^{1/p'} \left(\sum_{j=1}^{\infty} |a_{ij}|\,|x_j|^p\right)^{1/p} \qquad \text{(by equation (6.14))} \\
&\le C_1^{1/p'} \left(\sum_{j=1}^{\infty} C_2\,|x_j|^p\right)^{1/p} \qquad \text{(by equation (6.15))} \\
&= C_1^{1/p'} C_2^{1/p}\, \|x\|_p \;<\; \infty.
\end{aligned}
$$

This shows that the series of scalars $y_i = \sum_{j=1}^{\infty} a_{ij}x_j$ converges absolutely. Hence the vector $y = Ax = (y_i)_{i\in\mathbb{N}}$ is well-defined (i.e., it exists), although we do not yet know whether y belongs to ℓ^p.

Re-using some of the estimates made above we see that

$$
|y_i| \;=\; \left|\sum_{j=1}^{\infty} a_{ij}x_j\right| \;\le\; \sum_{j=1}^{\infty} |a_{ij}x_j| \;\le\; C_1^{1/p'} \left(\sum_{j=1}^{\infty} |a_{ij}|\,|x_j|^p\right)^{1/p}. \tag{6.16}
$$

Taking pth powers, we use this to estimate the ℓ^p-norm of $y = Ax$ as follows:

$$
\begin{aligned}
\|Ax\|_p^p \;=\; \|y\|_p^p \;=\; \sum_{i=1}^{\infty} |y_i|^p \;&\le\; \sum_{i=1}^{\infty} C_1^{p/p'} \sum_{j=1}^{\infty} |a_{ij}|\,|x_j|^p \\
&=\; C_1^{p/p'} \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} |a_{ij}|\,|x_j|^p \qquad (6.17) \\
&\le\; C_1^{p/p'} \sum_{j=1}^{\infty} C_2\,|x_j|^p \\
&=\; C_1^{p/p'} C_2\, \|x\|_p^p \;<\; \infty.
\end{aligned}
$$

The interchange in the order of summation that is made at the step corresponding to equation (6.17) is justified because all of the terms in the series are nonnegative (this is a discrete version of *Tonelli's Theorem*; such an interchange need not be valid if the series contains both positive and negative terms). We conclude that $Ax \in \ell^p$ whenever $x \in \ell^p$, i.e., A maps ℓ^p into ℓ^p. Further, by taking pth roots and then taking the supremum over all unit vectors in ℓ^p, we obtain a bound on the operator norm of A on ℓ^p:

$$
\|A\|_{\ell^p\to\ell^p} \;=\; \sup_{\|x\|_p=1} \|Ax\|_p \;\le\; \sup_{\|x\|_p=1} C_1^{1/p'} C_2^{1/p}\, \|x\|_p \;=\; C_1^{1/p'} C_2^{1/p}. \qquad \diamondsuit
$$

An entirely analogous result holds for bi-infinite matrices $A = [a_{ij}]_{i,j\in\mathbb{Z}}$ and bi-infinite vectors $x = (x_k)_{k\in\mathbb{Z}}$; we just need to replace the index set $\mathbb{N}$ with $\mathbb{Z}$ in the results above. We will present an important corollary of

Theorem 6.7.1 that holds in this bi-infinite setting. First, however, we make another formal definition.

Definition 6.7.2 (Convolution of Sequences). Let $x = (x_k)_{k\in\mathbb{Z}}$ and $y = (y_k)_{k\in\mathbb{Z}}$ be bi-infinite sequences of scalars. The *convolution* of x and y is formally defined to be the sequence $x * y$ whose kth component is

$$(x * y)_k \;=\; \sum_{j=-\infty}^{\infty} x_j\, y_{k-j}. \qquad \diamondsuit \tag{6.18}$$

As before, the word *formal* means that this definition need not always make sense. In particular, if the series given in equation (6.18) does not converge for some $k \in \mathbb{Z}$, then the convolution of x with y does not exist (give an example of x and y for which that happens). If the series does exist for every k, then $x * y$ exists and is the sequence whose kth component is defined by equation (6.18).

By rephrasing convolution in terms of an infinite matrix and applying Schur's Test, we obtain the following result. This is the discrete version of a more general theorem known as *Young's Inequality.* In this theorem, $\ell^p(\mathbb{Z})$ is the space of all bi-infinite sequences that are p-summable (if p is finite) or bounded (if $p = \infty$).

Theorem 6.7.3 (Young's Inequality). *Let $1 \le p \le \infty$ be given. If $x \in \ell^p(\mathbb{Z})$ and $y \in \ell^1(\mathbb{Z})$, then the convolution $x * y$ exists and belongs to $\ell^p(\mathbb{Z})$, and*

$$\|x * y\|_p \;\le\; \|x\|_p\, \|y\|_1. \tag{6.19}$$

Proof. Let A be the bi-infinite matrix

$$A \;=\; [y_{i-j}]_{i,j\in\mathbb{Z}}. \tag{6.20}$$

Note that the entries of A are constant along each diagonal. A matrix that has constant entries along diagonals is called a *Toeplitz matrix.*

For this matrix A, the bi-infinite analogues of the numbers C_1 and C_2 defined in equation (6.14) are

$$C_1 \;=\; \sup_{i\in\mathbb{Z}} \sum_{j=-\infty}^{\infty} |y_{i-j}| \quad \text{and} \quad C_2 \;=\; \sup_{j\in\mathbb{Z}} \sum_{i=-\infty}^{\infty} |y_{i-j}|.$$

By making a change of variables, we see that

$$\sum_{j=-\infty}^{\infty} |y_{i-j}| \;=\; \sum_{j=-\infty}^{\infty} |y_j| \;=\; \|y\|_1,$$

which is a finite constant independent of i. Therefore $C_1 = \|y\|_1$, and likewise $C_2 = \|y\|_1$. As C_1 and C_2 are both finite, Schur's Test therefore implies that Ax exists for each $x \in \ell^p$. But

$$(Ax)_i \;=\; \sum_{j=-\infty} y_{i-j}\, x_j \;=\; (x * y)_i,$$

so we conclude that the convolution

$$x * y \;=\; \big((x * y)_i\big)_{i\in\mathbb{Z}}$$

exists. Further, Schur's Test tells us that A maps ℓ^p boundedly into itself, and

$$\begin{aligned}
\|x * y\|_p \;=\; \|Ax\|_p \;&\le\; \|A\|_{\ell^p\to\ell^p}\, \|x\|_p \\
&\le\; C_1^{1/p'} C_2^{1/p}\, \|x\|_p \\
&=\; \|y\|_1^{1/p'}\, \|y\|_1^{1/p}\, \|x\|_p \\
&=\; \|y\|_1\, \|x\|_p.
\end{aligned}$$

This is true for every $x \in \ell^p(\mathbb{Z})$, so equation (6.19) holds. □

In particular, by taking $p = 1$ we see that $\ell^1(\mathbb{Z})$ *is closed under convolution.* That is, if x and y are any vectors in $\ell^1(\mathbb{Z})$, then their convolution $x*y$ exists and belongs to $\ell^1(\mathbb{Z})$ as well. Further, we have the following *submultiplicative* property of convolution:

$$\|x * y\|_1 \;\le\; \|x\|_1\, \|y\|_1, \qquad x, y \in \ell^1(\mathbb{Z}). \tag{6.21}$$

Remark 6.7.4. The entries of the matrix $A = [y_{i-j}]_{i,j\in\mathbb{Z}}$ defined in equation (6.20) depend only on the *difference* $i-j$ rather than on i and j individually. Such a matrix is said to be *Toeplitz*. What does a Toeplitz matrix look like? In particular, what does it look like if y belongs to c_{00}, i.e., if only finite many entries of y are nonzero? ♢

We have only discussed a discrete version of convolution. Although it may not be apparent at first glance, convolution is an important operation that plays a central role in many areas, including harmonic analysis in mathematics and signal processing in engineering. Some references that include more details on convolution include [DM72], [Ben97], [Kat04], [Heil11].

Problems

6.7.5. Complete the proof of Theorem 6.7.1 for the cases $p = 1$ and $p = \infty$.

6.7.6. Prove equation (6.21) directly, i.e., without appealing to Theorems 6.7.1 or 6.7.3. Show further that if every component of x and y is nonnegative, then $\|x * y\|_1 = \|x\|_1\, \|y\|_1$.

6.7.7. Prove the following properties of convolution of sequences in $\ell^1(\mathbb{Z})$.

(a) Commutativity: $x * y = y * x$ for all $x,\, y \in \ell^1(\mathbb{Z})$.

(b) Associativity: $(x * y) * z = x * (y * z)$ for all $x,\, y,\, z \in \ell^1(\mathbb{Z})$.

(c) Distributive Law: $x * (y + z) = x * y + x * z$ for all $x,\, y,\, z \in \ell^1(\mathbb{Z})$.

(d) Interaction with scalar multiplication: $c(x * y) = (cx) * y = x * (cy)$ for all $x,\, y \in \ell^1(\mathbb{Z})$ and scalars $c \in \mathbb{F}$.

(e) Existence of an Identity: There is a sequence $\delta \in \ell^1(\mathbb{Z})$ such that $x * \delta = x$ for every sequence $x \in \ell^1(\mathbb{Z})$.

6.7.8. Fix $1 \le p \le \infty$. Prove directly that if $x \in \ell^p(\mathbb{Z})$, and $y \in \ell^{p'}(\mathbb{Z})$, then $x * y$ exists and belongs to $\ell^\infty(\mathbb{Z})$, and $\|x * y\|_\infty \le \|x\|_p \, \|y\|_{p'}$.

6.8 Integral Operators

In this section we will present several more examples of operators. The operators that we will define here, which are called *integral operators*, act on measurable functions $f \colon E \to \mathbb{C}$. For simplicity of presentation, we will mostly focus on functions whose domain is the real line $\mathbb{R}$, although most of the results do carry over to other domains (see Section 6.8.4).

6.8.1 The Definition of an Integral Operator

Not surprisingly, the definition of an integral operator involves an integral. This is a Lebesgue integral, which we introduced and studied in detail in Volume 1. The integral in question will not always exist, so we make a "formal" definition of the integral operator. Following mathematical tradition, a *formal statement* often means an entirely *informal* statement. In particular, here it means that integral need not make sense for some f, and so for those f the value of $L_k f$ is simply undefined.

Definition 6.8.1 (Integral Operator). Let k be a fixed measurable function on $\mathbb{R}^2$. Then the *integral operator* L_k *with kernel* k is formally defined by

$$L_k f(x) \;=\; \int_{-\infty}^{\infty} k(x,y)\, f(y)\, dy. \tag{6.22}$$

That is, if f is a measurable function on $\mathbb{R}$, then $L_k f$ is the function defined by equation (6.22), as long as this integral is well-defined for a.e. $x \in \mathbb{R}$. Otherwise $L_k f$ is not defined. $\diamondsuit$

The use of the word *kernel* in this definition should not be confused with its other meaning as the nullspace of an operator. It should always be clear from context which meaning of "kernel" is intended.

Remark 6.8.2. Integral operators are natural generalizations of the ordinary matrix-vector product. To see this, let A be an $m \times n$ matrix with complex entries a_{ij} and let u be a vector in $\mathbb{C}^n$. Then $Au \in \mathbb{C}^m$, and its components are

$$(Au)_i \;=\; \sum_{j=1}^{n} a_{ij}\, u_j, \qquad i = 1, \dots, m.$$

Thus, the function values $k(x,y)$ of the kernel of an integral operator are entirely analogous to the entries a_{ij} of the matrix A, and the values $L_k f(x)$ of the function $L_k f$ are analogous to the entries $(Au)_i$ of the vector Au. However, while Au is defined for every vector u, it may be the case that $L_k f$ is only well-defined for certain functions f. $\diamondsuit$

As an illustration, we will consider integral operators whose kernels take the following especially simple form.

Definition 6.8.3. Given two functions g, h whose domain is the real line $\mathbb{R}$, we define $g \otimes h$ to be the function on $\mathbb{R}^2$ whose rule is

$$(g \otimes h)(x,y) \;=\; g(x)\, \overline{h(y)}, \qquad x, y \in \mathbb{R}. \tag{6.23}$$

We call $g \otimes h$ the *tensor product* of g and h. $\diamondsuit$

Sometimes the complex conjugate is omitted in the definition of a tensor product, or the complex conjugate is placed on g instead of h. It will be most convenient for our purposes to place the complex conjugate on the function h.

Now we look at an integral operator whose kernel k is a tensor product of two functions from $L^2(\mathbb{R})$. We will see that, in this case, $L_k f$ is well defined for every function $f \in L^2(\mathbb{R})$, and L_k defines a bounded linear operator that maps $L^2(\mathbb{R})$ into itself.

Example 6.8.4 (Tensor Product Kernels). Fix $g, h \in L^2(\mathbb{R})$ and set $k = g \otimes h$. Given $f \in L^2(\mathbb{R})$, the inner product $\langle f, h\rangle$ is well-defined. Further, $L_k f$ has the form

$$L_k f(x) \;=\; \int_{-\infty}^{\infty} g(x)\, \overline{h(y)}\, f(y)\, dy \;=\; g(x) \int_{-\infty}^{\infty} f(y)\, \overline{h(y)}\, dy \;=\; \langle f, h\rangle\, g(x).$$

That is,

$$L_k f \;=\; \langle f, h\rangle\, g, \qquad f \in L^2(\mathbb{R}).$$

Thus, for this choice of kernel, the integral operator L_k maps $L^2(\mathbb{R})$ into itself, and in fact L_k maps every vector in $L^2(\mathbb{R})$ to a scalar multiple of g. If either $g = 0$ or $h = 0$ then L_k is the zero operator; otherwise the range of L_k is the one-dimensional subspace spanned by g. The operator L_k is a bounded mapping on $L^2(\mathbb{R})$, because it follows from the Cauchy–Bunyakowski–Schwarz Inequality that

$$\|L_k f\|_2 \;=\; |\langle f, h\rangle|\, \|g\|_2 \;\le\; \|f\|_2\, \|h\|_2\, \|g\|_2 \;=\; C\, \|f\|_2,$$

where $C = \|h\|_2 \, \|g\|_2 < \infty$. $\diamondsuit$

In summary, if $k = g \otimes h$ where g, $h \in L^2(\mathbb{R})$, then L_k is a bounded linear operator on $L^2(\mathbb{R})$. If L_k is not the zero operator then its range is one-dimensional, so $\operatorname{rank}(L_k) = 1$ (even so, L_k is not a *functional* because its range is not $\mathbb{C}$). An operator whose range is one-dimensional is sometimes called a *rank-one operator*.

We often identify the tensor product function $k = g \otimes h$ with the integral operator $L_k = L_{g \otimes h}$ whose kernel is $k = g \otimes h$ in the following way.

Notation 6.8.5. Given g and h in $L^2(\mathbb{R})$, we let the symbols $g \otimes h$ denote either the tensor product function given in equation (6.23), or the operator whose rule is

$$(g \otimes h)(f) \;=\; \langle f, h \rangle \, g, \qquad f \in L^2(\mathbb{R}). \tag{6.24}$$

It is usually clear from context whether $g \otimes h$ is meant to denote a function or an operator. $\diamondsuit$

We can extend the operator notation $g \otimes h$ to an arbitrary Hilbert space H by simply replacing $L^2(\mathbb{R})$ with H in equation (6.24). That is, given any two vectors g, h in a Hilbert space H, we declare that $g \otimes h$ is the *operator* whose rule is

$$(g \otimes h)(f) \;=\; \langle f, h \rangle \, g, \qquad f \in H.$$

We call $g \otimes h$ the *tensor product* of g and h.

Example 6.8.6. (a) If g is a unit vector (i.e., $\|g\| = 1$) and we take $h = g$, then the tensor product operator $g \otimes g$ is the orthogonal projection of H onto the one-dimensional subspace spanned by g.

(b) The *outer product* of two column vectors u, $v \in \mathbb{C}^d$ is the matrix uv^{H}, where v^{H} is the *Hermitian*, or conjugate transpose, of v. The transformation induced by the matrix uv^{H} is

$$(uv^{\mathrm{H}})x \;=\; u\,(v^{\mathrm{H}}x) \;=\; (x \cdot v)\, u \;=\; (u \otimes v)(x), \qquad x \in \mathbb{C}^d.$$

Hence the tensor product $u \otimes v$ of u, $v \in \mathbb{C}^d$ is precisely the outer product of u and v. $\diamondsuit$

For a general kernel k, it can be much more difficult to determine classes of functions f for which $L_k f$ is well-defined. In the coming pages we will present several conditions on k that imply that L_k is well-defined and bounded on the Lebesgue space $L^p(\mathbb{R})$.

6.8.2 Hilbert–Schmidt Integral Operators

It is not always obvious how properties of a kernel k relate to properties of the integral operator L_k determined by k. Our first theorem in this direction will show that if the kernel k is a square-integrable function on $\mathbb{R}^2$, then L_k is a bounded mapping on $L^2(\mathbb{R})$. In the proof of this theorem, note that f belongs to $L^2(\mathbb{R})$ while $k \in L^2(\mathbb{R}^2)$. We use $\|f\|_2$ and $\|k\|_2$ to denote the L^2-norms of these functions, the domains $\mathbb{R}$ or $\mathbb{R}^2$ being clear from context.

Theorem 6.8.7 (Hilbert–Schmidt Integral Operators). *If $k \in L^2(\mathbb{R}^2)$, then the integral operator L_k defined by equation* (6.22) *is a bounded mapping of $L^2(\mathbb{R})$ into itself, and its operator norm satisfies $\|L_k\| \le \|k\|_2$.*

Proof. Suppose that $k \in L^2(\mathbb{R}^2)$, and fix $f \in L^2(\mathbb{R})$. Fubini's Theorem tells us that the function $k_x(y) = k(x,y)$ belongs to $L^2(\mathbb{R})$ for a.e. x. For those x's we have $k_x,\ \bar{f} \in L^2(\mathbb{R})$, and their inner product is precisely $L_kf(x)$ because

$$L_kf(x) \;=\; \int_{-\infty}^{\infty} k_x(y)\, f(y)\, dy \;=\; \langle k_x, \bar{f}\,\rangle.$$

Thus $L_kf(x)$ is defined for almost every x. We still must show that $L_kf \in L^2(\mathbb{R})$ and $\|L_kf\|_2 \le \|k\|_2\, \|f\|_2$.

Step 1. Suppose that f and k are both nonnegative a.e. Then $k(x,y)\, f(y)$ is a nonnegative measurable function on the domain $\mathbb{R}^2$, so it follows from Tonelli's Theorem that $L_kf(x) = \int k(x,y)\, f(y)\, dy$ is a measurable function of x. We estimate its L^2-norm by applying the Cauchy–Bunyakovski–Schwarz Inequality:

$$\begin{aligned}
\|L_kf\|_2^2 \;&=\; \int_{-\infty}^{\infty} |L_kf(x)|^2\, dx \\
&=\; \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} k(x,y)\, f(y)\, dy \right|^2 dx \\
&\le\; \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |k(x,y)|^2\, dy \right) \left(\int_{-\infty}^{\infty} |f(y)|^2\, dy \right) dx \\
&=\; \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |k(x,y)|^2\, dy \;\; \|f\|_2^2\, dx \\
&=\; \|f\|_2^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |k(x,y)|^2\, dy\, dx \\
&=\; \|k\|_2^2\, \|f\|_2^2 \;<\; \infty.
\end{aligned} \tag{6.25}$$

Therefore $L_kf \in L^2(\mathbb{R})$, and $\|L_kf\|_2 \le \|k\|_2\, \|f\|_2$.

Step 2. Now let $f \in L^2(\mathbb{R})$ and $k \in L^2(\mathbb{R}^2)$ be arbitrary functions. Write

$$f = (f_1^+ - f_1^-) + i(f_2^+ - f_2^-) \qquad \text{and} \qquad k = (k_1^+ - k_1^-) + i(k_2^+ - k_2^-),$$

where each function $f_\ell^\pm$ and $k_j^\pm$ is nonnegative a.e. By Step 1, each of the functions $L_{k_j^\pm}(f_\ell^\pm)$ is measurable and belongs to $L^2(\mathbb{R})$. Since $L_k f$ is a finite linear combination of the sixteen functions $L_{k_j^\pm}(f_\ell^\pm)$, we conclude that $L_k f$ is measurable and belongs to $L^2(\mathbb{R})$. Further, now that we know that $L_k f$ is measurable, we can repeat the same set of calculations that appear in equation (6.25), and conclude that $\|L_k f\|_2 \le \|k\|_2 \, \|f\|_2$. Since this inequality holds for all $f \in L^2(\mathbb{R})$, it follows that L_k maps $L^2(\mathbb{R})$ boundedly into itself, and its operator norm satisfies the estimate $\|L_k\| \le \|k\|_2$. □

6.8.3 Schur's Test

Our next theorem will give a different sufficient condition on the kernel k that ensures that L_k is a bounded operator on $L^2(\mathbb{R})$. Originally formulated in [Sch11], this result is known as *Schur's Test* (not to be confused with *Schur's Lemma*). It is complementary to Theorem 6.8.7, and which theorem to use depends on the situation at hand. For example, the kernel that appears in Exercise 6.8.11 below is not square-integrable, but it does satisfy the hypotheses of Schur's Test.

Theorem 6.8.8 (Schur's Test). *Let k be a measurable function on $\mathbb{R}^2$ that satisfies the mixed-norm conditions*

$$\begin{aligned} C_1 &= \operatorname*{ess\,sup}_{x\in\mathbb{R}} \int_{-\infty}^{\infty} |k(x,y)|\, dy \;<\; \infty, \\ C_2 &= \operatorname*{ess\,sup}_{y\in\mathbb{R}} \int_{-\infty}^{\infty} |k(x,y)|\, dx \;<\; \infty. \end{aligned} \tag{6.26}$$

Then the integral operator L_k defined by equation (6.22) *is a bounded mapping of $L^2(\mathbb{R})$ into itself, and its operator norm satisfies $\|L_k\| \le (C_1C_2)^{1/2}$.*

Proof. As in the proof of Theorem 6.8.7, measurability of $L_k f$ is most easily shown by first considering nonnegative f, k, and then extending to the general case. We omit the details and assume that $L_k f$ is measurable for all $f \in L^2(\mathbb{R})$. Then by applying the Cauchy–Bunyakovski–Schwarz Inequality we see that

$$\begin{aligned} \|L_k f\|_2^2 &= \int_{-\infty}^{\infty} |L_k f(x)|^2\, dx \\ &= \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} k(x,y)\, f(y)\, dy \right|^2 dx \\ &\le \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |k(x,y)|^{1/2} \cdot |k(x,y)|^{1/2}\, |f(y)|\, dy \right)^2 dx \end{aligned}$$

$$
\begin{aligned}
&\le \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |k(x,y)|\,dy \right) \left(\int_{-\infty}^{\infty} |k(x,y)|\,|f(y)|^2\,dy \right) dx \\
&\le \int_{-\infty}^{\infty} C_1 \int_{-\infty}^{\infty} |k(x,y)|\,|f(y)|^2\,dy\,dx \\
&= C_1 \int_{-\infty}^{\infty} |f(y)|^2 \int_{-\infty}^{\infty} |k(x,y)|\,dx\,dy \\
&\le C_1 \int_{-\infty}^{\infty} |f(y)|^2\, C_2\,dy \\
&= C_1 C_2\,\|f\|_2^2.
\end{aligned}
$$

We were allowed to interchange the order of integration in the preceding calculation because the integrand $|f(y)|^2\,|k(x,y)|$ is nonnegative, and therefore Tonelli's Theorem is applicable. It follows that L_k is bounded, and its operator norm satisfies $\|L_k\| \le (C_1C_2)^{1/2}$. $\square$

Schur's Test actually allows us to draw a stronger conclusion than what is given in Theorem 6.8.8. By using Hölder's Inequality instead of Cauchy–Bunyakovski–Schwarz, the next exercise shows that the integral operator L_k is defined and bounded on $L^p(\mathbb{R})$ for each index $1 \le p \le \infty$. Recall that p' denotes the dual index to p, i.e., p' is the extended real number that satisfies $\frac{1}{p} + \frac{1}{p'} = 1$.

Exercise 6.8.9. Show that if k satisfies the conditions given in equation (6.26), then L_k is a bounded mapping of $L^p(\mathbb{R})$ into itself for each index $1 \le p \le \infty$, and its operator norm satisfies $\|L_k\| \le C_1^{1/p'} C_2^{1/p}$. $\diamondsuit$

6.8.4 Integral Operators on Other Domains

So far in this section we have focused on functions whose domain is the real line, kernels that are defined on $\mathbb{R}^2$, and the corresponding integral operators. However, all of the results that we have derived so far can be extended to functions that are defined measurable subsets of $\mathbb{R}^d$. Specifically, let $E \subseteq \mathbb{R}^m$ and $F \subseteq \mathbb{R}^n$ be measurable sets, and let k be any measurable function that is defined on $E \times F$. Given a function f defined on F, we formally define $L_k f$ on the domain E by the rule

$$
L_k f(x) \;=\; \int_F k(x,y)\,f(y)\,dy, \qquad x \in E.
$$

The analogues of Theorem 6.8.7 and 6.8.8 for this setting are stated in the following exercise.

Exercise 6.8.10. Let $E \subseteq \mathbb{R}^m$ and $F \subseteq \mathbb{R}^n$ be measurable sets, and let k be a measurable function on $E \times F$. Prove the following statements.

(a) If $k \in L^2(E \times F)$, then L_k maps $L^2(F)$ boundedly into $L^2(E)$.

(b) If

$$\operatorname*{ess\,sup}_{x \in E} \int_F |k(x,y)|\, dy \;<\; \infty, \qquad \operatorname*{ess\,sup}_{y \in F} \int_E |k(x,y)|\, dx \;<\; \infty,$$

then L_k maps $L^p(F)$ boundedly into $L^p(E)$ for $1 \le p \le \infty$. $\diamondsuit$

6.8.5 Convolution as an Integral Operator

We introduced the convolution of integrable functions in Volume 1. In particular, $L^1(\mathbb{R})$ is closed under convolution. We will give a different derivation of some of the results of that exercise by viewing convolution as a special type of an integral operator.

Formally, the convolution of two functions f and g is the function $f * g$ defined by

$$(f * g)(x) \;=\; \int_{-\infty}^{\infty} f(y)\, g(x-y)\, dy,$$

as long as this integral makes sense. Comparing this to equation (6.22), we see that if we set $k(x,y) = g(x-y)$, then the integral operator whose kernel is k has the rule $L_k f \;=\; f * g$. However, even if g belongs to $L^2(\mathbb{R})$, the kernel $k(x,y) \;=\; g(x-y)$ will not belong to $L^2(\mathbb{R}^2)$, so Theorem 6.8.7 is not applicable to this situation. On the other hand, if we assume that g is *integrable* then Schur's Test can be applied, and it yields the results stated in the next exercise.

Exercise 6.8.11. Prove *Young's Inequality*: If $1 \le p \le \infty$, $f \in L^p(\mathbb{R})$, and $g \in L^1(\mathbb{R})$, then $f * g \in L^p(\mathbb{R})$ and

$$\|f * g\|_p \le \|f\|_p \, \|g\|_1. \qquad \diamondsuit$$

Taking $p = 1$, we see that if f and g both belong to $L^1(\mathbb{R})$, then so does $f * g$. Hence $L^1(\mathbb{R})$ is closed under convolution. Moreover, the convolution of integrable functions is commutative and associative and is *submultiplicative* in the sense that

$$\|f * g\|_1 \;\le\; \|f\|_1 \, \|g\|_1.$$

Using terminology that we will introduce later, this says that $L^1(\mathbb{R})$ is a commutative *Banach algebra* with respect to convolution.

Problems

6.8.12. Consider the kernel $k(x,y) = e^{2\pi i(x-y)}$ on the domain $[0,1]^2$. Prove that the corresponding integral operator $L_k \colon L^2[0,1] \to L^2[0,1]$ is an orthogonal projection.

6.8.13. The *Volterra operator* V is the integral operator

$$Vf(x) = \int_0^x f(y)\,dy.$$

Prove that V maps $L^p[0,1]$ continuously into itself for each $1 \le p \le \infty$, and likewise $V \colon C[0,1] \to C[0,1]$ is continuous.

6.8.14. Prove the following weighted version of Schur's Test. Assume that k is a measurable function on $\mathbb{R}^2$ and there are measurable functions u, $v \ge 0$ on $\mathbb{R}$ such that

$$\int_{-\infty}^{\infty} |k(x,y)|\,v(y)\,dy \le C_1\,u(x), \qquad \text{a.e. } x,$$

$$\int_{-\infty}^{\infty} |k(x,y)|\,u(x)\,dx \le C_2\,v(y), \qquad \text{a.e. } y.$$

Prove that the integral operator L_k whose kernel is k defines a bounded mapping of $L^2(\mathbb{R})$ into itself.

6.8.15. Given $g \in L^1(\mathbb{R})$, define $A_g \colon L^1(\mathbb{R}) \to L^1(\mathbb{R})$ by $A_g(f) = f * g$. Show that the operator norm of A_g is $\|A_g\| = \|g\|_1$.

6.9 The Dual of a Hilbert Space

An operator that maps a normed space X into the field of scalars $\mathbb{F}$ is called a *functional* (see Definition 6.1.1). Typically, we use lowercase Greek letters such as λ, μ, or ν to denote functionals. The set of all bounded linear functionals on X is called the dual space of X and is denoted by X^*.

Definition 6.9.1 (Dual Space). The set of all bounded linear functionals on a normed space X is called the *dual space* of X. We denote the dual space by

$$X^* = \mathcal{B}(X,\mathbb{F}) = \bigl\{A \colon X \to \mathbb{F} \;:\; A \text{ is bounded and linear}\bigr\}. \qquad \diamondsuit$$

Because the scalar field $\mathbb{F}$ is complete, Theorem 6.5.3 implies that X^* is complete (even if X is not complete).

Since a linear functional maps every vector in X to a *scalar*, it is in some sense quite "simple." We can often obtain useful information about X that may not be immediately apparent from direct examination of X by considering its dual space X^*. As the norm on $\mathbb{F}$ is absolute value, the operator norm of a functional $\mu \in X^*$ is

$$\|\mu\| = \sup_{\|x\|=1} |\mu(x)|.$$

We will prove that if H is a Hilbert space then H^* is isometrically isomorphic to H. Specifically, we will prove that each vector in H determines a functional in X^*, and conversely each functional in X^* is determined by a unique vector in H, and furthermore the relation between these two is isometric. To see how this correspondence works, fix a vector $y \in X$, and define a functional $\mu_y \colon H \to \mathbb{F}$ by

$$\mu_y(x) \;=\; \langle x, y\rangle, \qquad x \in H.$$

Since the inner product is linear in the first variable, μ_y is linear:

$$\mu_y(ax+bz) \;=\; \langle ax+bz, y\rangle \;=\; a\,\langle x,y\rangle + b\,\langle z,y\rangle \;=\; a\,\mu_y(x) + b\,\mu_y(z).$$

The Cauchy–Bunyakovski–Schwarz Inequality implies that

$$|\mu_y(x)| \;=\; |\langle x,y\rangle| \;\le\; \|x\|\,\|y\|, \qquad x \in H,$$

so by taking the supremum over the unit vectors in H we see that

$$\|\mu_y\| \;=\; \sup_{\|x\|=1} |\mu_y(x)| \;\le\; \sup_{\|x\|=1} \|x\|\,\|y\| \;=\; \|y\|.$$

Therefore μ_y is bounded, and hence is an element of H^*. Therefore we can define a mapping $T \colon H \to H^*$ by setting

$$T(y) \;=\; \mu_y, \qquad y \in H.$$

We will prove in the next theorem that T is an isometric isomorphism. There is one twist, however, as T is not a *linear* map but rather is *antilinear*. This is actually a natural consequence of the fact that the inner product is antilinear in the second variable. While we have only considered linear maps up to now, antilinear maps are very similar, and we use similar terminology for antilinear maps as for linear maps. In particular, an *antilinear isometric isomorphism* is a *antilinear* mapping that is both isometric and surjective. To emphasize, while μ_y is a *linear* function that sends x to $\mu_y(x)$, the mapping T is an *antilinear* function that sends y to μ_y.

Theorem 6.9.2 (Riesz Representation Theorem). *Let H be a Hilbert space. Using the notations given above, the following statements hold.*

(a) *If $y \in X$ then $\mu_y \in H^*$, and the operator norm of μ_y equals the norm of y:*

$$\|\mu_y\| \;=\; \sup_{\|x\|=1} |\langle x, y\rangle| \;=\; \|y\|.$$

(b) *If $\mu \in H^*$ then there exists a unique vector $y \in H$ such that $\mu = \mu_y$. That is, if $\mu\colon H \to \mathbb{F}$ is a bounded linear operator, then there exists a unique $y \in H$ such that*

$$\mu(y) \;=\; \langle x, y\rangle, \qquad x \in H.$$

(c) *The mapping $T\colon H \to H^*$ given by $T(y) = \mu_y$ is an antilinear isometric isomorphism of H onto H^*. In particular, if y, $z \in H$ and a, $b \in \mathbb{F}$, then*

$$\mu_{ay+bz} \;=\; T(ay+bz) \;=\; \bar{a}T(y) + \bar{b}T(z) \;=\; \bar{a}\mu_y + \bar{b}\mu_z.$$

Proof. (a) We have already seen that μ_y is linear and satisfies $\|\mu_y\| \le \|y\|$. If $y = 0$, then $\mu_y = 0$ and we are done. On the other hand, if $y \neq 0$ then $z = y/\|y\|$ is a unit vector, and

$$|\mu_y(z)| \;=\; |\langle z, y\rangle| \;=\; \frac{\langle y, y\rangle}{\|y\|} \;=\; \frac{\|y\|^2}{\|y\|} \;=\; \|y\|.$$

Since $\|z\| = 1$, this implies that

$$\|\mu_y\| \;=\; \sup_{\|x\|=1} |\mu_y(x)| \;\ge\; |\mu_y(z)| \;=\; \|y\|.$$

Combining this with the inequality $\|\mu_y\| \le \|y\|$ proved before, we see that $\|\mu_y\| = \|y\|$.

(b) Let μ be any functional in H^*. If $\mu = 0$ then $y = 0$ is the required vector, so we may assume that μ is not the zero operator. In this case μ does not map every vector to zero, so the kernel of μ is a subspace of H that is not all of H. Hence its orthogonal complement $\ker(\mu)^\perp$ contains more than just the zero vector. Let z be any nonzero vector in $\ker(\mu)^\perp$. Then $z \notin \ker(\mu)$, so $\mu(z) \neq 0$. Let

$$w \;=\; \frac{1}{\mu(z)}\, z \qquad \text{and} \qquad y \;=\; \frac{\mu(z)}{\|z\|^2}\, z.$$

Since μ is linear,

$$\mu(w) \;=\; \frac{1}{\mu(z)}\, \mu(z) \;=\; 1,$$

and we also have

$$\langle w, y\rangle \;=\; \frac{\mu(z)}{\mu(z)\, \|z\|^2}\, \langle z, z\rangle \;=\; 1.$$

Now let x be any vector in H. Then, since μ is linear,

$$\mu\big(x - \mu(x)w\big) \;=\; \mu(x) - \mu(x)\, \mu(w) \;=\; 0.$$

Therefore $x - \mu(x)w \in \ker(\mu)$. But $y \in \ker(\mu)^\perp$ (because y is a scalar multiple of z), so $x - \mu(x)w$ and y are orthogonal. Hence

$$0 \;=\; \bigl\langle x - \mu(x)w,\, y\bigr\rangle \;=\; \langle x, y\rangle - \mu(x)\,\langle w, y\rangle \;=\; \langle x, y\rangle - \mu(x).$$

Thus $\mu(x) = \langle x, y\rangle = \mu_y(x)$. As is true for every x, we have shown that $\mu = \mu_y$.

It only remains to show that y is unique. Suppose that we also had $\mu = \mu_z$ for some $z \in H$. Then for every vector $x \in H$,

$$\langle x, y - z\rangle \;=\; \langle x, y\rangle - \langle x, z\rangle \;=\; \mu_y(x) - \mu_z(x) \;=\; \mu(x) - \mu(x) \;=\; 0.$$

Since this is true for every x, including $x = y - z$ in particular, it follows that $y - z = 0$. Therefore y is unique.

(c) Parts (a) and (b) show that T is both surjective and norm-preserving. Therefore, we just have to show that T is antilinear. To do this, fix any $y \in H$ and $c \in \mathbb{F}$. Then, since the inner product is antilinear in the second variable, for each $x \in H$ we have

$$\mu_{cy}(x) \;=\; \langle x, cy\rangle \;=\; \bar{c}\,\langle x, y\rangle \;=\; \bar{c}\,\mu_y(x).$$

Therefore $T(cy) = \mu_{cy} = \bar{c}\,\mu_y = \bar{c}\,T_y$. A similar argument shows that

$$T(y + z) \;=\; \mu_{y+z} \;=\; \mu_y + \mu_z \;=\; T(y) + T(z),$$

so T is antilinear. $\square$

Problems

6.9.3. We say that a sequence $\{y_n\}_{n\in\mathbb{N}}$ of vectors in a Hilbert space H is a *Bessel sequence* if there exists a constant $B < \infty$ such that

$$\sum_{n=1}^{\infty} |\langle x, y_n\rangle|^2 \;\le\; B\,\|x\|^2, \qquad x \in H.$$

Prove the following statements.

(a) If $\{y_n\}_{n\in\mathbb{N}}$ is a Bessel sequence, then

$$Ax \;=\; \bigl\{\langle x, y_n\rangle\bigr\}_{n\in\mathbb{N}}, \qquad x \in H, \tag{6.27}$$

defines a bounded linear operator from A to ℓ^2.

(b) If $A\colon H \to \ell^2$ is a bounded linear operator, then there exists a Bessel sequence $\{y_n\}_{n\in\mathbb{N}}$ such that equation (6.27) holds.

6.10 The Dual of ℓ^p

Since ℓ^2 is Hilbert space, the Riesz Representation Theorem applies to ℓ^2, and it tells us that the dual space $(\ell^2)^*$ is (antilinearly) isometrically isomorphic to ℓ^2. Now we will try to characterize the dual of ℓ^p for p other than 2.

First we will show that every vector $y \in \ell^{p'}$ determines a bounded linear functional μ_y on ℓ^p. In contrast to the case $p = 2$, we will omit the complex conjugate in the definition of μ_y. This will give us the minor convenience that the isomorphism T that we create later will be linear instead of antilinear.

Theorem 6.10.1. *Fix $1 \le p \le \infty$. For each $y \in \ell^{p'}$, define $\mu_y \colon \ell^p \to \mathbb{F}$ by*

$$\mu_y(x) \;=\; \sum_{k=1}^{\infty} x_k\, y_k, \qquad x = (x_k)_{k\in\mathbb{N}} \in \ell^p. \tag{6.28}$$

Then $\mu_y \in (\ell^p)^*$, and the operator norm of μ_y equals the norm of y, i.e.,

$$\|\mu_y\| \;=\; \sup_{\|x\|=1} |\mu_y(x)| \;=\; \|y\|.$$

Proof. We will consider $1 < p < \infty$, and assign the proofs for $p = 1$ and $p = \infty$ as Problem 6.10.6.

If $1 < p < \infty$ then we also have $1 < p' < \infty$. If $x \in \ell^p$, then, since $y \in \ell^{p'}$, we can apply Hölder's Inequality to obtain the estimate

$$|\mu_y(x)| \;=\; \left|\sum_{k=1}^{\infty} x_k y_k\right| \;\le\; \|x\|_p \, \|y\|_{p'}.$$

Taking the supremum over all unit vectors x in ℓ^p, it follows that μ_y is bounded, and

$$\|\mu_y\| \;=\; \sup_{\|x\|=1} |\mu_y(x)| \;\le\; \sup_{\|x\|=1} \|x\|_p \, \|y\|_{p'} \;=\; \|y\|_{p'}.$$

We must prove that we also have $\|\mu_y\| \ge \|y\|_{p'}$.

If y is the zero sequence then we are done, so assume that not every component of y is zero. For each $k \in \mathbb{N}$, let α_k be a scalar such that $|\alpha_k| = 1$ and $\alpha_k\, y_k = |y_k|$. Set

$$z_k \;=\; \frac{\alpha_k\, |y_k|^{p'-1}}{\|y\|_{p'}^{p'-1}}, \qquad k \in \mathbb{N},$$

and let $z = (z_k)_{k\in\mathbb{N}}$. Then, using the fact that $(p'-1)\,p = p'$, we compute that

$$\|z\|_p^p \;=\; \sum_{k=1}^{\infty}\biggl(\frac{|y_k|^{p'-1}}{\|y\|_{p'}^{p'-1}}\biggr)^{p} \;=\; \sum_{k=1}^{\infty}\frac{|y_k|^{p'}}{\|y\|_{p'}^{p'}} \;=\; \frac{1}{\|y\|_{p'}^{p'}}\sum_{k=1}^{\infty}|y_k|^{p'} \;=\; 1.$$

Hence z is a unit vector in ℓ^p. Also,

$$\begin{aligned}
|\mu_y(z)| \;=\; \biggl|\sum_{k=1}^{\infty} z_k y_k\biggr| \;&=\; \sum_{k=1}^{\infty}\frac{\alpha_k\,|y_k|^{p'-1}\,y_k}{\|y\|_{p'}^{p'-1}}\\
&=\; \sum_{k=1}^{\infty}\frac{|y_k|^{p'}}{\|y\|_{p'}^{p'-1}}\\
&=\; \frac{\|y\|_{p'}^{p'}}{\|y\|_{p'}^{p'-1}}\\
&=\; \|y\|_{p'}.
\end{aligned}$$

Since z is one of the unit vectors, this shows that $\|\mu_y\| \ge \|y\|_{p'}$. □

As a corollary of Theorem 6.10.1, we see that the mapping $T\colon \ell^{p'} \to (\ell^p)^*$ defined by

$$T(y) \;=\; \mu_y, \qquad y \in \ell^{p'},$$

is a linear isometry (but we do not know yet whether T is surjective).

Remark 6.10.2. If we wished to mimic the ℓ^2 situation more closely, we could have included complex conjugates in equation (6.28) and defined $\mu_y(x) = \sum_{k=1}^{\infty} x_k\,\overline{y_k}$. In this case the mapping $T(y) = \mu_y$ would be an antilinear isometry, just as it was when we considered $p = 2$. ♢

Next we will prove that the mapping T is surjective *when p is finite*. Consequently $(\ell^p)^* \cong \ell^{p'}$ for $1 \le p < \infty$, i.e., the dual space of ℓ^p is isometrically isomorphic to $\ell^{p'}$ for finite p.

Theorem 6.10.3 (Dual of ℓ^p). *Fix $1 \le p < \infty$. Using the notation introduced before, the mapping $T\colon \ell^{p'} \to (\ell^p)^*$ given by $T(y) = \mu_y$ is surjective, and consequently T is an isometric isomorphism of $\ell^{p'}$ onto $(\ell^p)^*$.*

Proof. We proved that T is isometric in Theorem 6.10.1, so it only remains to prove that T is surjective. We will do this for $1 < p < \infty$, and assign the proof for $p = 1$ as Problem 6.10.6.

Choose any $\mu \in (\ell^p)^*$, i.e., assume that $\mu\colon \ell^p \to \mathbb{F}$ is a bounded linear functional. Define scalars

$$y_k \;=\; \mu(\delta_k), \qquad k \in \mathbb{N},$$

where δ_k is the kth standard basis vector. Then define a vector y by setting $y = (y_k)_{k\in\mathbb{N}}$. If $y_k = 0$ for every k then μ is the zero functional (why?), so

$\mu = \mu_y$ in this case. Therefore we can assume that not every y_k is zero. We will show that y belongs to $\ell^{p'}$ and $\mu = \mu_y$.

For each $k \in \mathbb{N}$, let α_k be a scalar such that $|\alpha_k| = 1$ and $\alpha_k y_k = |y_k|$. Set

$$z_k = \alpha_k |y_k|^{p'-1}, \qquad k \in \mathbb{N}.$$

Fix an integer $N \in \mathbb{N}$, and let

$$w_N = \sum_{k=1}^{N} z_k \delta_k = (z_1, \dots, z_N, 0, 0, \dots).$$

Then $w_N \in \ell^p$, and since μ is linear we have

$$\mu(w_N) = \sum_{k=1}^{N} z_k \, \mu(\delta_k) = \sum_{k=1}^{N} z_k \, y_k = \sum_{k=1}^{N} \alpha_k \, |y_k|^{p'-1} \, y_k = \sum_{k=1}^{N} |y_k|^{p'}.$$

Since μ is bounded, it follows that

$$\sum_{k=1}^{N} |y_k|^{p'} = \mu(w_N) \le \|\mu\| \, \|w_N\|_p = \|\mu\| \left(\sum_{k=1}^{N} |y_k|^{p'} \right)^{1/p}.$$

The sums on the preceding line are nonzero for all large enough N (because not every y_k is zero), so by dividing we obtain, for all large enough N,

$$\left(\sum_{k=1}^{N} |y_k|^{p'} \right)^{1/p'} = \left(\sum_{k=1}^{N} |y_k|^{p'} \right)^{1-\frac{1}{p}} = \frac{\displaystyle\sum_{k=1}^{N} |y_k|^{p'}}{\left(\displaystyle\sum_{k=1}^{N} |y_k|^{p'} \right)^{1/p}} \le \|\mu\|.$$

Letting $N \to \infty$, we conclude that

$$\|y\|_{p'} = \left(\sum_{k=1}^{\infty} |y_k|^{p'} \right)^{1/p'} \le \|\mu\| < \infty.$$

Hence $y \in \ell^{p'}$. Therefore μ_y, the functional determined by y, is a bounded linear functional on ℓ^p. Finally, if $x \in \ell^p$ then (since μ is linear and continuous),

$$\mu(x) = \mu\left(\sum_{k=1}^{\infty} x_k \delta_k \right) = \sum_{k=1}^{\infty} x_k \mu(\delta_k) = \sum_{k=1}^{\infty} x_k y_k = \mu_y(x).$$

Thus $\mu = \mu_y = T(y)$, so T is surjective. $\square$

In contrast, the next theorem states that T is *not surjective* when $p = \infty$. Therefore $(\ell^\infty)^*$ is not isometrically isomorphic to ℓ^1.

Theorem 6.10.4. *Using the notation from before, there exists a bounded linear functional $\nu \in (\ell^\infty)^*$ such that $\nu \neq \mu_y$ for every $y \in \ell^1$. Consequently the mapping $T\colon \ell^1 \to (\ell^\infty)^*$ defined in Lemma 6.10.1 is not surjective.* $\diamondsuit$

A typical proof of Theorem 6.10.4 employs the *Hahn–Banach Theorem* or another major theorem from functional analysis (for a sketch of how this works, see Problem 6.10.4).

Problem 6.10.7 gives another interesting characterization of a dual space. Specifically, it shows that the dual space of c_0 is isometrically isomorphic to ℓ^1, i.e., $(c_0)^* \cong \ell^1$.

Problems

6.10.5. Fix $1 \le p \le \infty$, and suppose that $A = [a_{ij}]_{i,j\in\mathbb{N}}$ is an infinite matrix such that each row $(a_{ij})_{j\in\mathbb{N}}$ belongs to $\ell^{p'}$. Prove that if $x \in \ell^p$, then the product Ax defined by equation (6.13) exists. Must A map ℓ^p into ℓ^q for some q?

6.10.6. (a) Complete the proof of Theorem 6.10.1 for $p = 1$ and $p = \infty$.

(b) Complete the proof of Theorem 6.10.3 for the case $p = 1$.

6.10.7. Given $y = (y_k)_{k\in\mathbb{N}} \in \ell^1$, define $\mu_y\colon c_0 \to \mathbb{F}$ by

$$\mu_y(x) \;=\; \sum_{k=1}^{\infty} x_k\, y_k, \qquad x = (x_k)_{k\in\mathbb{N}} \in c_0.$$

Prove that $T(y) = \mu_y$ is an isometric isomorphism of ℓ^1 onto ${c_0}^*$, and therefore ${c_0}^* \cong \ell^1$.

6.10.8. (a) Let c be the space of all convergent sequences, i.e.,

$$c \;=\; \Bigl\{x = (x_k)_{k\in\mathbb{N}} \in \ell^\infty : \lim_{k\to\infty} x_k \text{ exists}\Bigr\}.$$

Show that c is a closed subspace of ℓ^∞ and therefore is a Banach space with respect to the ℓ^∞-norm.

(b) Define $\mu\colon c \to \mathbb{F}$ by

$$\mu(x) \;=\; \lim_{k\to\infty} x_k, \qquad x = (x_k)_{k\in\mathbb{N}} \in c.$$

Show that $\mu \in c^*$, i.e., μ is a bounded linear functional on c.

(c) The *Hahn–Banach Theorem* implies that there exists a bounded linear functional $\nu \in (\ell^\infty)^*$ that *extends* μ, i.e., ν is a bounded linear functional on ℓ^∞ and $\nu(x) = \mu(x)$ for $x \in c$. Show that $\nu \neq \mu_y$ for any $y \in \ell^1$, where $\mu_y \in (\ell^\infty)^*$ is the functional defined by

$$\mu_y(x) \;=\; \sum_{k=1}^{\infty} x_k\, y_k, \qquad x = (x_k)_{k\in\mathbb{N}} \in \ell^\infty.$$

Conclude that the mapping $T(y) = \mu_y$ given in Theorem 6.10.4 is not surjective. Remark: For a discussion of the Hahn–Banach Theorem see [Heil11, Ch. 2]; for a proof of Hahn–Banach see [Con90] or [Fol99].

(d) Find a vector $\delta_0 \in c$ such that $\{\delta_n\}_{n\geq 0}$ is a Schauder basis for c.

(e) Show that $c^* \cong \ell^1$. Remark: We also have ${c_0}^* \cong \ell^1$ by Problem 6.10.7. Yet, even though c and c_0 the same dual space, c is not isometrically isomorphic to c_0 (for a proof, see [Heil11, Exer. 4.22]).

Chapter 7
More on Operator Theory

7.1 Notation for Linear Functionals

The Riesz Representation Theorem for Hilbert spaces tells us that every vector in H is uniquely identified with a continuous linear functional in H^*. However, because the identification $g \mapsto \mu_g$ is antilinear, the standard notation for denoting functions can be confusing in this context. If μ is a functional on H, then the standard way to denote the value of μ at f is to write $\mu(f)$. Unfortunately, if $\mu = \mu_g$ and we define $c\mu$ by $(c\mu)(f) = c\mu(f)$, then $c\mu$ is *not* the functional identified with cg. Instead, Theorem 6.9.2(c) tells us that the functional identified with cg is $\mu_{cg} = \bar{c}\mu_g$. If we want $c\mu$ to denote *this* functional then we must declare that $(c\mu)(f) = \bar{c}\,\mu(f)$. This is inconsistent with the standard interpretation of the meaning of $c\mu$. This unpleasant inconsistency is a consequence of the fact that an inner product is antilinear in the second variable.

To alleviate this problem, we introduce an alternative notation for functionals. Let X be a normed space. Given a functional μ on X, instead of writing $\mu(f)$ we will denote the value that μ takes at $f \in X$ by

$$\langle f, \mu \rangle.$$

If $X = H$ is a Hilbert space, then by the Riesz Representation we know that μ is canonically identified with some element $g \in H$, and the value that μ takes at f is the inner product of f with g. Hence in this case we have $\langle f, \mu \rangle = \langle f, g \rangle$, so our notation is consistent with our identification of μ with g. On the other hand, if X is a generic normed space then there need not be any inner product on X, and the symbols $\langle f, \mu \rangle$ simply denote the value that μ takes at f. To maintain consistency, we *declare* that the notation $\langle f, \mu \rangle$ is linear as a function of f but antilinear as a function of μ. That is, when we use this notation, the functional $c\mu$ is defined to be the functional whose rule is

$$\langle f, c\mu\rangle \;=\; \overline{c}\,\langle f,\mu\rangle, \qquad f \in X.$$

We will see below some concrete illustrations of why this notation is convenient, but first we give a precise formalization of both the standard notation $\mu(f)$ and our alternative notation $\langle f,\mu\rangle$.

Notation 7.1.1 (Notation for Linear Functionals). Let X be a normed linear space. Given a fixed linear functional $\mu\colon X \to \mathbb{C}$, we use the following two notations to denote the image of f under μ.

(a) We write $\mu(f)$ to denote the image of f under μ, with the understanding that this notation is linear in both f and μ, i.e.,

$$\mu(\alpha f + \beta g) \;=\; \alpha\mu(f) + \beta\mu(g)$$

and

$$(\alpha\mu + \beta\nu)(f) \;=\; \alpha\mu(f) + \beta\nu(f).$$

(b) We write $\langle f,\mu\rangle$ to denote the image of f under μ, with the understanding that this notation is linear in f but antilinear in μ, i.e.,

$$\langle \alpha f + \beta g, \mu\rangle \;=\; \alpha\langle f,\mu\rangle + \beta\langle g,\mu\rangle$$

while

$$\langle f, \alpha\mu + \beta\nu\rangle \;=\; \overline{\alpha}\langle f,\mu\rangle + \overline{\beta}\langle f,\nu\rangle. \tag{7.1}$$

This will be our preferred notation for linear functionals throughout the remainder of this volume. ◇

In summary, the notation $\langle f,\mu\rangle$ simply denotes the image of f under the linear functional μ. However, using this notation we consider the pairing of a vector f with a linear functional μ to be a generalization of the inner product on a Hilbert space, which means that it is a *sesquilinear form* on $X \times X^*$ that is linear as a function of f but antilinear as a function of μ. The notation $\mu(f)$ also denotes the image of f under μ, but this notation is a *bilinear form* that is linear both as a function of f and as a function of μ.

We give some examples that illustrate the use of this notation.

Example 7.1.2. Let $\delta\colon C_b(\mathbb{R}) \to \mathbb{C}$ be the functional constructed in Example 6.3.2. Using the standard notation, the rule for this functional is

$$\delta(f) \;=\; f(0), \qquad f \in C_b(\mathbb{R}).$$

Using our alternative notation, we write the rule for this functional as

$$\langle f,\delta\rangle \;=\; f(0), \qquad f \in C_b(\mathbb{R}).$$

The symbols $\delta(f)$ and $\langle f,\delta\rangle$ both denote the scalar $f(0)$.

When we use the standard notation, $i\delta$ denotes the functional whose rule is
$$(i\delta)(f) \;=\; i\,\delta(f) \;=\; i f(0).$$
When we use our alternative notation, $i\delta$ denotes the functional whose rule is
$$\langle f, i\delta\rangle \;=\; -i\,\delta(f) \;=\; -i f(0).$$
The relationship between δ and $i\delta$ depends on which notation we choose to use. This choice is purely a matter of convenience. For our purposes, the alternative notation is more convenient and will be used henceforth. $\diamondsuit$

Example 7.1.3. Let $\mu\colon X \to \mathbb{C}$ be a linear functional on a normed space X. According to Notation 7.1.1, $\mu(f)$ and $\langle f, \mu\rangle$ both denote the image of $f \in X$ under μ. Since μ is a functional, $\mu(f) = \langle f, \mu\rangle$ is a *scalar* for each $f \in X$. Therefore the operator norm of the functional μ expressed in these two notations is
$$\|\mu\| \;=\; \sup_{\|f\|=1} |\mu(f)| \;=\; \sup_{\|f\|=1} |\langle f, \mu\rangle|.$$
Using the sesquilinear form notation, we have
$$|\langle f, \mu\rangle| \;\le\; \|f\|\,\|\mu\|, \qquad f \in X.$$
In this notation, to prove that a bounded linear functional is continuous, we would assume that $f_n \to f$ in X, and then write
$$\lim_{n\to\infty} \bigl|\langle f, \mu\rangle - \langle f_n, \mu\rangle\bigr| \;=\; \lim_{n\to\infty} \bigl|\langle f - f_n, \mu\rangle\bigr| \;\le\; \lim_{n\to\infty} \|f - f_n\|\,\|\mu\| \;=\; 0.$$
This shows that $\langle f_n, \mu\rangle \to \langle f, \mu\rangle$ as $n \to \infty$, which implies that μ is continuous. $\diamondsuit$

7.2 The Dual of $L^p(E)$

There are several results that are referred to as a "Riesz Representation Theorem." We saw the Riesz Representation Theorem for Hilbert spaces in Theorem 6.9.2. Another Riesz Representation Theorem characterizes the dual space of $C_0(X)$ where X is a locally compact Hausdorff topological space. Specifically, $C_0(X)^*$ is isomorphic to $M_b(X)$, the space of bounded Radon measures on X. If we take $X = \mathbb{N}$ and place the discrete topology on $\mathbb{N}$, then $C_0(X) = c_0$ and $M_b(X) \cong \ell^1$. Hence $c_0{}^* \cong \ell^1$ (a direct proof of this fact is asked for in Problem 6.10.7).

The characterization of the dual space of $L^p(E)$ parallels the characterization for ℓ^p. First we give an exercise that derives some facts about linear functionals on $L^p(E)$.

Exercise 7.2.1. Let E be a measurable subset of $\mathbb{R}^d$ with $|E| > 0$, and fix $1 \le p \le \infty$. Given $g \in L^{p'}(E)$, define a functional $\mu_g \colon L^p(E) \to \mathbb{C}$ by

$$\langle f, \mu_g \rangle \;=\; \int_E f(x)\,\overline{g(x)}\,dx, \qquad f \in L^p(E). \tag{7.2}$$

Prove the following statements.

(a) $\mu_g \in L^p(E)^*$, and $\|\mu_g\| = \|g\|_{p'}$.

(b) The mapping $T \colon L^{p'}(E) \to L^p(E)^*$ defined by $T(g) = \mu_g$ is an antilinear isometry. $\diamondsuit$

If we identify g with μ_g, then Exercise 7.2.1 tells us that $L^{p'}(E) \subseteq L^p(E)^*$. We will see that, just as for the ℓ^p spaces, equality holds if p is finite, but not for $p = \infty$.

First we will need a lemma. To motivate this, fix $g \in L^{p'}(E)$. The fact that the operator norm of g acting as a functional on $L^p(E)$ equals to $L^{p'}$-norm of g tells us that

$$\|g\|_{p'} \;=\; \sup_{\|f\|_p = 1} |\langle f, g \rangle|, \tag{7.3}$$

where

$$\langle f, g \rangle \;=\; \int_e f(x)\,\overline{g(x)}\,dx.$$

However, suppose that we did not know that g was p'-integrable. Suppose we only knew that there was some appropriate subset S of $L^p(E)$ such that $\langle f, g \rangle$ was well-defined for each function $f \in S$ and

$$\sup_{f \in S,\, \|f\|_p = 1} |\langle f, g \rangle| \;<\; \infty.$$

Could we conclude from this that g belongs to $L^{p'}(E)$? This is a kind of converse to the fact given in equation (7.3), and it is the issue addressed in the following theorem. We state this result in a form that can be easily generalized to measures other than Lebesgue measure; for the specific case of Lebesgue measure we could simplify things by letting the set S be the set of all compactly supported simple functions, and statement (b) could be replaced by the assumption that $g \in L^1_{\mathrm{loc}}(E)$, i.e., $g\chi_K$ is integrable for every compact set $K \subseteq \mathbb{R}^d$.

Theorem 7.2.2. *Let E be a Lebesgue measurable subset of $\mathbb{R}^d$, and fix $1 \le p \le \infty$. Let S be the set of simple functions on E that vanish outside a set of finite measure, i.e.,*

$$S \;=\; \Bigl\{\phi \colon E \to \mathbb{C} : \phi \textit{ is simple and } |\{\phi \ne 0\}| < \infty\Bigr\}.$$

Suppose that:

(a) $g\colon E \to \mathbb{C}$ *is Lebesgue measurable,*

(b) $\phi\, g \in L^1(E)$ *for each* $\phi \in S$, *and*

(c) $M_g \;=\; \sup\biggl\{\biggl|\int_E \phi\, g\biggr| : \phi \in S,\ \|\phi\|_p = 1\biggr\} \;<\; \infty.$

Then $g \in L^{p'}(E)$ *and* $\|g\|_{p'} = M_g$.

Proof. By Hölder's Inequality,

$$\biggl|\int_E \phi\, g\biggr| \;\le\; \|\phi\|_p\, \|g\|_{p'},$$

and therefore $M_g \le \|g\|_{p'}$.

Let $\alpha(x)$ be a complex scalar of unit modulus such that

$$|g(x)| \;=\; \alpha(x)\, g(x), \qquad x \in E.$$

If $\phi \in S$ then $\phi \in L^p(E)$, so by the definition of M_g we have

$$\biggl|\int_E \phi\, g\biggr| \;\le\; M_g\, \|\phi\|_p.$$

We will extend this formula to bounded measurable functions f on E that vanish outside a set of finite measure.

Suppose that f is bounded and measurable, and that f is nonzero only on a set A of finite measure. Then $\chi_A \in S$, so $\chi_A\, g \in L^1(E)$ by hypothesis (b). Since f is bounded and zero outside of A, we therefore have $fg \in L^1(E)$.

By our standard approximation theorems we know that there exist simple functions ϕ_k such that $\phi_k \to f$ pointwise a.e. and $|\phi_k| \le |f|$. Each ϕ_k belongs to S, $\phi_k\, g \to fg$ pointwise a.e., and $|\phi_k\, g| \le |fg| \in L^1(E)$. The Dominated Convergence Theorem therefore implies that

$$\biggl|\int_E fg\biggr| \;=\; \lim_{k\to\infty} \biggl|\int_E \phi_k\, g\biggr| \;\le\; \lim_{k\to\infty} M_g\, \|\phi_k\|_p \;\le\; M_g\, \|f\|_p. \tag{7.4}$$

Case $1 < p < \infty$. In this case we have $1 < p' < \infty$. Let ϕ_k be simple functions on E such that $0 \le \phi_k \nearrow |g|^{p'}$. If necessary, by replacing ϕ_k with $\phi_k \cdot \chi_{E\cap[-k,k]^d}$ we may assume that each ϕ_k vanishes outside a set of finite measure. Since each ϕ_k is nonnegative, we can define

$$g_k \;=\; \alpha \cdot \phi_k^{1/p}.$$

Motivation:

$$|g_k| \;=\; \phi_k^{1/p} \;=\; \phi_k^{1-\frac{1}{p'}} \;\approx\; \bigl(|g|^{p'}\bigr)^{1-\frac{1}{p'}} \;=\; |g|^{p'-1},$$

and therefore

$$|g_k\, g| \;\approx\; |g|^{p'}.$$

More precisely, since $|g_k| = \phi_k^{1/p}$, we have

$$\|g_k\|_p \;=\; \biggl(\int_E |g_k|^p\biggr)^{1/p} \;=\; \biggl|\int_E \phi_k\biggr|^{1/p} \;=\; \|\phi_k\|_1^{1/p}$$

and

$$\phi_k \;=\; \phi_k^{1/p}\,\phi_k^{1/p'} \;\le\; \phi_k^{1/p}\,|g| \;=\; \phi_k^{1/p}\,\alpha\, g \;=\; g_k\, g.$$

Further, each g_k is bounded and vanishes outside a set of finite measure, so by equation (7.4) we have

$$\begin{aligned} \|\phi_k\|_1^{1/p}\,\|\phi_k\|_1^{1/p'} \;=\; \|\phi_k\|_1 \;&=\; \int_E \phi_k \\ &\le\; \int_E g_k\, g \;\le\; M_g\,\|g_k\|_p \;=\; M_g\,\|\phi_k\|_1^{1/p}. \end{aligned}$$

If $g = 0$ then there is nothing to prove, so assume that g is not the zero function. Then ϕ_k is not the zero function for all large enough k, so we can divide through in the equation above to obtain

$$\|\phi_k\|_1^{1/p'} \;\le\; M_g.$$

Applying Fatou's Lemma,

$$\|g\|_{p'}^{p'} = \int_E |g|^{p'} = \int_E \liminf_{k\to\infty} |\phi_k| \;\le\; \liminf_{k\to\infty} \int_E |\phi_k| \;=\; \liminf_{k\to\infty} \|\phi_k\|_1 \;\le\; M_g^{p'}.$$

Thus $g \in L^{p'}(E)$ and $\|g\|_{p'} \le M_g$.

Case $p = 1$, $p' = \infty$. Fix $\varepsilon > 0$, and define

$$A \;=\; \bigl\{|g| \ge M_g + \varepsilon\bigr\}.$$

If $|A| > 0$, choose $B \subseteq A$ with $0 < |B| < \infty$. Define

$$f \;=\; \alpha \cdot \chi_B.$$

Then f is bounded and vanishes outside a set of finite measure, so by equation (7.4) we have

$$\biggl|\int_E fg\biggr| \;\le\; M_g\,\|f\|_1 \;=\; |B|.$$

However,

$$\int_E fg \;=\; \int_B \alpha\, g \;=\; \int_B |g| \;\ge\; (M_g + \varepsilon)\,|B|.$$

Therefore

$$(M_g + \varepsilon)\,|B| \ \le\ M_g\,|B|,$$

which is a contradiction. Therefore we must have $|A| = 0$. Hence $|g| \le M_g + \varepsilon$ a.e. Since this is true for every $\varepsilon > 0$, we conclude that $\|g\|_\infty \le M_g < \infty$.

Case $p = \infty$, $p' = 1$. Set

$$g_k \ =\ \alpha\,\chi_{E\cap[-k,k]^d}.$$

Then g_k is bounded and vanishes outside a set of finite measure, so

$$\int_{E\cap[-k,k]^d} |g| \ =\ \int_E g_k\, g \ \le\ M_g\,\|g_k\|_\infty \ =\ M_g.$$

Since this is true for every k,

$$\|g\|_1 \ =\ \int_E |g| \ \le\ M_g \ <\ \infty.$$

Hence $g \in L^1(E)$. □

Remark 7.2.3. Theorem 7.2.2 extends to arbitrary positive measure spaces (X, Σ, μ) if we assume either that the set $S_g = \{g \ne 0\}$ is σ-finite, or that the measure μ is semifinite. See Theorem 6.14 in Folland for details. ♢

Theorem 7.2.4 (Dual Space of $L^p(E)$). *Let E be a Lebesgue measurable subset of $\mathbb{R}^d$, and fix $1 \le p < \infty$. For each $g \in L^{p'}(E)$, define μ_g by*

$$\langle f, \mu_g\rangle \ =\ \int_E f(x)\,\overline{g(x)}\,dx, \qquad f \in L^p(E). \tag{7.5}$$

Then the mapping $T\colon L^{p'}(E) \to L^p(E)^*$ defined by $T(g) = \mu_g$ is an antilinear isometric isomorphism of $L^{p'}(E)$ onto $L^p(E)^*$.

Proof. We already know that T is an antilinear isometric map of $L^{p'}(E)$ into $L^p(E)^*$. Therefore T is injective, and so we only need to prove that T is surjective.

Case 1: $|E| < \infty$. Suppose that $\mu \in L^p(E)^*$, i.e., $\mu\colon L^p(E) \to \mathbb{C}$ is bounded and linear. Define

$$\nu(A) \ =\ \langle \chi_A, \mu\rangle, \qquad \text{measurable } A \subseteq E.$$

Our immediate goal is to show that ν is a complex measure on $(E, \mathcal{L}_E)$, where $\mathcal{L}_E$ is the σ-algebra of Lebesgue measurable subsets of E.

First, since E has finite measure we have $\chi_A \in L^p(E)$ for each measurable $A \subseteq E$. Therefore $\nu(A)$ is a well-defined complex scalar for each $A \subseteq E$, and also $\nu(\emptyset) = \langle 0, \mu\rangle = 0$.

Second, suppose that $A_1, A_2, \dots$ are disjoint measurable subsets of E, and let $A = \cup A_j$. Then

$$\chi_A(x) \;=\; \sum_{j=1}^{\infty} \chi_{A_j}(x), \qquad x \in E,$$

since for each x the series on the right has at most one nonzero term.

Exercise: Show that the series $\chi_A = \sum_{j=1}^{\infty} \chi_{A_j}$ converges *in L^p-norm* and not just pointwise (the fact that $|E| < \infty$ is important here).

Consequently, using both the linearity *and* the continuity of μ, we have

$$\nu(A) \;=\; \langle \chi_A, \mu\rangle \;=\; \biggl\langle \sum_{j=1}^{\infty} \chi_{A_j},\, \mu \biggr\rangle \;=\; \sum_{j=1}^{\infty} \langle \chi_{A_j}, \mu\rangle \;=\; \sum_{j=1}^{\infty} \nu(A_j).$$

Therefore ν is countably additive, and hence is a complex measure on E.

Moreover, if $A \subseteq E$ satisfies $|A| = 0$, then $\chi_A = 0$ a.e., so we have

$$\nu(A) \;=\; \langle \chi_A, \mu\rangle \;=\; \langle 0, \mu\rangle \;=\; 0.$$

Hence $\nu \ll dx$.

Because dx is a bounded measure on E, the Radon–Nikodym Theorem implies that there exists a $g \in L^1(E)$ such that $d\nu = \overline{g}\, dx$. We will show that $g \in L^{p'}(E)$ and that $\mu = \mu_g = T(g)$, which will show that T is surjective.

Now, the symbols $d\nu = \overline{g}\, dx$ mean that

$$\nu(A) \;=\; \int_A \overline{g(x)}\, dx, \qquad \text{all measurable } A \subseteq E.$$

Consequently, for each measurable $A \subseteq E$ we have

$$\langle \chi_A, \mu\rangle \;=\; \nu(A) \;=\; \int_E \chi_A\, d\nu \;=\; \int_E \chi_A\, \overline{g}.$$

By linearity, we therefore have for each simple function ϕ on E that

$$\langle \phi, \mu\rangle \;=\; \int_E \phi\, \overline{g}.$$

Note that $\phi \in L^p(E)$ since ϕ is bounded and $|E| < \infty$. Since μ is bounded on $L^p(E)$, we therefore have that

$$\biggl| \int_E \phi\, \overline{g} \biggr| \;=\; |\langle \phi, \mu\rangle| \;\le\; \|\mu\|\, \|\phi\|_p.$$

It therefore follows from Theorem 7.2.2 that $g \in L^{p'}(E)$. Hence μ_g is a bounded linear functional on $L^p(E)$, i.e., $\mu_g \in L^p(E)^*$. For each simple function ϕ we have

$$\langle \phi, \mu_g \rangle \;=\; \int_E \phi\, \overline{g} \;=\; \langle \phi, \mu \rangle,$$

i.e., μ and μ_g agree on the simple functions. Since these operators are continuous and since the simple functions are dense in $L^p(E)$, we conclude that $\mu = \mu_g$. This finishes the proof for this case.

Case 2: $|E| = \infty$. Set $E_k = E \cap [-k,k]^d$ (the important fact here being that Lebesgue measure is σ-finite). Let μ_k be the restriction of μ to $L^p(E_k)$. Then $\mu_k \in L^p(E_k)^*$, so it follows from Case 1 that there exists a function $g_k \in L^{p'}(E_k)$ such that $\mu_k = \mu_{g_k}$. By taking functions to be zero outside of E_k, we can consider $L^p(E_k)$ to be a subspace of $L^p(E)$. Therefore

$$\langle f, \mu \rangle \;=\; \langle f, \mu_k \rangle \;=\; \int_{E_k} f\, \overline{g_k}, \qquad f \in L^p(E_k).$$

Furthermore,

$$\|g_k\|_{p'} \;=\; \|\mu_k\| \;\le\; \|\mu\|,$$

where μ_k is the operator norm of μ restricted to $L^p(E_k)$, and $\|\mu\|$ is the operator norm of μ on $L^p(E)$.

Now, if f is any function in $L^p(E_k)$ then $f \in L^p(E_{k+1})$ as well, and since f is zero outside of E_k, we have

$$\int_{E_k} f\, \overline{g_k} \;=\; \langle f, \mu \rangle \;=\; \int_{E_{k+1}} f\, \overline{g_{k+1}} \;=\; \int_{E_k} f\, \overline{g_{k+1}}.$$

Consequently, we must have $g_k = g_{k+1}$ for almost every $x \in E_k$. Hence we can define a function g on E by setting $g(x) = g_k(x)$ if $x \in E_k$. Then g is measurable, and if $1 < p < \infty$ then Fatou's Lemma implies that

$$\|g\|_{p'} \;\le\; \liminf_{k \to \infty} \|g_k\|_{p'} \;\le\; \|\mu\| \;<\; \infty.$$

If $p = 1$ then $p' = \infty$ and

$$\|g\|_\infty \;=\; \sup_k \|g_k\|_\infty \;\le\; \|\mu\| \;<\; \infty.$$

In any case, we conclude that $g \in L^{p'}(E)$.

Finally, if $f \in L^p(E)$ then $f\chi_{E_k} \in L^p(E_k)$, so

$$\langle f\chi_{E_k}, \mu \rangle \;=\; \int_{E_k} f\, \overline{g_k} \;=\; \int_{E_k} f\, \overline{g}.$$

However, $f\chi_{E_k} \to f$ in L^p-norm and μ is continuous, so

$$\langle f\chi_{E_k}, \mu \rangle \;\to\; \langle f, \mu \rangle.$$

On the other hand, $f\,\overline{g} \in L^1(E)$ by Hölder's Inequality, so by the Lebesgue Dominated Convergence Theorem,

$$\int_{E_k} f\,\overline{g} \;\to\; \int_E f\,\overline{g}.$$

Hence

$$\langle f, \mu\rangle \;=\; \int_E f\,\overline{g} \;=\; \langle f, \mu_g\rangle.$$

This is true for all $f \in L^p(E)$, so we conclude that $\mu = \mu_g$. $\square$

Remark 7.2.5. Theorem 7.2.4 extends to arbitrary positive measure spaces (X, Σ, μ). For $1 < p < \infty$ we can even prove this without any restrictions on μ, i.e., even if μ is not σ-finite. However, for $p = 1$ we must assume that μ is σ-finite. See Theorem 6.15 in Folland for details. $\diamondsuit$

Theorem 7.2.4 tells us that, for indices in the range $1 \le p < \infty$, the dual space of $L^p(E)$ is isometrically isomorphic to $L^{p'}(E)$. If $p = \infty$, then the map $T\colon L^1(E) \to L^\infty(E)^*$ given by $T(g) = \mu_g$ is an antilinear isometry, but it is not surjective. In this sense, $L^1(E)$ has a canonical image within $L^\infty(E)^*$, but there are functionals in $L^\infty(E)^*$ that do not correspond to elements of $L^1(E)$.

In light of Theorem 7.2.4, we usually identify $L^{p'}(E)$ with $L^p(E)^*$ when p is finite, and we also identify $L^1(E)$ with its image in $L^\infty(E)^*$. Abusing notation, we write

$$L^p(E)^* \;=\; L^{p'}(E) \quad \text{for } 1 \le p < \infty,$$

and

$$L^1(E) \;\subsetneq\; L^\infty(E)^*,$$

with the understanding that these hold in the sense of the identification of $g \in L^{p'}(E)$ with $\mu_g \in L^p(E)^*$.

Problems

7.2.6. Let X be a normed space, and suppose that the series $f = \sum f_n$ converges in X. Show that if $\mu \in X^*$, then $\sum \langle f_n, \mu\rangle$ is a convergent series of scalars and $\langle f, \mu\rangle = \sum \langle f_n, \mu\rangle$.

7.2.7. Let H be a Hilbert space, and fix $\mu \in H^*$. Show that if $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal sequence in H, then $\bigl\{\langle e_n, \mu\rangle\bigr\}_{n\in\mathbb{N}} \in \ell^2$.

7.2.8. Let X be a normed space.

(a) Given $\mu \in X^*$, $\mu \ne 0$, show that if $h \notin \ker(\mu)$ then every $f \in X$ can be written uniquely as $f = g + ch$ where $g \in \ker(\mu)$ and $c \in \mathbb{C}$.

(b) Let μ be a linear functional on X. Prove that

$$\mu \text{ is bounded} \quad \Longleftrightarrow \quad \ker(\mu) \text{ is closed.}$$

7.2.9. Let H be a Hilbert space, and assume that $B\colon H \times H \to \mathbb{C}$ is a sesquilinear operator, i.e., it is linear in the first variable and antilinear in the second. Suppose that there is a constant $C > 0$ such that

$$|B(f,g)| \leq C\,\|f\|\,\|g\|, \qquad f,g \in H.$$

Prove that there is a bounded linear operator $A \in \mathcal{B}(H)$ such that

$$B(f,g) = \langle f, Ag\rangle \qquad \text{for all } f,g \in H.$$

7.2.10. Fix $0 < p < 1$, and recall that $L^p(\mathbb{R})$ is a complete metric space with respect to the distance $\mathrm{d}(f,g) = \|f-g\|_p^p$. This problem will determine the dual space of $L^p(\mathbb{R})$.

(a) Let μ be a linear functional on $L^p(\mathbb{R})$. Prove that μ is continuous if and only if it is bounded in the sense that there exists a constant $C > 0$ such that $|\langle f, \mu\rangle| \leq C\,\|f\|_p$ for all $f \in L^p(\mathbb{R})$.

(b) Suppose that μ is a continuous linear functional on $L^p(\mathbb{R})$. Set

$$F(x) = \bigl\langle \chi_{[0,x]},\, \mu\bigr\rangle, \qquad x \in \mathbb{R},$$

where if x is negative we let $[0,x] = [x,0]$. Prove that F must be constant.

(c) Show that $L^p(\mathbb{R})^* = \{0\}$.

Problems

7.2.11. Let $E \subseteq \mathbb{R}$ be measurable. Given $1 \leq p \leq \infty$ and a measurable *weight function* $w\colon E \to (0,\infty)$, the *weighted L^p space* $L^p_w(\mathbb{R})$ consists of all measurable functions $f\colon E \to \mathbb{C}$ for which $\|f\|_{p,w} = \|fw\|_p < \infty$. Prove that $L^p_w(\mathbb{R})$ is a Banach space with respect to the norm $\|\cdot\|_{p,w}$. Prove that if $1 \leq p < \infty$ then $L^p_w(E)^* \cong L^{p'}_{1/w}(E)$.

7.3 Banach Algebras

By definition, a vector space is closed under the operations of vector addition and scalar multiplication. Some vector spaces are also closed under a third "multiplication-like" operation that inputs a pair of vectors f, g and yields a "product" vector fg. If this operation satisfies the properties given in the next definition, then we call such a space an *algebra*.

Definition 7.3.1 (Algebra). An *algebra* is a vector space $\mathcal{A}$ such that for each pair of vectors $f, g \in \mathcal{A}$ there exists a unique vector $fg \in \mathcal{A}$ such that the following conditions are satisfied for all $f, g, h \in \mathcal{A}$ and $c \in \mathbb{C}$:

(a) $(fg)h = f(gh)$,

(b) $f(g+h) = fg + fh$ and $(f+g)h = fh + gh$, and

(c) $c(fg) = (cf)g = f(cg)$.

If $\mathcal{A}$ is an algebra and $fg = gf$ for all $f, g \in \mathcal{A}$, then we say that $\mathcal{A}$ is *commutative*. If there exists an element $e \in \mathcal{A}$ such that $ef = fe = f$ for every $f \in \mathcal{A}$, then $\mathcal{A}$ is an *algebra with identity*. ♢

An algebra need not be a normed space. However, if a normed space is an algebra *and* its product is submultiplicative in the sense given in the following definition, then we will called it a normed algebra.

Definition 7.3.2 (Banach Algebra). Suppose that $\mathcal{A}$ is a normed space that is an algebra. If the algebra product is *submultiplicative*, which means that

$$\forall\, f, g \in \mathcal{A}, \quad \|fg\| \;\le\; \|f\|\,\|g\|,$$

then we say that $\mathcal{A}$ is a *normed algebra*. A *Banach algebra* is a complete normed algebra, i.e., it is a normed algebra that is a Banach space. ♢

Here are some examples of Banach algebras.

Example 7.3.3. (a) Exercise 6.8.11 introduced the *convolution* of functions. If $f, g \in L^1(\mathbb{R})$, then their convolution is the function $f * g$ given by

$$(f * g)(x) \;=\; \int_{-\infty}^{\infty} f(y)\, g(x-y)\, dy.$$

Exercise 6.8.11 showed that $f * g \in L^1(\mathbb{R})$ and

$$\|f * g\|_1 \;\le\; \|f\|_1\, \|g\|_1.$$

Furthermore, convolution is both associative and commutative on $L^1(\mathbb{R})$. Therefore $L^1(\mathbb{R})$ is a commutative Banach algebra with respect to convolution. However, $L^1(\mathbb{R})$ *does not* contain an identity element for convolution.

(b) If X is a Banach space, then $\mathcal{B}(X)$, the space of all bounded linear operators that map X into itself, is also a Banach space. The composition of bounded operators is bounded, so $\mathcal{B}(X)$ is closed with respect to composition of operators. By Lemma 6.5.4, we have the submultiplicative property

$$\|AB\| \;\le\; \|A\|\,\|B\|$$

for all A, $B \in \mathcal{B}(X)$. Although composition of operators is associative, it is not commutative. Therefore $\mathcal{B}(X)$ is a noncommutative Banach algebra with respect to composition. Moreover, the identity operator I is an identity element for this algebra.

(c) $C_b(\mathbb{R})$ is a Banach space, and it is also closed with respect to pointwise products of functions. Further, we have the submultiplicative property

$$\|fg\|_\infty \ \le \ \|f\|_\infty \, \|g\|_\infty.$$

Pointwise products are both associative and commutative, and the constant function 1 is an identity for pointwise products. Therefore $C_b(\mathbb{R})$ is a commutative Banach algebra with identity, with respect to the operation of pointwise products.

(d) $C_0(\mathbb{R})$ is a commutative Banach algebra *without identity*, with respect to pointwise products of functions. $\diamondsuit$

A *subalgebra* of a Banach algebra is a subspace that is closed with respect to the product operation of the algebra. However, more important than subalgebras are *ideals*, which satisfy a more stringent closure requirement than subalgebras. Ideals are the "black holes" of the algebra, sucking any product of an algebra element with an ideal element into the ideal.

Definition 7.3.4 (Ideals). Let $\mathcal{A}$ be a Banach algebra, and let I be a subspace of $\mathcal{A}$.

(a) I is a *left ideal* in $\mathcal{A}$ if $fg \in I$ whenever $f \in \mathcal{A}$ and $g \in I$.

(b) I is a *right ideal* in $\mathcal{A}$ if $gf \in I$ whenever $f \in \mathcal{A}$ and $g \in I$.

(c) I is a *two-sided ideal* in $\mathcal{A}$, or simply an *ideal*, if $fg,\ gf \in I$ whenever $f \in \mathcal{A}$ and $g \in I$. $\diamondsuit$

For example, the subspace $C_c(\mathbb{R})$ is an ideal in $C_0(\mathbb{R})$ with respect to pointwise multiplication of functions. By Exercise 5.3.14, $C_c(\mathbb{R})$ is dense in $C_0(\mathbb{R})$. However, not all ideals are dense. For example, if E is any nonempty subset of $\mathbb{R}$, then

$$I \ = \ \bigl\{f \in C_0(\mathbb{R}) : f(x) = 0 \text{ for all } x \in E\bigr\}$$

is a proper, closed ideal in $C_0(\mathbb{R})$.

Given a vector f in a commutative Banach algebra $\mathcal{A}$, the next exercise shows how to construct the smallest ideal that contains f. What changes are needed if $\mathcal{A}$ is not commutative?

Exercise 7.3.5. Let $\mathcal{A}$ be a commutative Banach algebra.

(a) Prove that if $f \in \mathcal{A}$, then $f\mathcal{A} = \{fg : g \in \mathcal{A}\}$ is an ideal in $\mathcal{A}$. We say that $f\mathcal{A}$ is the *ideal generated by* f.

(b) Show by example that f need not belong to $f\mathcal{A}$. $\diamondsuit$

In addition to the product operation, some Banach algebras have yet another operation, similar to the complex conjugation operation on the field of complex numbers.

Definition 7.3.6 (Involution). An *involution* on a Banach algebra $\mathcal{A}$ is a function $f \mapsto f^*$ that maps $\mathcal{A}$ into itself and satisfies the following conditions for all vectors $f, g \in \mathcal{A}$ and all scalars $c \in \mathbb{C}$:

(a) $(f^*)^* = f$,

(b) $(fg)^* = g^* f^*$,

(c) $(f+g)^* = f^* + g^*$, and

(d) $(cf)^* = \bar{c} f^*$. $\diamondsuit$

Here are some examples.

Exercise 7.3.7. (a) Given $f \in C_b(\mathbb{R})$, define

$$f^*(x) \;=\; \overline{f(x)},$$

i.e., f^* is the complex conjugate of f. Show that $f \mapsto f^*$ defines an involution on the Banach algebras $C_b(\mathbb{R})$ and $C_0(\mathbb{R})$, with respect to pointwise products of functions.

(b) Given $f \in L^1(\mathbb{R})$, define

$$\widetilde{f}(x) \;=\; \overline{f(-x)}. \tag{7.6}$$

Prove that $f \mapsto \widetilde{f}$ defines an involution on $L^1(\mathbb{R})$, with respect to convolution of functions. $\diamondsuit$

We will see an example of an involution on a noncommutative Banach algebra in Section 8.2, when we introduce the adjoint operation on $\mathcal{B}(H)$, where H is a Hilbert space. In fact, Theorem 8.2.6 will show that this adjoint has the further property that $\|A^*A\| = \|A\|^2$ for every $A \in \mathcal{B}(H)$. According to the following definition, this says that $\mathcal{B}(H)$ is an example of a noncommutative C^*-*algebra*.

Definition 7.3.8 (C^*-algebra). A C^*-*algebra* is a Banach algebra $\mathcal{A}$ with an involution that satisfies $\|f^*f\| = \|f\|^2$ for all $f \in \mathcal{A}$. $\diamondsuit$

For example, $C_b(\mathbb{R})$ and $C_0(\mathbb{R})$ are C^*-algebras with respect to the involution $f^*(x) = \overline{f(x)}$. Problem 7.3.14 asks whether $L^1(\mathbb{R})$ is a C^*-algebra with respect to involution.

Problems

7.3.9. Given $0 < \alpha < 1$, let $C^\alpha(\mathbb{R})$ be the space of Hölder continuous functions introduced in Problem 5.3.17. This is a Banach space with respect to the norm

$$\|f\|_{C^\alpha} \;=\; \|f\|_{\mathrm{u}} \;+\; \sup_{x \neq y} \frac{|f(x)-f(y)|}{|x-y|^\alpha}. \tag{7.7}$$

Prove that $C^\alpha(\mathbb{R})$ is a Banach algebra with respect to the operation of pointwise products of functions.

7.3.10. Let I be a proper ideal in a commutative Banach algebra $\mathcal{A}$. Show that if $\mathcal{A}$ has an identity element e, then $e \notin I$.

7.3.11. Let $\mathcal{A}$ be a commutative Banach algebra with identity e. We say that a vector $f \in \mathcal{A}$ is *invertible* if there exists a vector $f^{-1} \in \mathcal{A}$ such that $ff^{-1} = e$. Show that if an ideal I contains an invertible element, then $I = \mathcal{A}$.

7.3.12. Show that if I is an ideal in a commutative Banach algebra $\mathcal{A}$, then its closure $\overline{I}$ is also an ideal.

7.3.13. Define

$$C_{ub}(\mathbb{R}) \;=\; \bigl\{f \in C_b(\mathbb{R}) : f \text{ is uniformly continuous}\bigr\}.$$

Prove that $C_{ub}(\mathbb{R})$ is a closed subalgebra of $C_b(\mathbb{R})$. Is it an ideal in $C_b(\mathbb{R})$?

7.3.14. Given $f \in L^1(\mathbb{R})$, let $\widetilde{f}(x) = \overline{f(-x)}$.

(a) Show that if $f \geq 0$ a.e., then $\|f * \widetilde{f}\,\|_1 = \|f\|_1^2$.

(b) Is $L^1(\mathbb{R})$ a C^*-algebra with respect to the involution $f \mapsto \widetilde{f}$?

7.4 The Fourier Transform on $L^1(\mathbb{R})$

The Fourier transform was discussed in more detail in Chapter 8 of Volume 1, and any proofs that are omitted in the discussion here are presented in detail in Volume 1.

If $f \in L^1(\mathbb{R})$ then its *Fourier transform* is the function $\widehat{f}$ defined by

$$\widehat{f}(\xi) \;=\; \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i \xi x}\, dx, \qquad \xi \in \mathbb{R}. \tag{7.8}$$

Because f is integrable we can see immediately that $\widehat{f}$ is bounded:

$$|\widehat{f}(\xi)| \;\leq\; \int_{-\infty}^{\infty} |f(x)\, e^{-2\pi i \xi x}|\, dx \;=\; \int_{-\infty}^{\infty} |f(x)|\, dx \;=\; \|f\|_1.$$

The *Riemann–Lebesgue Lemma* tells us that $\widehat{f} \in C_0(\mathbb{R})$, and furthermore

$$\|\widehat{f}\,\|_\infty \;\le\; \|f\|_1. \tag{7.9}$$

Consequently we can view the Fourier transform as an operator $\mathcal{F}$ that maps $L^1(\mathbb{R})$ into $C_0(\mathbb{R})$, i.e., $\mathcal{F}\colon L^1(\mathbb{R}) \to C_0(\mathbb{R})$ is defined by

$$\mathcal{F}(f) \;=\; \widehat{f}, \qquad f \in L^1(\mathbb{R}).$$

Equation (7.9) implies that $\mathcal{F}$ is bounded, and its operator norm satisfies

$$\|\mathcal{F}\| \;=\; \|\mathcal{F}\|_{L^1 \to C_0} \;\le\; 1.$$

By direct calculation, the Fourier transform of the box function $\chi = \chi_{[-\frac{1}{2},\frac{1}{2}]}$ is the *sinc function*

$$\widehat{\chi}(\xi) \;=\; \mathrm{sinc}(\xi) \;=\; \frac{\sin \pi\xi}{\pi\xi}. \tag{7.10}$$

Since $\|\chi\|_1 = 1$ and $\|\widehat{\chi}\,\|_\infty = 1$, it follows that the operator norm of $\mathcal{F}$ is precisely $\|\mathcal{F}\| = 1$. On the other hand, the Fourier transform of the square wave $\psi = \chi_{[0,\frac{1}{2})} - \chi_{[-\frac{1}{2},0]}$ is

$$\widehat{\psi}(\xi) \;=\; -2i\,\frac{\sin^2(\pi\xi/2)}{\pi\xi},$$

and we have the strict inequality

$$\|\widehat{\psi}\,\|_\infty \;<\; 1 \;=\; \|\psi\|_1.$$

Consequently, the Fourier transform $\mathcal{F}\colon L^1(\mathbb{R}) \to C_0(\mathbb{R})$ is not isometric.

The range of $\mathcal{F}$ is usually denoted by

$$A(\mathbb{R}) \;=\; \mathrm{range}(\mathcal{F}) \;=\; \bigl\{\widehat{f} : f \in L^1(\mathbb{R})\bigr\}.$$

By the Riemann–Lebesgue Lemma, $A(\mathbb{R}) \subseteq C_0(\mathbb{R})$, and the following result shows that the space of compactly supported, twice-differentiable functions is contained in $A(\mathbb{R})$. Therefore $A(\mathbb{R})$ is a dense subspace of $C_0(\mathbb{R})$.

Lemma 7.4.1. $C_c^2(\mathbb{R}) \subseteq A(\mathbb{R}) \subseteq C_0(\mathbb{R})$.

Proof. The smoother that f is, the faster that $\widehat{f}$ decays at infinity. In particular, if $f \in C_c^2(\mathbb{R})$ then it can be shown that there is a constant $C > 0$ such that

$$|\widehat{f}(\xi)| \;\le\; \frac{C}{|\xi|^2}, \qquad \xi \ne 0. \tag{7.11}$$

As $\widehat{f}$ is also bounded, it follows that $\widehat{f} \in L^1(\mathbb{R})$. The *inverse Fourier transform* of f is

$$\check{f}(\xi) \;=\; \widehat{f}(-\xi), \tag{7.12}$$

so we also have $\check{f} \in L^1(\mathbb{R})$. As both f and $\widehat{f}$ are integrable, the *Inversion Formula* implies that the Fourier transform of the integrable function $g = \check{f}$ is

$$\widehat{g} \;=\; (\check{f})^\wedge \;=\; f.$$

Consequently $f \in A(\mathbb{R})$. □

Although $A(\mathbb{R})$ is dense in $C_0(\mathbb{R})$, it can be shown that $A(\mathbb{R}) \neq C_0(\mathbb{R})$. Thus, while the Fourier transform maps $L^1(\mathbb{R})$ into $C_0(\mathbb{R})$, it is not surjective. For example, the function

$$F(\xi) \;=\; \begin{cases} 1/\ln\xi, & \xi > e, \\ \xi/e, & -e \le \xi \le e, \\ -1/\ln\xi, & \xi < -e, \end{cases}$$

belongs to $C_0(\mathbb{R}) \backslash A(\mathbb{R})$. There even exist compactly supported continuous functions that do not belong to $A(\mathbb{R})$. One example, constructed in [Her85], is

$$B(\xi) \;=\; \begin{cases} \frac{1}{n}\,\sin(2\pi 4^n \xi), & \frac{1}{2^{n+1}} \le |\xi| \le \frac{1}{2^n}, \\ 0, & \xi \le 0 \text{ or } |\xi| > \frac{1}{2}. \end{cases}$$

The letter B is for "butterfly"; see the illustration in Figure 7.1.

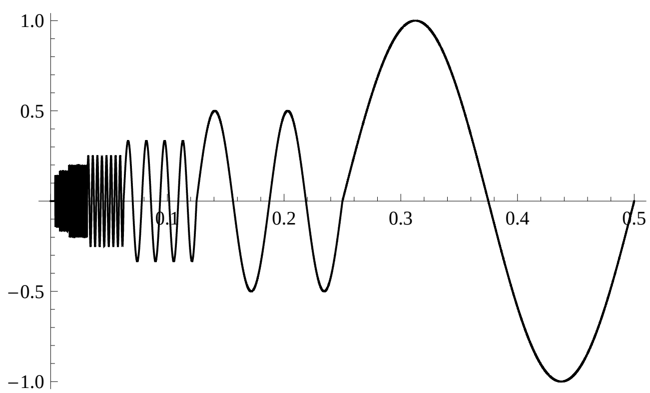


Fig. 7.1 Graph of the butterfly function.

7.5 The Fourier Transform on $L^2(\mathbb{R})$

Equation (7.8) defines the Fourier transform of an integrable function. A function in $L^2(\mathbb{R})$ need not be integrable, so equation (7.8) does not allow us to define the Fourier transform of functions in $L^2(\mathbb{R})$. On the other hand, the Fourier transform is *densely defined* on $L^2(\mathbb{R})$ in the sense of Notation 6.5.5. For example, every function $f \in C_c^2(\mathbb{R})$ is integrable and $C_c^2(\mathbb{R})$ is dense in $L^2(\mathbb{R})$, so $\widehat{f}$ is defined for all f in a dense subspace of $L^2(\mathbb{R})$. The following result shows that the Fourier transform maps this dense subspace into $L^2(\mathbb{R})$, and in fact it does so isometrically with respect to the L^2-norm. This exercise uses facts about convolution that we derived in Section 6.8.5.

Lemma 7.5.1. *If $f \in C_c^2(\mathbb{R})$, then $\widehat{f} \in L^2(\mathbb{R})$. Furthermore,*

$$\|\widehat{f}\|_2 \;=\; \|f\|_2, \qquad f \in C_c^2(\mathbb{R}). \tag{7.13}$$

Proof. If $f \in C_c^2(\mathbb{R})$ then f is integrable, so $\widehat{f}$ is bounded; in fact, $\widehat{f} \in C_0(\mathbb{R})$ by the Riemann–Lebesgue Lemma. Additionally, $\widehat{f}$ satisfies the decay condition that appears in equation (7.11), so $\widehat{f} \in L^2(\mathbb{R})$.

To prove that equation (7.13) holds, we consider the *involution* of f, which is the function $\widetilde{f}(x) = \overline{f(-x)}$. This is an integrable function, and by making a simple change of variables we compute that

$$\big(\widetilde{f}\,\big)^{\wedge}(\xi) \;=\; \overline{\widehat{f}(\xi)}.$$

We also need the *autocorrelation*

$$g \;=\; f * \widetilde{f},$$

the convolution of f with $\widetilde{f}$. The convolution of integrable functions is integrable by Exercise 6.8.11, so $g \in L^1(\mathbb{R})$ (additionally, g is continuous since f and $\widetilde{f}$ are both continuous). Since the Fourier transform converts convolution to multiplication, it follows that

$$\widehat{g}(\xi) \;=\; \big(f * \widetilde{f}\,\big)^{\wedge}(\xi) \;=\; \widehat{f}(\xi)\,\overline{\widehat{f}(\xi)} \;=\; |\widehat{f}(\xi)|^2 \;\in\; L^1(\mathbb{R}).$$

Thus g and $\widehat{g}$ are both integrable, so the Inversion Formula implies that $g(x) = \big(\widehat{g}\,\big)^{\vee}(x)$ for every x, where

$$\overset{\vee}{h}(\gamma) \;=\; \widehat{h}(-\gamma) \;=\; \int_{-\infty}^{\infty} h(t)\, e^{2\pi i \gamma t}\, dt, \qquad \gamma \in \mathbb{R},$$

denotes the inverse Fourier transform of an integrable function h. Evaluating g at $x = 0$, we see that

$$g(0) \;=\; \big(\widehat{g}\,\big)^{\vee}(0) \;=\; \int_{-\infty}^{\infty} \widehat{g}(\xi)\, d\xi \;=\; \int_{-\infty}^{\infty} |\widehat{f}(\xi)|^2\, d\xi \;=\; \|\widehat{f}\,\|_2^2.$$

On the other hand, by the definition of convolution we also have

$$g(0) \;=\; \big(f * \widetilde{f}\,\big)(0) \;=\; \int_{-\infty}^{\infty} f(y)\, \widetilde{f}(0-y)\, dy \;=\; \int_{-\infty}^{\infty} f(y)\, \overline{f(y)}\, dy \;=\; \|f\|_2^2.$$

Consequently $\|\widehat{f}\,\|_2 = \|f\|_2$. □

Thus the Fourier transform maps the dense subspace $C_c^2(\mathbb{R})$ of $L^2(\mathbb{R})$ isometrically into (but not onto) $L^2(\mathbb{R})$. In the language of Notation 6.5.5, $\mathcal{F}$ is a bounded, densely defined mapping of $L^2(\mathbb{R})$ into itself. Exercise 6.5.7 applies to precisely this type of situation, and it tells us that there is an extension of the Fourier transform to a bounded operator that is defined on all of $L^2(\mathbb{R})$.

Theorem 7.5.2. (a) *There is a unique bounded operator $\mathcal{F}\colon L^2(\mathbb{R}) \to L^2(\mathbb{R})$ that extends the Fourier transform in the sense that*

$$\mathcal{F}(f) \;=\; \widehat{f} \quad \textit{for all } f \in C_c^2(\mathbb{R}).$$

(b) *The operator $\mathcal{F}$ satisfies $\mathcal{F}(f) = \widehat{f}$ for all $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$.*

(c) *$\mathcal{F}$ is an isometry on $L^2(\mathbb{R})$, and therefore the following* Plancherel Equality *holds:*

$$\|f\|_2 \;=\; \|\mathcal{F}(f)\|_2, \qquad f \in L^2(\mathbb{R}). \tag{7.14}$$

(d) *$\mathcal{F}$ is a unitary mapping of $L^2(\mathbb{R})$ onto itself (i.e., $\mathcal{F}$ is an isometric isomorphism that maps the Hilbert space $L^2(\mathbb{R})$ onto itself).*

Proof. (a) By applying Exercise 6.5.7 to Lemma 7.5.1, we see that there is a unique operator $\mathcal{F}\colon L^2(\mathbb{R}) \to L^2(\mathbb{R})$ that coincides with the Fourier transform on $C_c^2(\mathbb{R})$.

(b) For clarity in this part of the proof, we will let $\mathcal{F}_1\colon L^1(\mathbb{R}) \to C_0(\mathbb{R})$ be the Fourier transform operator defined in Section 7.4, and we will let $\mathcal{F}_2\colon L^2(\mathbb{R}) \to L^2(\mathbb{R})$ be the operator created in part (a). Each of these operators is continuous on their respective domains, and by construction they coincide on $C_c^2(\mathbb{R})$. Now we want to prove that $\mathcal{F}_1$ and $\mathcal{F}_2$ coincide on the largest domain that they have in common, which is $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$.

Choose any function $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Since $C_c^2(\mathbb{R})$ is dense in both of these spaces, there exists a sequence $\{f_n\}_{n\in\mathbb{N}}$ contained in $C_c^2(\mathbb{R})$ such that $f_n \to f$ in both L^1-norm and L^2-norm. As $\mathcal{F}_1$ is continuous on $L^1(\mathbb{R})$, we have $\mathcal{F}_1(f_n) \to \mathcal{F}_1(f)$, and similarly $\mathcal{F}_2(f_n) \to \mathcal{F}_2(f)$ since $\mathcal{F}_2$ is continuous on $L^2(\mathbb{R})$. However, $\mathcal{F}_1(f_n) = \mathcal{F}_2(f_n)$ for every n, so $\mathcal{F}_1(f) = \mathcal{F}_2(f)$.

Thus $\mathcal{F}_1$ and $\mathcal{F}_2$ coincide on $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, and from now on we just use the symbol $\mathcal{F}$ to denote both $\mathcal{F}_1$ and $\mathcal{F}_2$.

(c) This follows from Problem 6.5.12, or can be done directly.

(d) Since $\mathcal{F}\colon L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is an isometry, its range is a closed subspace of $L^2(\mathbb{R})$. Consequently, if we show that the range is dense in $L^2(\mathbb{R})$ then it must be true that $\mathrm{range}(\mathcal{F}) = L^2(\mathbb{R})$.

To do this, fix any function $f \in C_c^2(\mathbb{R})$. Then $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, and it follows from equation (7.8) that $\widehat{f} \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Consequently

$$\check{f}(\xi) \;=\; \widehat{f}(-\xi) \;\in\; L^1(\mathbb{R}) \cap L^2(\mathbb{R}),$$

and the Inversion Formula implies that

$$\mathcal{F}(\check{f}) \;=\; \big(\widehat{f}\big)^{\vee} \;=\; f.$$

Hence f belongs to the range of $\mathcal{F}$. We have shown that $C_c^2(\mathbb{R}) \subseteq \mathrm{range}(\mathcal{F})$, so $\mathrm{range}(\mathcal{F})$ is both dense and closed in $L^2(\mathbb{R})$. Therefore $\mathrm{range}(\mathcal{F}) = L^2(\mathbb{R})$, so $\mathcal{F}$ is a surjective, isometric map of the Hilbert space $L^2(\mathbb{R})$ onto itself. An isometric isomorphism that maps a Hilbert spaces onto another Hilbert space is called a *unitary mapping*. $\square$

In summary, the Fourier transform is a bounded but not surjective mapping of $L^1(\mathbb{R})$ into $C_0(\mathbb{R})$, and it is a unitary map of $L^2(\mathbb{R})$ onto itself. If f is integrable, then its Fourier transform $\mathcal{F}(f) = \widehat{f}$ is defined by the integral

$$\widehat{f}(\xi) \;=\; \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i \xi x}. \tag{7.15}$$

The Fourier transform is defined on $L^2(\mathbb{R})$, and if $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ then $\widehat{f}$ is defined by equation (7.15), but if $f \in L^2(\mathbb{R}) \backslash L^1(\mathbb{R})$ then f is not integrable and its Fourier transform is not given by equation (7.15). Instead, we must follow the procedure given in Exercise 6.5.7. Specifically, if $\{f_n\}_{n\in\mathbb{N}}$ is any sequence in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ that converges to f in both L^1-norm and L^2-norm, then $\{\widehat{f_n}\}_{n\in\mathbb{N}}$ is a convergent sequence in $L^2(\mathbb{R})$, and we *define* $\mathcal{F}(f)$ to be the limit of that sequence. We call $\mathcal{F}(f)$ the *Fourier transform* of f, and we usually denote using the same symbols that we use for the Fourier transform of an integrable function, i.e., we simply write $\widehat{f}$ instead of $\mathcal{F}(f)$. Using this notation, the Plancherel Equality stated in equation (7.14) can be rewritten as

$$\int_{-\infty}^{\infty} |f(x)|^2\, dx \;=\; \int_{-\infty}^{\infty} |\widehat{f}(\xi)|^2\, d\xi, \qquad f \in L^2(\mathbb{R}).$$

Since $\mathcal{F}\colon L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is unitary, it has a unitary inverse operator $\mathcal{F}^{-1}\colon L^2(\mathbb{R}) \to L^2(\mathbb{R})$. If $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ then $\mathcal{F}^{-1}(f)$ is precisely $\check{f}$, the inverse Fourier transform defined in equation (7.12).For this reason, given any function $f \in L^2(\mathbb{R})$ we usually denote $\mathcal{F}^{-1}(f)$ simply by $\check{f}$.

Example 7.5.3. Let

$$s(x) \;=\; \operatorname{sinc}(x) \;=\; \frac{\sin \pi x}{\pi x}$$

denote the sinc function. This function is not integrable, so its Fourier transform is not given by equation (7.15). On the other hand, the sinc function does belong to $L^2(\mathbb{R})$, so Theorem 7.5.2 tells us that there is a well-defined Fourier transform $\widehat{s}$ of s, and $\widehat{s} \in L^2(\mathbb{R})$. Moreover, because Theorem 7.5.2 proves that the Fourier transform is a unitary mapping of $L^2(\mathbb{R})$ onto itself, we can even compute $\widehat{s}$ without having to construct a sequence of functions in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ that converge to s. This is because we saw in equation (7.10) that the Fourier transform of the box function $\chi = \chi_{[-\frac{1}{2},\frac{1}{2}]}$ is

$$\widehat{\chi} \;=\; s.$$

That is, $\mathcal{F}(\chi) = s$. As $\chi \in L^2(\mathbb{R})$ and $\mathcal{F}$ is unitary, we immediately infer that $\chi = \mathcal{F}^{-1}(s)$. Moreover, s is even so its Fourier transform and inverse Fourier transform coincide. Consequently,

$$\widehat{s} \;=\; \mathcal{F}(s) \;=\; \mathcal{F}^{-1}(s) \;=\; \chi,$$

i.e., the Fourier transform of the sinc function is the box function. Even so, we cannot compute $\widehat{s}$ by making use of the integral formula given in equation (7.15). $\diamondsuit$

It is possible to extend the Fourier transform beyond $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$. The process of *interpolation* allows us to define the Fourier transform of any function in $L^p(\mathbb{R})$ for indices in the range $1 \le p \le 2$. We can even go further and define the Fourier transform of every *tempered distribution*.

Problems

7.5.4. (a) Show that if $f \in L^1(\mathbb{R})$ and $\widehat{f} \in L^2(\mathbb{R})$, then $f \in L^2(\mathbb{R})$.

(b) Use part (a) to show that the Plancherel Equality holds for functions in $L^1(\mathbb{R})$, i.e., if $f \in L^1(\mathbb{R})$ then

$$\int_{-\infty}^{\infty} |f(x)|^2 \, dx \;=\; \int_{-\infty}^{\infty} |\widehat{f}(\xi)|^2 \, d\xi,$$

in the sense that one side is finite if and only if the other side is finite and in this case they are equal, otherwise both are infinite.

(c) Exhibit a function $f \in L^1(\mathbb{R}) \setminus L^2(\mathbb{R})$ such that $\widehat{f} \notin L^1(\mathbb{R})$.

7.6 The Fourier Transform as an Integral Operator

If we restate equation (7.8) as

$$\widehat{f}(\xi) \;=\; L_k f(\xi) \;=\; \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i \xi x}\, dx, \qquad f \in L^1(\mathbb{R}), \tag{7.16}$$

we see that the Fourier transform on $L^1(\mathbb{R})$ is the integral operator whose kernel is $k(x,\xi) = e^{-2\pi i \xi x}$. Unfortunately, this kernel k does not fit any of the boundedness results we proved in this section, since it is not square-integrable on $\mathbb{R}^2$, and it does not satisfy the hypotheses of Schur's Test. This illustrates some of the limitations of our results for integral operators.

On the other hand we do have $k \in L^\infty(\mathbb{R}^2)$, and we can immediately deduce from this that the Fourier transform is a bounded mapping of $L^1(\mathbb{R})$ into $L^\infty(\mathbb{R})$. However, it takes more work to prove that the range $A(\mathbb{R})$ is contained in $C_0(\mathbb{R})$; this is the Riemann–Lebesgue Lemma. Further, although we showed in Section 7.5 that we can extend the Fourier transform to a unitary operator on $L^2(\mathbb{R})$, it is not an integral operator on $L^2(\mathbb{R})$ since equation (7.16) only defines the Fourier transform of functions in $L^1(\mathbb{R})$.

Problems

7.6.1. Given $g \in L^1(\mathbb{R})$, define $A_g \colon L^1(\mathbb{R}) \to L^1(\mathbb{R})$ by $A_g(f) = f * g$. Show that the operator norm of A_g is $\|A_g\| = \|g\|_1$.

Chapter 8
Operators on Hilbert Spaces

In this chapter we will study operators that map one Hilbert space into another. The fact that we now have an inner product to work with allows us to derive much more detailed results than we could for operators on generic normed spaces.

Throughout this chapter the letters H and K will usually be Hilbert spaces, but some results only require them to be inner product spaces (we state the precise hypotheses in each theorem). Each of H and K has its own inner product, but typically we just write $\langle \cdot, \cdot \rangle$ for the inner product and let context determine whether this is the inner product on H or on K. For example, if x and y both belong to H, then we know that $\langle x, y \rangle$ must be the inner product from H, while if they belong to K then $\langle x, y \rangle$ has to denote the inner product on K. If there is a chance of confusion, then we may write $\langle x, y \rangle_H$ and $\langle x, y \rangle_K$.

8.1 Unitary Operators

Recall from Definition 6.6.1 that an *isometry* is an operator A that preserves the norms of vectors (i.e., $\|Ax\| = \|x\|$ for every x). However, if H and K are inner product spaces, then in addition to considering whether a mapping $A\colon H \to K$ preserve norms, we can also ask whether it will preserve inner products. We will say that A is *inner-product preserving* if

$$\langle Ax, Ay \rangle \;=\; \langle x, y \rangle, \qquad \text{all } x, y \in H.$$

Taking $x = y$, we see that any operator that preserves inner products must preserve norms as well, and therefore is an isometry. The next lemma proves that the converse holds for linear operators on inner product spaces.

Lemma 8.1.1. *Let H and K be inner product spaces. If $A\colon H \to K$ is a linear operator, then A is an isometry if and only if A is inner-product preserving.*

Proof. $\Leftarrow$. If A preserves inner products, then $\|Ax\|^2 = \langle Ax, Ax\rangle = \langle x, x\rangle = \|x\|^2$, so A preserves norms as well.

$\Rightarrow$. Assume that A is an isometry. Given $x,\, y \in H$, we use the Polar Identity and the fact that A preserves norms to compute that

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

Thus $\mathrm{Re}\langle Ax,Ay\rangle = \mathrm{Re}\langle x,y\rangle$. If $\mathbb{F}=\mathbb{R}$ then we are done. If $\mathbb{F}=\mathbb{C}$, then a similar calculation based on expanding $\|x+iy\|^2$ shows that $\mathrm{Im}\langle Ax,Ay\rangle = \mathrm{Im}\langle x,y\rangle$. $\square$

Recall from Definition 6.6.5 that an *isometric isomorphism* is a linear isometry that is a bijection (in fact, every linear isometry is automatically injective, so a linear isometry is a bijection if and only if it is surjective). We give the following name to isometric isomorphisms whose domain and codomain are *Hilbert spaces.*

Definition 8.1.2 (Unitary Operator). Let H and K be Hilbert spaces.

(a) If $A\colon H \to K$ is an isometric isomorphism, then we say that A is a *unitary operator*.

(b) If there exists a unitary operator $A\colon H \to K$, then we say that H and K are *unitarily isomorphic*, and in this case we write $H \cong K$. $\diamondsuit$

Here is an example of a unitary operator.

Example 8.1.3 (Analysis Operator). Let H be any separable, infinite-dimensional Hilbert space. Then H has an orthonormal basis, say $\{e_n\}_{n\in\mathbb{N}}$. If $x \in H$, then we know from the Plancherel Equality that

$$\sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2 \;=\; \|x\|^2. \tag{8.1}$$

Therefore the sequence $\{\langle x,e_n\rangle\}_{n\in\mathbb{N}}$ belongs to the space ℓ^2, and so we can define an operator $T\colon H \to \ell^2$ by setting

$$Tx \;=\; \bigl\{\langle x, e_n\rangle\bigr\}_{n\in\mathbb{N}}.$$

That is, T maps the vector x to the sequence of inner products of x with the basis elements e_n. This operator T is called the *analysis operator* associated with the orthonormal basis $\{e_n\}_{n\in\mathbb{N}}$.

The fact that the inner product is linear in the first variable implies that T is linear. By equation (8.1), the ℓ^2-norm of the sequence Tx equals the norm of the vector x:

$$\|Tx\|_2 \;=\; \Big\|\big\{\langle x, e_n\rangle\big\}_{n\in\mathbb{N}}\Big\|_2 \;=\; \left(\sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2\right)^{1/2} \;=\; \|x\|.$$

Hence T is an isometry. Finally, if $c = (c_n)_{n\in\mathbb{N}}$ is any sequence in ℓ^2 then we know from Theorem 5.6.3 that the series $x = \sum c_n e_n$ converges, and that theorem also implies that $c_n = \langle x, e_n\rangle$ for every n. Therefore $Tx = c$, so T is surjective. Thus T is a surjective isometric isomorphism, so it is a unitary operator that maps H onto ℓ^2. $\diamondsuit$

Next we will construct a particular family of bounded operators, and determine which operators in the family are unitary. As in Example 8.1.3, we begin with an orthonormal basis $\{e_n\}_{n\in\mathbb{N}}$ for a separable, infinite-dimensional Hilbert space H. Each $x \in H$ can be uniquely written as

$$x \;=\; \sum_{n=1}^{\infty} \langle x, e_n\rangle\, e_n, \tag{8.2}$$

and our operator will create a new vector from this by modifying the coefficients in this series.

Let $\lambda = (\lambda_n)_{n\in\mathbb{N}}$ be any sequence of scalars. The idea behind our construction is to modify the series in equation (8.2) by rescaling each term in the series by a factor of λ_n. That is, we would like to create a new vector, which we will call $M_\lambda x$, by defining

$$M_\lambda x \;=\; \sum_{n=1}^{\infty} \lambda_n \langle x, e_n\rangle\, e_n. \tag{8.3}$$

However, in order for this definition to make sense we have to ensure that the series on the line above converges. The following theorem shows that if the sequence λ is bounded, then the series in equation (8.3) converges for every x and M_λ defines a bounded mapping of H into H. Furthermore, if each λ_n has unit modulus then M_λ is unitary (for the case $\mathbb{F} = \mathbb{R}$, this amounts to requiring that λ be a sequence of ± 1's).

Theorem 8.1.4. *Let H be a separable, infinite-dimensional Hilbert space, let $\{e_n\}_{n\in\mathbb{N}}$ be an orthonormal basis for H, and let $\lambda = (\lambda_n)_{n\in\mathbb{N}}$ be a sequence of scalars.*

(a) *If λ is a bounded sequence, then the series defining $M_\lambda x$ in equation* (8.3) *converges for every $x \in H$, and the operator $M_\lambda \colon H \to H$ defined in this way is bounded.*

(b) *If $|\lambda_n| = 1$ for every n, then $M_\lambda \colon H \to H$ is a unitary operator.*

Proof. (a) Assume that λ is a bounded sequence, i.e., $\lambda \in \ell^\infty$. By definition of the ℓ^∞-norm, we have $|\lambda_n| \le \|\lambda\|_\infty$ for every n. Therefore, given any $x \in H$,

$$\begin{aligned}
\sum_{n=1}^{\infty} |\lambda_n \langle x, e_n\rangle|^2 &\le \sum_{n=1}^{\infty} \|\lambda\|_\infty^2\, |\langle x, e_n\rangle|^2 && (\text{since } |\lambda_n| \le \|\lambda\|_\infty)\\
&= \|\lambda\|_\infty^2 \sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2\\
&= \|\lambda\|_\infty^2\, \|x\|^2 && (\text{Plancherel's Equality})\\
&< \infty.
\end{aligned}$$

This implies, by Theorem 5.6.3, that the series $\sum \lambda_n \langle x, e_n\rangle\, e_n$ converges, and so we can define $M_\lambda x$ by equation (8.3). Moreover, it follows from the fact that the inner product is linear in the first variable that M_λ is a linear operator.

Applying the Plancherel Equality again, we see that

$$\|M_\lambda x\|^2 \;=\; \sum_{n=1}^{\infty} |\lambda_n \langle x, e_n\rangle|^2 \;\le\; \|\lambda\|_\infty^2\, \|x\|^2.$$

Therefore M_λ is a bounded operator. In fact, by taking the supremum of $\|M_\lambda x\|$ over the unit vectors we see that the operator norm of M_λ satisfies

$$\|M_\lambda\| \;=\; \sup_{\|x\|=1} \|M_\lambda x\| \;\le\; \|\lambda\|_\infty.$$

The reader should verify (this is part of Problem 8.1.9) that we actually have $\|M_\lambda\| = \|\lambda\|_\infty$.

(b) If $|\lambda_n| = 1$ for every n then λ is a bounded sequence, so M_λ is a bounded operator by part (a). Additionally, by the Plancherel Equality,

$$\|M_\lambda x\|^2 \;=\; \sum_{n=1}^{\infty} |\lambda_n \langle x, e_n\rangle|^2 \;=\; \sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2 \;=\; \|x\|^2,$$

so M_λ is isometric. To show that M_λ is surjective, fix any $x \in H$ and define

$$y \;=\; \sum_{n=1}^{\infty} \overline{\lambda_n}\, \langle x, e_n\rangle\, e_n$$

(why does this series converge?). Then

$$M_\lambda y \;=\; \sum_{n=1}^{\infty} \lambda_n \overline{\lambda}_n \langle x, e_n\rangle\, e_n \;=\; \sum_{n=1}^{\infty} |\lambda_n|^2 \langle x, e_n\rangle\, e_n \;=\; \sum_{n=1}^{\infty} \langle x, e_n\rangle\, e_n \;=\; x.$$

Thus M_λ is surjective, and therefore it is a unitary operator. $\square$

We sometimes refer to M_λ, or operators like it, as a "multiplication operator," or a "multiplier." To give a concrete example, take $H = \ell^2$ and assume that $\{e_n\}_{n\in\mathbb{N}}$ is the standard basis for ℓ^2. Then M_λ is simply the following componentwise multiplication:

$$M_\lambda x \;=\; (\lambda_1 x_1,\, \lambda_2 x_2,\, \dots), \qquad x = (x_1, x_2, \dots) \in \ell^2.$$

Problems 8.1.9 and 8.1.10 develop some further properties of M_λ for various choices of λ.

Problems

8.1.5. Let H and K be inner product spaces, and let $A\colon H \to K$ be a linear operator. Prove that $\|A\| = \sup\bigl\{|\langle Ax, y\rangle| : \|x\| = \|y\| = 1\bigr\}$.

8.1.6. (a) Prove that every separable infinite-dimensional Hilbert space is unitarily isomorphic to ℓ^2, and use this to prove that any two separable infinite-dimensional Hilbert spaces are unitarily isomorphic. That is, prove that if H and K are two separable infinite-dimensional Hilbert spaces, then $H \cong K$.

(b) Are any two finite-dimensional Hilbert spaces unitarily isomorphic? If not, fill in the blank so that the following statement is true, and provide a proof: If H and K are two finite-dimensional Hilbert spaces, then H and K are unitarily isomorphic if and only if ____________.

8.1.7. Let $M \neq \{0\}$ be a closed subspace of a Hilbert space H, and let P be the orthogonal projection of H onto M (see Definition 5.5.8). Prove the following statements.

(a) $\|x - Px\| = \operatorname{dist}(x, M)$ for every $x \in H$.

(b) P is linear, $\|Px\| \le \|x\|$ for every $x \in H$, and $\|P\| = 1$.

(c) $P^2 = P$ (an operator that equals its own square is *idempotent*.)

(d) $\ker(P) = M^\perp$ and $\operatorname{range}(P) = M$.

(e) $I - P$ is the orthogonal projection of H onto $M^\perp$.

8.1.8. (a) Let X and Y be normed spaces. Given $A_n,\, A \in \mathcal{B}(X, Y)$, we say that A_n *converges to A in the strong operator topology* if $A_n x \to Ax$ for every

$x \in X$. Show that if $\|A - A_n\| \to 0$, then A_n converges to A in the strong operator topology.

(b) Let $\{e_n\}_{n\in\mathbb{N}}$ be an orthonormal basis for a Hilbert space H, and let P_N be the orthogonal projection of H onto $\operatorname{span}\{e_1, \dots, e_N\}$. Prove that $P_N \to I$ in the strong operator topology, but $\|I - P_N\| \not\to 0$ as $N \to \infty$.

8.1.9. Let $\lambda = (\lambda_n)_{n\in\mathbb{N}}$ be a bounded sequence of scalars, and let M_λ be the bounded operator discussed in Theorem 8.1.4. Set $\delta = \inf |\lambda_n|$, and prove the following statements.

(a) $\|M_\lambda\| = \|\lambda\|_\infty$.

(b) M_λ is injective if and only if $\lambda_n \neq 0$ for every n.

(c) M_λ is surjective if and only if $\delta > 0$.

(d) If $\delta = 0$ but $\lambda_n \neq 0$ for every n, then $\operatorname{range}(M_\lambda)$ is a proper, dense subspace of H.

(e) If $\delta > 0$ then M_λ is a topological isomorphism, but M_λ is unitary if and only if $|\lambda_n| = 1$ for every n.

8.1.10. Assume that $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for a Hilbert space H, and let $\lambda = (\lambda_n)_{n\in\mathbb{N}}$ be any sequence of scalars. For those $x \in H$ *for which the following series converges*, define

$$M_\lambda x \;=\; \sum_{n=1}^{\infty} \lambda_n \langle x, e_n\rangle\, e_n. \tag{8.4}$$

Prove the following statements.

(a) The series in equation (8.4) converges if and only if x belongs to

$$\operatorname{domain}(M_\lambda) \;=\; \biggl\{x \in H : \sum_{n=1}^{\infty} |\lambda_n \langle x, e_n\rangle|^2 < \infty\biggr\}.$$

Hint: If $\lambda = (\lambda_n)_{n\in\mathbb{N}} \notin \ell^\infty$, then there exists a subsequence $(\lambda_{n_k})_{k\in\mathbb{N}}$ such that $|\lambda_{n_k}| \geq k$ for each k. Let $c_{n_k} = 1/k$ and define all other c_n to be zero.

(b) If $\lambda \notin \ell^\infty$, then $\operatorname{domain}(M_\lambda)$ is a dense but proper subspace of H, and $M_\lambda \colon \operatorname{domain}(M_\lambda) \to H$ is an unbounded linear operator.

(c) $\operatorname{domain}(M_\lambda) = H$ if and only if $\lambda \in \ell^\infty$.

8.2 Adjoints of Operators on Hilbert Spaces

Let A be an $m \times n$ matrix with scalar entries, which we write as

$$A = \begin{bmatrix} a_{ij} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ a_{31} & a_{32} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}. \tag{8.5}$$

The *transpose* of A is the $n \times m$ matrix A^{T} obtained by interchanging rows with columns in A:

$$A^{\mathrm{T}} = \begin{bmatrix} a_{ji} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} & a_{31} & \cdots & a_{m1} \\ a_{12} & a_{22} & a_{31} & \cdots & a_{m2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & a_{31} & \cdots & a_{mn} \end{bmatrix}.$$

The *Hermitian*, or *conjugate transpose*, of A is the $n \times m$ matrix A^{H} obtained by taking the complex conjugate of each entry of the transpose of A:

$$A^{\mathrm{H}} = \overline{A^{\mathrm{T}}} = \begin{bmatrix} \overline{a_{ji}} \end{bmatrix} = \begin{bmatrix} \overline{a_{11}} & \overline{a_{21}} & \overline{a_{31}} & \cdots & \overline{a_{m1}} \\ \overline{a_{12}} & \overline{a_{22}} & \overline{a_{31}} & \cdots & \overline{a_{m2}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \overline{a_{1n}} & \overline{a_{2n}} & \overline{a_{31}} & \cdots & \overline{a_{mn}} \end{bmatrix}. \tag{8.6}$$

Note that if $\mathbb{F} = \mathbb{R}$ then $A^{\mathrm{H}} = A^{\mathrm{T}}$.

Letting $x \cdot y$ denote the dot product of vectors, we have the following fundamental equality that relates A to A^{H}. We assign the proof of this result as Problem 8.2.9.

Theorem 8.2.1. *If A is an $m \times n$ matrix, then*

$$Ax \cdot y = x \cdot A^{H} y, \qquad x \in \mathbb{F}^n, \ y \in \mathbb{F}^m. \qquad \diamondsuit \tag{8.7}$$

We will use the Riesz Representation Theorem to prove that if $A \colon H \to K$ is any bounded linear operator between Hilbert spaces H and K, then there is a bounded linear operator $A^* \colon K \to H$, called the *adjoint* of A, that is related to A in a manner similar to the relationship between the matrices A and A^{H} that appears in equation (8.7).

Theorem 8.2.2 (Adjoints). *Let H and K be Hilbert spaces, and assume that $A \colon H \to K$ is a bounded linear operator. Then there exists a bounded linear operator $A^* \colon K \to H$ such that*

$$\langle Ax, y \rangle = \langle x, A^* y \rangle, \qquad x \in H, \ y \in K. \tag{8.8}$$

Moreover, A^* is the unique *operator from K to H that satisfies equation* (8.8).

Proof. Before beginning the proof, note that Ax and y both belong to K, and therefore the notation $\langle Ax, y\rangle$ in equation (8.2.2) implicitly denotes the inner product of vectors in K. Similarly, since x and A^*x belong to H, $\langle x, A^*y\rangle$ implicitly means the inner product of vectors in H.

Given a vector $y \in K$, define a functional $\mu_y \colon H \to \mathbb{F}$ by

$$\mu_y(x) \;=\; \langle Ax, y\rangle, \qquad x \in H.$$

Then μ_y is linear because A is linear and the inner product is linear in the first variable. Applying the Cauchy–Bunyakowski–Schwarz Inequality and recalling that $\|Ax\| \le \|A\|\,\|x\|$ for every x, we compute that

$$|\mu_y(x)| \;=\; |\langle Ax, y\rangle| \;\le\; \|Ax\|\,\|y\| \;\le\; \|A\|\,\|x\|\,\|y\|.$$

Therefore μ_y is bounded. Indeed, by taking the supremum over the unit vectors in H we see that the operator norm of μ_y satisfies

$$\|\mu_y\| \;=\; \sup_{\|x\|=1} |\mu_y(x)| \;\le\; \|A\|\,\|y\| \;<\; \infty.$$

Hence $\mu_y \in H^*$. The Riesz Representation Theorem (Theorem 6.9.2) therefore implies that there exists a unique vector $y^* \in H$ such that

$$\mu_y(x) \;=\; \langle x, y^*\rangle, \qquad x \in H.$$

Let

$$A^*y \;=\; y^*.$$

Since each $y \in K$ determines a unique $y^* \in H$, we have defined a mapping $A^* \colon K \to H$. By definition, if $x \in H$ and $y \in K$, then

$$\langle Ax, y\rangle \;=\; \mu_y(x) \;=\; \langle x, y^*\rangle \;=\; \langle x, A^*y\rangle.$$

Hence equation (8.8) holds. We must show that A^* is linear, bounded, and unique.

Choose any vectors $y, z \in K$ and scalars $a, b \in \mathbb{F}$. Then for every $x \in H$ we have

$$\begin{aligned}
\langle x, A^*(ay+bz)\rangle \;&=\; \langle Ax, ay+bz\rangle && \text{(definition of } A^*\text{)}\\
&=\; \overline{a}\,\langle Ax, y\rangle + \overline{b}\,\langle Ax, z\rangle && \text{(antilinearity in 2nd variable)}\\
&=\; \overline{a}\,\langle x, A^*y\rangle + \overline{b}\,\langle x, A^*z\rangle && \text{(definition of } A^*\text{)}\\
&=\; \langle x, aA^*y + bA^*z\rangle && \text{(antilinearity in 2nd variable).}
\end{aligned}$$

Since this is true for every $x \in H$, it follows that

$$A^*(ay+bz) \;=\; aA^*y + bA^*z.$$

Therefore A^* is linear.

Now we show that A^* is bounded. Given $y \in K$, we compute that

$$\begin{aligned}
\|A^*y\| &= \sup_{\|x\|=1} |\langle x, A^*y\rangle| \\
&= \sup_{\|x\|=1} |\langle Ax, y\rangle| && \text{(by definition of } A^*) \\
&\le \sup_{\|x\|=1} \|Ax\|\,\|y\| && \text{(CBS Inequality)} \\
&= \|A\|\,\|y\| && \text{(definition of operator norm).}
\end{aligned}$$

Hence A^* is bounded. In fact, by taking the supremum over the unit vectors $y \in K$ we see that

$$\|A^*\| = \sup_{\|y\|=1} \|A^*y\| \le \sup_{\|y\|=1} \|A\|\,\|y\| = \|A\|. \tag{8.9}$$

It only remains to show that A^* is unique. Suppose that $B\colon K \to H$ also satisfies $\langle Ax, y\rangle = \langle x, By\rangle$ for all $x \in H$ and $y \in K$. Then

$$\langle x, A^*y\rangle = \langle Ax, y\rangle = \langle x, By\rangle$$

for all $x \in H$ and $y \in K$. Holding y fixed, it follows that $\langle x, A^*y - By\rangle = 0$ for every x. This implies that $A^*y - By = 0$, and therefore $A^*y = By$. This is true for every y, so A^* and B are the same operator. $\square$

We introduce a name for the operator A^*.

Definition 8.2.3 (Adjoint Operator). Given Hilbert spaces H and K, the *adjoint* of $A \in \mathcal{B}(H,K)$ is the unique operator $A^* \in \mathcal{B}(K,H)$ that satisfies equation (8.8). $\diamondsuit$

We will give some examples, beginning with the operator corresponding to an $m \times n$ matrix.

Example 8.2.4. In this example our Hilbert spaces are the Euclidean spaces $\mathbb{F}^m$ and $\mathbb{F}^n$, whose inner product is the dot product of vectors.

Let A be the $m \times n$ matrix given in equation (8.5). If $x \in \mathbb{F}^n$, then Ax is a vector in $\mathbb{F}^m$. Thus A determines an operator, which we will also call A, that sends $x \in \mathbb{F}^n$ to $Ax \in \mathbb{F}^n$. Likewise the matrix A^{H} defined in equation (8.6) determines an operator that maps $y \in \mathbb{F}^m$ to $A^{\mathrm{H}}y \in \mathbb{F}^n$.

By Theorem 8.2.2, there is *unique* operator $A^*\colon \mathbb{F}^m \to \mathbb{F}^n$ that satisfies $Ax \cdot y = x \cdot A^*y$ for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. Yet by Theorem 8.2.1 we have $Ax \cdot y = x \cdot A^{\mathrm{H}}y$ for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. Hence A^{H} and A^* are the same operator, and therefore are determined by the same matrix, the Hermitian of A given in equation (8.6).

When dealing with matrices, we usually call $A^* = A^{\mathrm{H}}$ the Hermitian of A instead of the adjoint of A. Note that if $\mathbb{F} = \mathbb{R}$ then $A^* = A^{\mathrm{H}} = A^{\mathrm{T}}$ is simply the transpose of A. $\diamondsuit$

Next we compute the adjoint of the operator M_λ that was introduced in Theorem 8.1.4.

Example 8.2.5. Assume that $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for a separable Hilbert space H, and $\lambda = (\lambda_n)_{n\in\mathbb{N}}$ is a bounded sequence of scalars. Let $M_\lambda\colon H \to H$ be the multiplication operator defined in Theorem 8.1.4, i.e.,

$$M_\lambda x \;=\; \sum_{n=1}^{\infty} \lambda_n \,\langle x, e_n\rangle\, e_n, \qquad x \in H.$$

Since M_λ is bounded and linear, Theorem 8.2.2 implies that it has an adjoint $M_\lambda^*\colon H \to H$ that is also bounded and linear. Letting $\overline{\lambda} = \bigl(\overline{\lambda_n}\bigr)_{n\in\mathbb{N}}$ denote the sequence of complex conjugates of the λ_n, another bounded linear mapping is

$$M_{\overline{\lambda}} x \;=\; \sum_{n=1}^{\infty} \overline{\lambda_n}\,\langle x, e_n\rangle\, e_n, \qquad x \in H.$$

We will show that $M_\lambda^* = M_{\overline{\lambda}}$. Given vectors $x, y \in H$, we have

$$\begin{aligned}
\langle x, M_\lambda^* y\rangle \;&=\; \langle M_\lambda x, y\rangle && \text{(definition of the adjoint)}\\
&=\; \Biggl\langle \sum_{n=1}^{\infty} \lambda_n \,\langle x, e_n\rangle\, e_n,\; y \Biggr\rangle && \text{(definition of } M_\lambda)\\
&=\; \sum_{n=1}^{\infty} \bigl\langle \lambda_n \,\langle x, e_n\rangle\, e_n,\; y \bigr\rangle && \text{(continuity of the inner product)}\\
&=\; \sum_{n=1}^{\infty} \lambda_n \,\langle x, e_n\rangle\, \langle e_n, y\rangle && \text{(linearity in the first variable)}\\
&=\; \sum_{n=1}^{\infty} \bigl\langle x,\; \overline{\lambda_n}\,\overline{\langle e_n, y\rangle}\, e_n \bigr\rangle && \text{(antilinearity in the second variable)}\\
&=\; \sum_{n=1}^{\infty} \bigl\langle x,\; \overline{\lambda_n}\,\langle y, e_n\rangle\, e_n \bigr\rangle && \text{(conjugate symmetry)}\\
&=\; \Biggl\langle x,\; \sum_{n=1}^{\infty} \overline{\lambda_n}\,\langle y, e_n\rangle\, e_n \Biggr\rangle && \text{(continuity of the inner product)}\\
&=\; \langle x, M_{\overline{\lambda}}\, y\rangle.
\end{aligned}$$

That is, $M_\lambda^* y = M_{\overline{\lambda}}\, y$ for every $y \in H$. This shows that M_λ^* equals the multiplication operator $M_{\overline{\lambda}}$. $\diamondsuit$

The next theorem lays out some of the properties of adjoint operators on Hilbert spaces.

Theorem 8.2.6. *Let H, K, and L be Hilbert spaces. Given $A \in \mathcal{B}(H,K)$, the following statements hold.*

(a) *If* $x \in H$ *and* $y \in K$, *then* $\langle A^*y, x\rangle = \langle y, Ax\rangle$.
(b) $(A^*)^* = A$ *and* $(cA)^* = \bar{c}A^*$ *for all scalars* c.
(c) *If* $B \in \mathcal{B}(H,K)$, *then* $(A+B)^* = A^* + B^*$.
(d) *If* $B \in \mathcal{B}(K,L)$, *then* $(BA)^* = A^*B^*$.
(e) $\ker(A) = \mathrm{range}(A^*)^\perp$ *and* $\ker(A^*) = \mathrm{range}(A)^\perp$.
(f) $\ker(A)^\perp = \overline{\mathrm{range}}(A^*)$ *and* $\ker(A^*)^\perp = \overline{\mathrm{range}}(A)$.
(g) A *is injective if and only if* $\mathrm{range}(A^*)$ *is dense in* H.
(h) $\|A\| = \|A^*\| = \|A^*A\|^{1/2} = \|AA^*\|^{1/2}$.

Proof. We will prove some portion of these statements, and assign the proof of the rest as Problem 8.2.12.

(a) If $x \in H$ and $y \in K$ then by conjugate symmetry and the definition of the adjoint we have

$$\langle A^*y, x\rangle \;=\; \overline{\langle x, A^*y\rangle} \;=\; \overline{\langle Ax, y\rangle} \;=\; \langle y, Ax\rangle.$$

(b) If $x \in H$ and $y \in K$ then $\big\langle (A^*)^*x,\, y\big\rangle = \langle x, A^*y\rangle = \langle Ax, y\rangle$. This implies that $(A^*)^* = A$.

(e) Fix any vector $x \in \ker(A)$. If $z \in \mathrm{range}(A^*)$, then $z = A^*y$ for some $y \in K$. Since $Ax = 0$, we therefore have

$$\langle x, z\rangle \;=\; \langle x, A^*y\rangle \;=\; \langle Ax, y\rangle \;=\; 0.$$

This shows that $x \in \mathrm{range}(A^*)^\perp$, and therefore $\ker(A) \subseteq \mathrm{range}(A^*)^\perp$.

Conversely, choose any vector $x \in \mathrm{range}(A^*)^\perp$. Then for every $z \in H$ we have $A^*z \in \mathrm{range}(A^*)$, and therefore

$$\langle Ax, z\rangle \;=\; \langle x, A^*z\rangle \;=\; 0.$$

This implies that $Ax = 0$, so $x \in \ker(A)$ and hence $\mathrm{range}(A^*)^\perp \subseteq \ker(A)$.

(h) We saw in equation (8.9) that $\|A^*\| \le \|A\|$. By applying equation (8.9) to the operator A^*, we see that we also have $\|(A^*)^*\| \le \|A^*\|$. Since part (b) established that $(A^*)^* = A$, it follows that $\|A^*\| = \|A\|$. $\square$

Here are some properties of the adjoint of a topological isomorphism or a unitary mapping.

Theorem 8.2.7. *Let* H *and* K *be Hilbert spaces, and let* $A \in \mathcal{B}(H,K)$ *be given.*

(a) *If* A *is a topological isomorphism, then* A^* *is also a topological isomorphism, and* $(A^{-1})^* = (A^*)^{-1}$.
(b) A *is unitary if and only if* A *is a bijection and* $A^* = A^{-1}$.

Proof. (a) Suppose that A is a topological isomorphism, i.e., A is a linear bijection and both A and A^{-1} are bounded. Problem 6.6.17 therefore implies that there exists a constant $c > 0$ such that

$$\|Ax\| \geq c\,\|x\|, \qquad x \in H. \tag{8.10}$$

Fix any nonzero vector $y \in K$. Then $y = Ax$ for some $x \in H$, so

$$\begin{aligned}
\|y\|^2 \;=\; \|Ax\|^2 &= \langle Ax, Ax\rangle && \text{(definition of the norm)}\\
&= \langle A^*Ax, x\rangle && \text{(definition of } A^*\text{)}\\
&\leq \|A^*Ax\|\,\|x\| && \text{(CBS Inequality)}\\
&\leq \|A^*Ax\|\,\frac{\|Ax\|}{c} && \text{(by equation (8.10))}\\
&= \|A^*y\|\,\frac{\|y\|}{c} && (y = Ax)
\end{aligned}$$

Dividing by $\|y\|$ and rearranging, we conclude that

$$\|A^*y\| \;\geq\; c\,\|y\|, \qquad \text{all nonzero } y \in K. \tag{8.11}$$

This same inequality holds trivially if $y = 0$. Problem 6.2.10 therefore implies that A^* is injective and $\mathrm{range}(A^*)$ is a closed subspace of H. Consequently $\mathrm{range}(A^*) = \overline{\mathrm{range}}(A^*)$. On the other hand, since A is injective, Theorem 8.2.6(g) implies that $\mathrm{range}(A^*)$ is dense in H. This tells us that $\overline{\mathrm{range}}(A^*) = H$. Hence $\mathrm{range}(A^*) = \overline{\mathrm{range}}(A^*) = H$, so A^* is surjective.

The remainder of the proof of this part (showing that A^* and $(A^*)^{-1}$ are both bounded, and $(A^*)^{-1} = (A^{-1})^*$) is assigned as Problem 8.2.15.

(b) We assign this proof as Problem 8.2.15. □

8.2.1 Adjoints of Unbounded Operators

Adjoints can be defined for unbounded operators, although now we must be careful with domains. For example, consider the differentiation operator $Df = f'$. This operator is not defined on all of $L^2(\mathbb{R})$, but instead is densely defined in the sense of Notation 6.5.5. For example, D maps the dense subspace

$$S = \bigl\{f \in C_0^1(\mathbb{R}) : f, f' \in L^2(\mathbb{R})\bigr\}$$

into $L^2(\mathbb{R})$. Although $D\colon S \to L^2(\mathbb{R})$ is unbounded, if we fix f, $g \in S$ then fg is differentiable and $fg \in C_0(\mathbb{R})$, so integration by parts yields

$$\langle Df, g\rangle \;=\; \int_{-\infty}^{\infty} f'(x)\,\overline{g(x)}\,dx \;=\; -\int_{-\infty}^{\infty} f(x)\,\overline{g'(x)}\,dx \;=\; -\langle f, Dg\rangle.$$

Hence $D^* = -D$ as an operator mapping S into $L^2(\mathbb{R})$.

The important point in the preceding example is that if we fix $g \in S$, then $f \mapsto \langle Df, g\rangle$ is a bounded linear functional on S, even though D is an unbounded operator. The following exercise extends this to general operators.

Exercise 8.2.8. Suppose that S is a dense subspace of H, and $A\colon S \to K$ is a linear, but not necessarily bounded, operator. Let

$$S^* \;=\; \bigl\{g \in K : f \mapsto \langle Af, g\rangle \text{ is a bounded linear functional on } S\bigr\}.$$

Show that there is an operator $A^*\colon S^* \to H$ such that

$$\langle Af, g\rangle \;=\; \langle f, A^*g\rangle, \qquad f \in S,\; g \in S^*. \qquad \diamondsuit$$

If necessary, we can always restrict a densely defined operator to a smaller but still dense domain. Given an operator A mapping some dense subspace of H into H, if we can find some dense subspace S on which A is defined and satisfies $\langle Af, g\rangle = \langle f, Ag\rangle$ for f, $g \in S$, then we say that A is *self-adjoint*.

Problems

8.2.9. (a) Let A be an $m \times n$ matrix, as in equation (8.5). Let $\{e_1, \dots, e_n\}$ be the standard basis for $\mathbb{F}^n$, and let $\{f_1, \dots, f_m\}$ be the standard basis for $\mathbb{F}^m$. Prove that $Ae_i \cdot f_j = a_{ij}$ for each $i = 1, \dots, n$ and $j = 1, \dots, m$.

(b) Prove Theorem 8.2.1.

8.2.10. Let L, $R\colon \ell^2 \to \ell^2$ be the left-shift and right-shift operators from Examples 6.6.3 and 6.6.4. Prove that $L^* = R$ and $R^* = L$.

8.2.11. Let $\{e_n\}_{n\in\mathbb{N}}$ be an orthonormal basis for a Hilbert space H, and let $T\colon H \to \ell^2$ be the *analysis operator* defined in Example 8.1.3. Find a formula for the adjoint $T^*\colon \ell^2 \to H$ (which is called the *synthesis operator* for $\{e_n\}_{n\in\mathbb{N}}$).

8.2.12. Prove the remaining implications in Theorem 8.2.6.

8.2.13. Given Hilbert spaces H and K, define $T\colon \mathcal{B}(H,K) \to \mathcal{B}(K,H)$ by $T(A) = A^*$. Prove that T is an antilinear isometric isomorphism (i.e., T is antilinear, isometric, and surjective).

8.2.14. Given $A \in \mathcal{B}(H)$, where H is a Hilbert space, show that

$$\ker(A) \;=\; \ker(A^*A) \qquad \text{and} \qquad \overline{\mathrm{range}}(A^*A) \;=\; \overline{\mathrm{range}}(A^*).$$

8.2.15. Complete the proof of Theorem 8.2.7.

8.2.16. (a) Show that the adjoint defines an involution on $\mathcal{B}(H)$ in the sense of Definition 7.3.6.

(b) Show that $\mathcal{B}(H)$ is a C^*-algebra, in the sense of Definition 7.3.8, with respect to the adjoint operation.

8.2.17. Fix $\lambda \in \ell^\infty$, and let M_λ be the multiplication operator defined in Theorem 8.1.4. Find M_λ^*, and characterize those λ such that M_λ is self-adjoint, positive, or positive definite.

8.2.18. Fix $k \in L^2(\mathbb{R}^2)$. By Theorem 6.8.7, the integral operator L_k with kernel k is a bounded linear map of $L^2(\mathbb{R})$ onto itself. Prove that the adjoint operator $(L_k)^*$ is the integral operator L_{k^*} whose kernel is $k^*(x,y) = \overline{k(y,x)}$. Characterize those kernels $k \in L^2(\mathbb{R}^2)$ that correspond to self-adjoint operators L_k.

8.3 Self-Adjoint Operators

Now we take $H = K$ and focus on operators that map a Hilbert space H into itself. If $A\colon H \to H$ is bounded and linear, then its adjoint $A^*\colon H \to H$ is also bounded and linear. Therefore it is possible that A and A^* might be the same operator. We have a special name for this situation.

Definition 8.3.1 (Self-Adjoint Operator). Let H be a Hilbert space. An operator $A \in \mathcal{B}(H)$ is *self-adjoint* or *Hermitian* if $A = A^*$. Equivalently, A is self-adjoint if

$$\forall\, x, y \in H, \quad \langle Ax, y\rangle \;=\; \langle x, Ay\rangle. \qquad \diamondsuit$$

Remark 8.3.2. We will focus on self-adjoint operators, but we mention a somewhat broader class that is often useful. An operator A that *commutes* with its adjoint (but need not be equal to its adjoint) is called a *normal operator*. That is, A is normal if $AA^* = A^*A$. All self-adjoint operators are normal, but not all normal operators are self-adjoint (see Problem 8.3.17). $\diamondsuit$

To illustrate self-adjointness, we consider square matrices.

Example 8.3.3. Let $\mathbb{F} = \mathbb{C}$, and let A be an $n \times n$ complex matrix. We saw earlier that the adjoint of A is its Hermitian matrix $A^* = A^{\mathrm{H}} = \overline{A^{\mathrm{T}}}$. Hence, a matrix A is self-adjoint if and only if $A = A^{\mathrm{H}}$, i.e., A equals its Hermitian matrix. When dealing with matrices, it is customary to say that A *is Hermitian* instead of A *is self-adjoint*.

If $\mathbb{F} = \mathbb{R}$ then all of the entries of A are real, and the complex conjugate in the definition of A^{H} is irrelevant. Hence the adjoint of A is its transpose A^{T} in this case. Thus, a real matrix is self-adjoint if and only if $A = A^{\mathrm{T}}$. A matrix that equals its transpose is said to be *symmetric*. $\diamondsuit$

A self-adjoint operator need not be a topological isomorphism; in fact a self-adjoint operator need not be either injective or surjective. For example, the zero function $0\colon H \to H$ is self-adjoint.

Example 8.3.4. Let $\lambda = (\lambda_n)_{n\in\mathbb{N}}$ be a bounded sequence, and let M_λ be the operator defined in Theorem 8.1.4. We saw in Example 8.2.5 that the adjoint of M_λ is $M_\lambda^* = M_{\overline{\lambda}}$, where $\overline{\lambda} = \big(\overline{\lambda_n}\big)_{n\in\mathbb{N}}$. We will determine when M_λ is self-adjoint.

Since $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal basis we have $\langle e_m, e_n\rangle = \delta_{mn}$ (the Kronecker delta that is 1 when $m = n$ and zero otherwise). Therefore

$$M_\lambda e_m \;=\; \sum_{n=1}^{\infty} \lambda_n \, \langle e_m, e_n\rangle \, e_n \;=\; \sum_{n=1}^{\infty} \lambda_n \, \delta_{mn} \, e_n \;=\; \lambda_m \, e_m.$$

A similar calculation shows that $M_{\overline{\lambda}} e_m = \overline{\lambda_m}\, e_m$. Consequently, if M_λ is self-adjoint then

$$\lambda_m \, e_m \;=\; M_\lambda e_m \;=\; M_\lambda^* e_m \;=\; M_{\overline{\lambda}} e_m \;=\; \overline{\lambda_m}\, e_m.$$

Since $e_m \neq 0$, it follows that $\lambda_m = \overline{\lambda_m}$, which tells us that λ_m is real. This is true for every m, so we have shown that if M_λ is self-adjoint then λ is a sequence of real scalars. Conversely, if λ is a sequence of real scalars, then $\overline{\lambda} = \lambda$ and therefore $M_\lambda^* = M_{\overline{\lambda}} = M_\lambda$. Thus

M_λ is self-adjoint if and only if each λ_n is real.

Note that if $\mathbb{F} = \mathbb{R}$ then each λ_n is real by definition, so M_λ is automatically self-adjoint in this case. However, if $F = \mathbb{C}$ then M_λ need not be self-adjoint. For example, if $\lambda = (i^n)_{n\in\mathbb{N}} = (i, -1, -i, 1, \dots)$, then M_λ is not self-adjoint. On the other hand, since $|\lambda_n| = |i^n| = 1$, Theorem 8.1.4 implies that M_λ is unitary for this choice of λ. $\diamondsuit$

The following properties of self-adjoint operators are an immediate consequence of Theorem 8.2.6.

Corollary 8.3.5. *Let H be a Hilbert space. If $A \in \mathcal{B}(H)$ is self-adjoint, then the following statements hold.*

(a) $\ker(A) = \mathrm{range}(A)^\perp$.

(b) $\ker(A)^\perp = \overline{\mathrm{range}}(A)$.

(c) *A is injective if and only if* $\mathrm{range}(A)$ *is dense in H.* $\diamondsuit$

Example 8.3.6. Let $\lambda = (\frac{1}{n})_{n\in\mathbb{N}}$, so M_λ has the form

$$M_\lambda x \;=\; \sum_{n=1}^{\infty} \frac{1}{n} \, \langle x, e_n\rangle \, e_n, \qquad x \in H.$$

Since λ is a real sequence, we know from Example 8.3.4 that M_λ is self-adjoint. Further, M_λ is injective (why?), so Corollary 8.3.5 implies that $\operatorname{range}(M_\lambda)$ is dense in H. However, M_λ is not surjective (for example, the sequence $y = \bigl(\frac{1}{n}\bigr)_{n\in\mathbb{N}} \in \ell^2$ does not belong to the range of M_λ). Therefore for this choice of λ we see that $\operatorname{range}(M_\lambda)$ is a dense but proper subspace of H. ♢

The next theorem is one of the few results in this volume where the choice of scalar field makes a significant difference. In particular, this theorem shows that if $\mathbb{F} = \mathbb{C}$, then a bounded operator $A\colon H \to H$ is is self-adjoint if and only if the inner product $\langle Ax, x\rangle$ is real for every vector $x \in H$. The statement of this theorem becomes *false* if the scalar field is real (see Problem 8.3.12). However, as long as we are working over the complex field, this result gives us a convenient way to recognize self-adjoint operators. When we want to distinguish between a Hilbert space over the real field and one over the complex field, it is sometimes convenient to say that H is a *real Hilbert space* if it is a Hilbert space and $\mathbb{F} = \mathbb{R}$, and it is a *complex Hilbert space* if it is a Hilbert space and $\mathbb{F} = \mathbb{C}$.

Theorem 8.3.7. *Let $\mathbb{F} = \mathbb{C}$, and assume that H is a complex Hilbert space. If $A \in \mathcal{B}(H)$, then*

$$A \text{ is self-adjoint} \quad\Longleftrightarrow\quad \langle Ax, x\rangle \in \mathbb{R} \quad \text{for all } x \in H.$$

Proof. $\Rightarrow$. Assume that A is self-adjoint, and fix $x \in H$. Using the conjugate symmetry of the inner product, the definition of the adjoint, and the fact that $A = A^*$, we compute that

$$\overline{\langle Ax, x\rangle} \;=\; \langle x, Ax\rangle \;=\; \langle A^*x, x\rangle \;=\; \langle Ax, x\rangle.$$

Therefore $\langle Ax, x\rangle$ is real.

$\Leftarrow$. Assume that $\langle Ax, x\rangle$ is real for all $x \in H$. Given $x, y \in H$, we have

$$\langle A(x+y), x+y\rangle \;=\; \langle Ax+Ay, x+y\rangle \;=\; \langle Ax, x\rangle + \langle Ax, y\rangle + \langle Ay, x\rangle + \langle Ay, y\rangle.$$

Since the three inner products $\langle A(x+y), x+y\rangle$, $\langle Ax, x\rangle$, and $\langle Ay, y\rangle$ are all real, we conclude that $\langle Ax, y\rangle + \langle Ay, x\rangle$ is real and therefore equals its own complex conjugate. Consequently,

$$\begin{aligned}
\langle Ax, y\rangle + \langle Ay, x\rangle \;&=\; \overline{\langle Ax, y\rangle + \langle Ay, x\rangle} && (z = \overline{z} \text{ when } z \text{ is real}) \\
&=\; \overline{\langle Ax, y\rangle} + \overline{\langle Ay, x\rangle} && (\overline{w+z} = \overline{w} + \overline{z}) \\
&=\; \langle y, Ax\rangle + \langle x, Ay\rangle && (\text{conjugate symmetry}).
\end{aligned} \tag{8.12}$$

Similarly, after examining the equation

$$\langle A(x+iy), x+iy\rangle \;=\; \langle Ax, x\rangle - i\,\langle Ax, y\rangle + i\,\langle Ay, x\rangle + \langle Ay, y\rangle,$$

we find that

$$\langle Ax, y\rangle - \langle Ay, x\rangle \;=\; -\langle y, Ax\rangle + \langle x, Ay\rangle. \tag{8.13}$$

Adding equations (8.12) and (8.13) gives $2\langle Ax, y\rangle = 2\langle x, Ay\rangle$. Combining this with the definition of the adjoint, we obtain

$$\langle Ax, y\rangle \;=\; \langle x, Ay\rangle \;=\; \langle A^*x, y\rangle.$$

Since this is true for all vectors x, $y \in H$, it follows that $A = A^*$. $\square$

Next we give a useful formula for computing the operator norm of a self-adjoint map. This theorem holds for both real and complex Hilbert spaces.

Theorem 8.3.8. *Let H be a Hilbert space. If $A \in \mathcal{B}(H)$ is self-adjoint, then*

$$\|A\| \;=\; \sup_{\|x\|=1} |\langle Ax, x\rangle|.$$

Proof. For convenience of notation, let

$$M \;=\; \sup_{\|x\|=1} |\langle Ax, x\rangle|. \tag{8.14}$$

If z is any nonzero vector in H then $z/\|z\|$ is a unit vector, so by applying equation (8.14) and rearranging (and noting that the inequality is trivial if $z = 0$), we see that

$$|\langle Az, z\rangle| \;\le\; M\,\|z\|^2, \qquad \text{all } z \in H. \tag{8.15}$$

Now choose any vector $x \in H$. Then, by the Cauchy–Bunyakovski–Schwarz Inequality,

$$|\langle Ax, x\rangle| \;\le\; \|Ax\|\,\|x\| \;\le\; \|A\|\,\|x\|^2.$$

Taking the supremum over all unit vectors x, we conclude that

$$M \;=\; \sup_{\|x\|=1} |\langle Ax, x\rangle| \;\le\; \sup_{\|x\|=1} \|A\|\,\|x\|^2 \;=\; \|A\|.$$

To prove the opposite inequality, let x and y be unit vectors in H. Expanding the inner products, cancelling terms, and using the fact that $A = A^*$, it follows that

$$\begin{aligned}
\big\langle A(x+y),\, x+y\big\rangle \;-\; \big\langle A(x-y),\, x-y\big\rangle \;&=\; 2\,\langle Ax, y\rangle \;+\; 2\,\langle Ay, x\rangle\\
&=\; 2\,\langle Ax, y\rangle \;+\; 2\,\overline{\langle y, Ax\rangle}\\
&=\; 2\,\langle Ax, y\rangle \;+\; 2\,\overline{\langle Ax, y\rangle}\\
&=\; 4\,\mathrm{Re}\langle Ax, y\rangle.
\end{aligned}$$

Therefore

$$\begin{aligned}
4\bigl|\mathrm{Re}\langle Ax,y\rangle\bigr| &= \bigl|\bigl\langle A(x+y),\, x+y\bigr\rangle \;-\; \bigl\langle A(x-y),\, x-y\bigr\rangle\bigr| \\
&\le \bigl|\bigl\langle A(x+y),\, x+y\bigr\rangle\bigr| \;+\; \bigl|\bigl\langle A(x-y),\, x-y\bigr\rangle\bigr| \\
&\le M\,\|x+y\|^2 \;+\; M\,\|x-y\|^2 \qquad \text{(by equation (8.15))} \\
&= 2M\left(\|x\|^2+\|y\|^2\right) \qquad\qquad \text{(Parallelogram Law)} \\
&= 4M.
\end{aligned}$$

Thus

$$|\mathrm{Re}\langle Ax,y\rangle| \;\le\; M \quad \text{for all unit vectors } x,y. \tag{8.16}$$

At this point, if $\mathbb{F}=\mathbb{R}$ then we apply Problem 8.1.5 and conclude that

$$\|A\| \;=\; \sup_{\|x\|=\|y\|=1} |\langle Ax,y\rangle| \;=\; \sup_{\|x\|=\|y\|=1} |\mathrm{Re}\langle Ax,y\rangle| \;\le\; M.$$

Therefore we are done if the scalar field is real.

On the other hand, if $\mathbb{F}=\mathbb{C}$ then we fix unit vectors $x,\, y \in H$ and let α be a complex scalar with unit modulus such that $\alpha\,\langle Ax,y\rangle = |\langle Ax,y\rangle|$. Then

$$|\langle Ax,y\rangle| \;=\; \alpha\,\langle Ax,y\rangle \;=\; \langle Ax,\overline{\alpha}y\rangle. \tag{8.17}$$

Since $\overline{\alpha}y$ is a unit vector and since $\langle Ax,\overline{\alpha}y\rangle$ is a real number, by combining equations equations (8.16) and (8.17) we obtain

$$|\langle Ax,y\rangle| \;=\; \langle Ax,\overline{\alpha}y\rangle \;=\; \mathrm{Re}\langle Ax,\overline{\alpha}y\rangle \;\le\; M.$$

Applying Problem 8.1.5, we find that $\|A\| \le M$. □

We prove a useful corollary for operators on a complex Hilbert space. This result is false as stated if we take $\mathbb{F}=\mathbb{R}$, but it does hold for operators on a real Hilbert space if we add, as an explicit extra hypothesis, the assumption that $A = A^*$.

Corollary 8.3.9. *Let $\mathbb{F}=\mathbb{C}$, and assume that H is a complex Hilbert space. If $A \in \mathcal{B}(H)$, then*

$$\langle Ax,x\rangle \;=\; 0 \text{ for every } x \in H \qquad \Longleftrightarrow \qquad A = 0. \qquad \diamondsuit$$

Proof. $\Rightarrow$. Assume $\langle Ax,x\rangle = 0$ for every x. Then $\langle Ax,x\rangle$ is always real, so Theorem 8.3.7 implies that A is self-adjoint. Therefore, by Theorem 8.3.8,

$$\|A\| \;=\; \sup_{\|x\|=1} |\langle Ax,x\rangle| \;=\; 0. \qquad \square$$

Now we define positive and positive definite operators, which are special types of self-adjoint operators.

Definition 8.3.10 (Positive and Positive Definite Operators). Assume that $A \in \mathcal{B}(H)$ is a self-adjoint operator on a Hilbert space H.

(a) We say that A is *positive* (or *nonnegative*) if $\langle Ax, x\rangle \ge 0$ for every $x \in H$. In this case we write $A \ge 0$.

(b) We say that A is *positive definite* (or *strictly positive*) if $\langle Ax, x\rangle > 0$ for every nonzero $x \in H$. In this case we write $A > 0$. $\diamondsuit$

For complex Hilbert spaces, Theorem 8.3.7 implies that the assumption of self-adjointness in Definition 8.3.10 is redundant. We state this precisely as the following lemma.

Lemma 8.3.11. *Assume $\mathbb{F} = \mathbb{C}$, and let H be a complex Hilbert space. If $A \in \mathcal{B}(H)$ is a bounded linear operator, then the following statements hold.*

(a) *A is positive $\iff$ $\langle Ax, x\rangle \ge 0$ for every $x \in H$.*

(b) *A is positive definite $\iff$ $\langle Ax, x\rangle > 0$ for every $x \ne 0$.* $\diamondsuit$

Problems

8.3.12. Give an example of a real Hilbert space H and a nonzero linear operator $A \in \mathcal{B}(H)$ such that $\langle Ax, x\rangle = 0$ for every $x \in H$.

8.3.13. Let H be a Hilbert space, and suppose that A, $B \in \mathcal{B}(H)$ are self-adjoint. Prove that ABA and BAB are self-adjoint, but AB is self-adjoint if and only if $AB = BA$. Exhibit self-adjoint operators A, B that do not commute.

8.3.14. Let H be a Hilbert space. Show that if $A \in \mathcal{B}(H)$, then $A + A^*$, AA^*, A^*A, and $AA^* - A^*A$ are all self-adjoint, and AA^* and A^*A are positive.

8.3.15. Let M be a closed subspace of a Hilbert space H, and fix $P \in \mathcal{B}(H)$. Show that P is the orthogonal projection of H onto M if and only if $P^2 = P$, $P^* = P$, and $\mathrm{range}(P) = M$.

8.3.16. Let M be a closed subspace of a Hilbert space H, and let P be the orthogonal projection of H onto M. Let $\{e_n\}_{n\in\mathbb{N}}$, $\{f_m\}_{m\in I}$, and $\{g_m\}_{m\in J}$, be orthonormal bases for H, M, and $M^\perp$, respectively. Prove that

$$\sum_{n=1}^{\infty} \|Pe_n\|^2 \;=\; \sum_{n=1}^{\infty}\biggl(\sum_{m\in I} |\langle Pe_n, f_m\rangle|^2 \;+\; \sum_{m\in J} |\langle Pe_n, g_m\rangle|^2\biggr),$$

and use this to show that $\sum \|Pe_n\|^2 = \dim(M)$.

8.3.17. Let H be a Hilbert space. We say that an operator $A \in \mathcal{B}(H)$ is *normal* if A commutes with its adjoint, i.e., if $AA^* = A^*A$. Prove the following statements.

(a) A is normal if and only if $\|Ax\| = \|A^*x\|$ for every $x \in H$.

(b) If λ is any bounded sequence of scalars then the multiplication operator M_λ introduced in Theorem 8.1.4 is normal.

8.3.18. Let H and K be Hilbert spaces. Given $A \in \mathcal{B}(H,K)$, prove the following statements.

(a) $A^*A \in \mathcal{B}(H)$ and $AA^* \in \mathcal{B}(K)$ are self-adjoint and positive.

(b) A^*A is positive definite $\iff$ A is injective $\iff$ A^* has dense range.

(c) AA^* is positive definite $\iff$ A^* is injective $\iff$ A has dense range.

8.3.19. The *Hilbert matrix* is

$$\mathcal{H} = \begin{bmatrix} 1 & 1/2 & 1/3 & 1/4 & \cdots \\ 1/2 & 1/3 & 1/4 & 1/5 & \\ 1/3 & 1/4 & 1/5 & 1/6 & \\ 1/4 & 1/5 & 1/6 & 1/7 & \\ \vdots & & & & \ddots \end{bmatrix}.$$

Define

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 1/2 & 1/2 & 0 & 0 & \\ 1/3 & 1/3 & 1/3 & 0 & \\ 1/4 & 1/4 & 1/4 & 1/4 & \\ \vdots & & & & \ddots \end{bmatrix} \quad \text{and} \quad L = \begin{bmatrix} 1 & 1/2 & 1/3 & 1/4 & \cdots \\ 1/2 & 1/2 & 1/3 & 1/4 & \\ 1/3 & 1/3 & 1/3 & 1/4 & \\ 1/4 & 1/4 & 1/4 & 1/4 & \\ \vdots & & & & \ddots \end{bmatrix}.$$

For this problem, take as given the fact that C determines a bounded map of ℓ^2 into ℓ^2 (compare Section 6.7). Prove the following statements.

(a) $L = CC^*$, so $L \geq 0$ (i.e., L is a positive operator).

(b) $I - (I-C)(I-C)^* = \operatorname{diag}(1, 1/2, 1/3, 1/4, \dots)$, the diagonal matrix with entries $1, 1/2, \dots$ on the diagonal.

(c) $\|(I-C)\|^2 = \|(I-C)(I-C)^*\| \leq 1$.

(d) $\|C\| \leq 2$ and $\|L\| \leq 4$.

Remark: It is a fact (though not so easy to prove) that if A, B are symmetric matrices and $a_{ij} \leq b_{ij}$ for all $i, j \in \mathbb{N}$, then $\|A\| \leq \|B\|$. Consequently, $\|\mathcal{H}\| \leq \|L\| \leq 4$. It is known that the operator norm of the Hilbert matrix is precisely $\|\mathcal{H}\| = \pi$ [Cho83].

8.4 Review of Compact Sets in Metric Spaces

A *compact set* is a special type of closed set. The abstract definition of a compact set is phrased in terms of coverings by open sets, but we will also derive several equivalent reformulations of the definition for subsets of metric spaces. One of these will show if K is a compact set, then every infinite sequence $\{x_n\}_{n\in\mathbb{N}}$ contained in K has a subsequence that converges to an element of K. We will also prove that in the Euclidean space $\mathbb{F}^d$, a subset is compact if and only if it is both closed and bounded. However, in other metric spaces the question of whether a set is compact or not can be a more subtle issue. For example, we will see that the closed unit disk in ℓ^1 is a *closed and bounded set that is not compact*!

By a *cover* of a set S, we mean a collection of sets $\{E_i\}_{i\in I}$ whose union contains S. If each set E_i is open, then we call $\{E_i\}_{i\in I}$ an *open cover* of S. The index set I may be finite or infinite (even uncountable). If I is finite then we call $\{E_i\}_{i\in I}$ a *finite cover* of S. Thus a *finite open cover* of S is a collection of finitely many open sets whose union contains S.

Definition 8.4.1 (Compact Set). A subset K of a metric space X is *compact* if every covering of K by open sets has a finite subcovering. Stated precisely, K is compact if it is the case that whenever

$$K \subseteq \bigcup_{i\in I} U_i,$$

where $\{U_i\}_{i\in I}$ is any collection of open subsets of X, there exist *finitely* many indices $i_1, \dots, i_N \in I$ such that

$$K \subseteq \bigcup_{k=1}^{N} U_{i_k}.$$

If X itself is a compact set, then we say that X is a *compact metric space*. $\diamondsuit$

In order for a set K to be called *compact*, it must be the case that *every* open cover of K has a finite subcover. If there is even one particular open covering of K that does not have a finite subcover, then K is not compact.

Example 8.4.2. (a) Consider the interval $(0,1]$ in the real line $\mathbb{R}$. For each integer $k \in \mathbb{N}$, let U_k be the open interval

$$U_k = \left(\tfrac{1}{k}, 2\right).$$

Then $\{U_k\}_{k\in\mathbb{N}}$ is an open cover of $(0,1]$, but we cannot cover the interval $(0,1]$ using only finitely many of the sets U_k (why not?). Consequently the open cover $\{U_k\}_{k\in\mathbb{N}}$ contains no finite subcover, so $(0,1]$ is not compact.

Even though $(0,1]$ is not compact, there do exist *some* open coverings that have finite subcoverings. For example, if we set $U_1 = (0,1)$ and $U_2 = (0,2)$, then $\{U_1, U_2\}$ is an open cover of $(0,1]$, and $\{U_2\}$ is a finite subcover. However, as we showed above, it is not true that *every* open cover of $(0,1]$ contains a finite subcover. This is why $(0,1]$ is not compact.

(b) Now consider the interval $[1,\infty)$. Even though this interval is a closed subset of $\mathbb{R}$, we will prove that it is not compact. To see this, set $U_k = (k-1, k+1)$ for each $k \in \mathbb{N}$. Then $\{U_k\}_{k\in\mathbb{N}}$ is an open cover of $[1,\infty)$, but we cannot cover $[0,\infty)$ using only finitely many of the intervals U_k. Hence there is at least one open cover of $[0,\infty)$ that has no finite subcover, so this set is not compact. $\diamondsuit$

We prove next that compact subsets of a metric space are both closed and bounded.

Lemma 8.4.3. *If K is a compact subset of a metric space X, then K is closed and bounded.*

Proof. Suppose that K is a compact set, and fix any particular point $x \in X$. Then the union of the open balls $B_n(x)$ over $n \in \mathbb{N}$ is all of X (why?), so $\{B_n(x)\}_{n\in\mathbb{N}}$ an open cover of K. This cover must have a finite subcover, say

$$\{B_{n_1}(x),\, B_{n_2}(x),\, \dots,\, B_{n_M}(x)\}.$$

By reordering these sets if necessary we can assume that $n_1 < \cdots < n_M$. With this ordering the balls are *nested*:

$$B_{n_1}(x) \;\subseteq\; B_{n_2}(x) \;\subseteq\; \cdots \;\subseteq\; B_{n_M}(x).$$

Since K is contained in the union of these balls it is therefore contained in the ball with largest radius, i.e., $K \subseteq B_{n_M}(x)$. Therefore K is bounded.

It remains to show that K is closed. If $K = X$ then we are done, so assume that $K \neq X$. Fix any point y in $K^{\mathrm{C}} = X\backslash K$. If x is a point in K then $x \neq y$, so by the *Hausdorff property* of metric spaces there must exist disjoint open sets U_x, V_x such that $x \in U_x$ and $y \in V_x$. The collection $\{U_x\}_{x\in K}$ is an open cover of K, so it must contain some finite subcover. That is, there must exist finitely many points $x_1, \dots, x_N \in K$ such that

$$K \;\subseteq\; U_{x_1} \cup \cdots \cup U_{x_N}. \tag{8.18}$$

Each V_{x_j} is disjoint from U_{x_j}, so it follows from equation (8.18) that the set

$$V \;=\; V_{x_1} \cap \cdots \cap V_{x_N}$$

is entirely contained in the complement of K. Thus, V is an open set that satisfies

$$y \;\in\; V \;\subseteq\; K^{\mathrm{C}}.$$

This shows that K^{C} is open, so we conclude that K is closed. $\square$

While every compact subset of a metric space is closed and bounded, there can exist sets that are closed and bounded but not compact. For example, we will see that the "closed unit disk" in ℓ^1 is closed and bounded, but *not compact*. On the other hand, we prove next that, *in the Euclidean space* $\mathbb{F}^d$, a set is compact *if and only if* it closed and bounded.

Theorem 8.4.4 (Heine–Borel Theorem). *If $K \subseteq \mathbb{F}^d$, then K is compact if and only if K is closed and bounded.*

Proof. For simplicity of presentation we will assume in this proof that $\mathbb{F} = \mathbb{R}$, but the same ideas can used to prove the result when $\mathbb{F} = \mathbb{C}$ (this is Problem 8.4.18).

$\Rightarrow$. According to Lemma 8.4.3, every compact set is closed and bounded.

$\Leftarrow$. First we will show that the closed cube $Q = [-R, R]^d$ is a compact subset of $\mathbb{R}^d$. If Q was not compact, then there would exist an open cover $\{U_i\}_{i\in I}$ of Q that has no finite subcover. By bisecting each side of Q, we can divide Q into 2^d closed subcubes whose union is Q. At least one of these subcubes cannot be covered by finitely many of the sets U_i (why not?). Call that subcube Q_1. Then subdivide Q_1 into 2^d closed subcubes. At least one of these, which we call Q_2, cannot be covered by finitely many U_i. Continuing in this way we obtain nested decreasing closed cubes $Q \supseteq Q_1 \supseteq Q_2 \supseteq \cdots$. Appealing to Problem 8.4.16, $\cap Q_k$ is nonempty. Hence there is a point $x \in Q$ that belongs to every Q_k. This point must belong to some set U_i (since these sets cover Q), and since U_i is open there is some $r > 0$ such that $B_r(x) \subseteq U_i$. However, the sidelengths of the cubes Q_k decrease to zero as k increases, so if we choose k large enough then we will have $x \in Q_k \subseteq B_r(x) \subseteq U_i$ (check this!). Hence Q_k is covered by a single set U_i, which is a contradiction. Therefore Q must be compact.

Now let K be an arbitrary closed and bounded subset of $\mathbb{R}^d$. Then K is contained in $Q = [-R, R]^d$ for some $R > 0$. Hence K is a closed subset of the compact set Q. Therefore K is compact by Problem 8.4.11. □

We will give several equivalent reformulations of compactness in terms of the following concepts.

Definition 8.4.5. Let E be a subset of a metric space X.

(a) E is *sequentially compact* if every sequence $\{x_n\}_{n\in\mathbb{N}}$ of points from E contains a convergent subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ whose limit belongs to E.

(b) E is *totally bounded* if given any $r > 0$ we can cover E by finitely many open balls of radius r. That is, for each $r > 0$ there must exist finitely many points $x_1, \dots, x_N \in X$ such that

$$E \subseteq \bigcup_{k=1}^{N} B_r(x_k). \qquad \diamondsuit$$

Here are some examples to illustrate sequential compactness.

Example 8.4.6. (a) Every finite subset of a metric space is sequentially compact (why?).

(b) The interval $(0,1]$ is not sequentially compact. To see why, consider the sequence $\{\frac{1}{n}\}_{n\in\mathbb{N}}$. This is convergent sequence contained in $(0,1]$, but it does not converge to an element of $(0,1]$. Likewise, no subsequence converges to an element of $(0,1]$.

(c) Suppose that E is a nonempty subset of metric space that is not closed. Then E does not contain all of its accumulation points. Hence there must be some point $x \in X$ that is an accumulation point of E but does not belong to E. onsequently there exist points $x_n \in E$ such that $x_n \to x$. Hence $\{x_n\}_{n\in\mathbb{N}}$ is an infinite sequence in E that converges to a point outside of E. Consequently (why?), no subsequence can converge to an element of E, so E is not sequentially compact. Therefore we have proved, by a contrapositive argument, that all sequentially compact sets are closed.

(d) The interval $[1,\infty)$ is closed, but it is not sequentially compact because the sequence $\mathbb{N} = \{n\}_{n\in\mathbb{N}}$ has no convergent subsequences. Thus there are closed sets that are not sequentially compact.

(e) Let D be the closed unit disk in ℓ^1, i.e.,

$$D \;=\; \{x \in \ell^1 : \|x\|_1 \le 1\}.$$

This set is both closed and bounded, but it is not sequentially compact because the standard basis $\{\delta_n\}_{n\in\mathbb{N}}$ is contained in D but contains no convergent subsequences (prove this). $\diamondsuit$

Here are some examples illustrating total boundedness.

Example 8.4.7. (a) The interval $[0,1)$ is totally bounded. Given $r > 0$, let n be large enough that $\frac{1}{n} < r$. Open balls in one dimension are just open intervals, and the finitely many balls with radius $\frac{1}{n}$ centered at the points $0, \frac{1}{n}, \frac{2}{n}, \dots, 1$ cover $[0,1)$. Consequently the balls with radius r centered at these points cover $[0,1)$. Therefore $[0,1)$ is totally bounded (even though it is not closed).

(b) The interval $[1,\infty)$ is not totally bounded because we cannot cover it with finitely many balls with a fixed finite radius.

(c) Based on our intuition from finite dimensions, it may seem that every bounded set is totally bounded. However, according to Problem 8.4.20, the closed unit disk in ℓ^1 is not totally bounded, even though it is both closed and bounded. $\diamondsuit$

We will need the following lemma.

Lemma 8.4.8. *Let E be a sequentially compact subset of a metric space X. If $\{U_i\}_{i\in I}$ is an open cover of E, then there exists a number $\delta > 0$ such that if B is an open ball of radius δ that intersects E, then there is an $i \in I$ such that $B \subseteq U_i$.*

Proof. Let $\{U_i\}_{i\in I}$ be an open cover of E. We want to prove that there is a $\delta > 0$ such that

$$B_\delta(x) \cap E \neq \varnothing \quad \Longrightarrow \quad B_\delta(x) \subseteq U_i \text{ for some } i \in I.$$

Suppose that no $\delta > 0$ has this property. Then for each positive integer n, there must exist some open ball with radius $\frac{1}{n}$ that intersects E but is not contained in any U_i. Call this open ball G_n. For each n, choose a point $x_n \in G_n \cap E$. Then since $\{x_n\}_{n\in\mathbb{N}}$ is a sequence in E and E is sequentially compact, there must be a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ that converges to a point $x \in E$. Since $\{U_i\}_{i\in I}$ is a cover of E, we must have $x \in U_i$ for some $i \in I$, and since U_i is open there must exist some $r > 0$ such that $B_r(x) \subseteq U_i$. Now choose k large enough that we have both

$$\frac{1}{n_k} < \frac{r}{3} \qquad \text{and} \qquad \mathrm{d}\big(x, x_{n_k}\big) < \frac{r}{3}.$$

Keeping in mind that G_{n_k} contains x_{n_k}, G_{n_k} is an open ball with radius $1/n_k$, the distance from x to x_{n_k} is less than $r/3$, and $B_r(x)$ has radius r, it follows (why?) that $G_{n_k} \subseteq B_r(x) \subseteq U_i$. But this is a contradiction. □

Now we prove some reformulations of compactness that hold for subsets of complete metric spaces.

Theorem 8.4.9. *If K is a subset of a* complete *metric space X, then the following three statements are equivalent.*

(a) *K is compact.*

(b) *K is sequentially compact.*

(c) *K is totally bounded and complete (i.e., every Cauchy sequence of points from K converges to a point in K).*

Proof. (a) $\Rightarrow$ (b). We will prove the contrapositive statement. Suppose that K is not sequentially compact. Then there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ in K that has no subsequence that converges to an element of K. Choose any point $x \in K$. If every open ball centered at x contains infinitely many of the points x_n, then by considering radii $r = \frac{1}{k}$ we can construct a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ that converges to x. This is a contradiction, so there must exist some open ball centered at x that contains only finitely many x_n. If we call this ball B_x, then $\{B_x\}_{x\in K}$ is an open cover of K that contains no finite subcover (because any finite union $B_{x_1} \cup \cdots \cup B_{x_M}$ can contain only finitely many of the x_n). Consequently K is not compact.

(b) $\Rightarrow$ (c). Suppose that K is sequentially compact, and suppose that $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in K. Then since K is sequentially compact, there is a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ that converges to some point $x \in K$. Hence $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy and has a convergent subsequence. This implies (why?) that $x_n \to x$. Therefore K is complete.

Suppose that K is not totally bounded. Then there is a radius $r > 0$ such that K cannot be covered by finitely many open balls of radius r centered at points of X. Choose any point $x_1 \in K$. Since K cannot be covered by a single r-ball, K cannot be a subset of $B_r(x_1)$. Hence there exists a point $x_2 \in K \setminus B_r(x_1)$. In particular, $\mathrm{d}(x_2, x_1) \geq r$. But K cannot be covered by two r-balls, so there must exist a point x_3 that belongs to $K \setminus \bigl(B_r(x_1) \cup B_r(x_2)\bigr)$. In particular, we have both $\mathrm{d}(x_3, x_1) \geq r$ and $\mathrm{d}(x_3, x_2) \geq r$. Continuing in this way we obtain a sequence of points $\{x_n\}_{n\in\mathbb{N}}$ in K that has no convergent subsequence, which is a contradiction.

(c) $\Rightarrow$ (b). Assume that K is complete and totally bounded, and let $\{x_n\}_{n\in\mathbb{N}}$ be any sequence of points in K. Since K can be covered by finitely many open balls of radius $\frac{1}{2}$, at least one of those balls must contain infinitely many of the points x_n. That is, there is some infinite subsequence $\{x_n^{(1)}\}_{n\in\mathbb{N}}$ of $\{x_n\}_{n\in\mathbb{N}}$ that is contained in an open ball of radius $\frac{1}{2}$. The Triangle Inequality therefore implies that

$$\forall\, m, n \in \mathbb{N}, \quad \mathrm{d}\bigl(x_m^{(1)}, x_n^{(1)}\bigr) \;<\; 1.$$

Similarly, since K can be covered by finitely many open balls of radius $\frac{1}{4}$, there is some subsequence $\{x_n^{(2)}\}_{n\in\mathbb{N}}$ of $\{x_n^{(1)}\}_{n\in\mathbb{N}}$ such that

$$\forall\, m, n \in \mathbb{N}, \quad \mathrm{d}\bigl(x_m^{(2)}, x_n^{(2)}\bigr) \;<\; \frac{1}{2}.$$

Continuing by induction, for each $k > 1$ we find a subsequence $\{x_n^{(k)}\}_{n\in\mathbb{N}}$ of $\{x_n^{(k-1)}\}_{n\in\mathbb{N}}$ such that $\mathrm{d}\bigl(x_m^{(k)}, x_n^{(k)}\bigr) < \frac{1}{k}$ for all m, $n \in \mathbb{N}$.

Now consider the diagonal subsequence $\{x_k^{(k)}\}_{k\in\mathbb{N}}$. Given $\varepsilon > 0$, let n be large enough that $\frac{1}{n} < \varepsilon$. If $j \geq k > n$, then $x_j^{(j)}$ is one element of the sequence $\{x_n^{(k)}\}_{n\in\mathbb{N}}$ (why?), say $x_j^{(j)} = x_n^{(k)}$. Hence

$$\mathrm{d}\bigl(x_j^{(j)}, x_k^{(k)}\bigr) \;=\; \mathrm{d}\bigl(x_n^{(k)}, x_k^{(k)}\bigr) \;<\; \frac{1}{k} \;<\; \frac{1}{n} \;<\; \varepsilon.$$

Thus $\{x_k^{(k)}\}_{k\in\mathbb{N}}$ is a Cauchy subsequence of the original sequence $\{x_n\}_{n\in\mathbb{N}}$. Since K is complete, this subsequence must converge to some element of E. Hence K is sequentially compact.

(b) $\Rightarrow$ (a). Assume that K is sequentially compact. Since we have already proved that statement (b) implies statement (c), we know that K is complete and totally bounded.

Suppose that $\{U_i\}_{i\in I}$ is any open cover of K. By Lemma 8.4.8, there exists a $\delta > 0$ such that if B is an open ball of radius δ that intersects E, then there is an $i \in I$ such that $B \subseteq U_i$. However, K is totally bounded, so we can cover K by finitely many open balls of radius δ. Each of these balls is contained in some U_i, so K is covered by finitely many U_i. $\square$

The type of argument that was used to prove the implication (c) $\Rightarrow$ (b) of Theorem 8.4.9 is often called a *Cantor diagonalization argument*. We will see another argument of this type in the proof of Theorem 8.5.6.

We saw in Example 8.4.6 that the closed unit disk in ℓ^1 is not sequentially compact. Applying Theorem 8.4.9, it follows that D is neither compact nor totally bounded, even though it is both closed and bounded.

Problem 8.4.21 constructs an interesting compact subset of ℓ^1 known as a *Hilbert cube*.

Problems

8.4.10. (a) Prove directly from the definition that every finite subset of a metric space, including the empty set, is compact.

(b) In $X = \mathbb{R}$, exhibit an open cover of $E = \{\frac{1}{n}\}_{n\in\mathbb{N}}$ that has no finite subcover.

8.4.11. Let K be a compact subset of a metric space X, and let E be a subset of K. Prove that E is closed if and only if E is compact.

8.4.12. Let X be a metric space. Prove that an arbitrary intersection of compact subsets of X is compact, and a finite union of compact subsets of X is compact.

8.4.13. (a) Suppose that F and K are nonempty, disjoint subsets of a metric space X. Prove that if F is closed and K is compact, then $\operatorname{dist}(F,K) > 0$, where $\operatorname{dist}(F,K) = \inf\{\mathrm{d}(x,y) : x \in F, y \in K\}$.

(b) Show by example that the distance between two nonempty disjoint closed sets E and F can be zero.

8.4.14. Prove that a compact metric space is both complete and separable.

8.4.15. Let X and Y be metric spaces. The *graph* of a function $f\colon X \to Y$ is $\operatorname{graph}(f) = \{(x, f(x)) : x \in X\}$. Prove that if f is continuous and X is compact, then $\operatorname{graph}(f)$ is a compact subset of $X \times Y$.

8.4.16. (a) Let $I_1 \supseteq I_2 \supseteq \cdots$ be a nested decreasing sequence of closed intervals in $\mathbb{R}$. Prove that $\cap I_n \neq \varnothing$.

(b) Let $Q_1 \supseteq Q_2 \supseteq \cdots$ be a nested decreasing sequence of closed cubes in $\mathbb{R}^d$. Prove that $\cap Q_n \neq \varnothing$.

Remark: This problem is a special case of Problem 8.4.17, but it is instructive to prove it directly.

8.4.17. (a) Prove the *Cantor Intersection Theorem* for metric spaces: If $K_1 \supseteq K_2 \supseteq \cdots$ is a nested decreasing sequence of nonempty compact subsets of a metric space X, then $\cap K_n \neq \varnothing$.

(b) Show by example that it is possible for the intersection of a nested decreasing sequence of nonempty *closed* sets to be empty.

8.4.18. We take $\mathbb{F} = \mathbb{C}$ in this problem.

(a) Given $R > 0$, prove that

$$Q \;=\; \bigl\{(a_1 + ib_1, \dots, a_d + ib_d) : -R \le a_i, b_i \le R\bigr\}$$

is a compact subset of $\mathbb{C}^d$.

(b) Prove the Heine–Borel Theorem (Theorem 8.4.4) for the case $\mathbb{F} = \mathbb{C}$.

8.4.19. Let E be a totally bounded subset of a metric space X.

(a) Prove directly that $\overline{E}$ is totally bounded.

(b) Show that if X is complete, then $\overline{E}$ is compact.

8.4.20. The closed unit disk in ℓ^1 is $D = \{x \in \ell^1 : \|x\|_1 \le 1\}$. Give direct proofs of the following statements.

(a) D is a closed and bounded subset of ℓ^1.

(b) D is not sequentially compact.

(c) D is not totally bounded.

(d) The collection of open balls $\bigl\{B_2(\pm\delta_n)\bigr\}_{n\in\mathbb{N}}$ covers D, but does not have a finite subcover, so D is not compact.

8.4.21. Let K be the "rectangular box" in ℓ^1 that has sides of length $\frac{1}{2}, \frac{1}{4}, \dots,$ i.e.,

$$K \;=\; \bigl\{y = (y_k)_{k\in\mathbb{N}} : 0 \le y_k \le 2^{-k}\bigr\}.$$

This problem will show that K, which is known as a *Hilbert cube*, is a compact subset of ℓ^1.

a) Suppose that $\{z_k\}_{k\in\mathbb{N}}$ is a sequence of elements of K and z_k converges *componentwise* to $y \in \ell^1$. Prove that $z_k \to y$ in ℓ^1-norm.

(b) Suppose that $\{x_n\}_{n\in\mathbb{N}}$ is an infinite sequence of elements of K. Use a Cantor diagonalization argument to prove that there is a subsequence $\{x_k^{(k)}\}_{k\in\mathbb{N}}$ that converges componentwise to a sequence $y = (y(k))_{k\in\mathbb{N}}$.

(c) Prove that K is sequentially compact in ℓ^1 (and therefore is also compact and totally bounded).

(d) Is there anything special about the numbers 2^{-k}, or can we replace them with numbers $a_k > 0$ that satisfy $\sum a_k < \infty$?

8.5 Compact Operators on Hilbert Spaces

We will study the special class of *compact operators* on Hilbert spaces in this section. In order to simplify the presentation, we introduce a notation for the closed unit disk in a Hilbert space.

Notation 8.5.1. The *closed unit ball* or *closed unit disk* in H will be denoted by

$$D_H \;=\; \overline{B_1(0)} \;=\; \bigl\{x \in H : \|x\| \le 1\bigr\}. \qquad \diamondsuit$$

The closed unit disk is both closed and bounded. However, D_H is compact if and only if H is finite dimensional.

Assume $T\colon H \to K$ is a continuous linear operator, and consider $T(D_H)$, the image of the closed unit disk under T. What type of set is $T(D_H)$? If H is finite-dimensional, then D_H is a compact subset of H, and therefore $T(D_H)$ is compact since continuous functions maps compact sets to compact sets. However, if H is infinite-dimensional then D_H is not compact, and $T(D_H)$ need not be compact either (for example, think about what happens if T is the identity operator). Since D_H is a bounded set and T is a bounded operator, we do know that $T(D_H)$ is a bounded subset of K, but in general $T(D_H)$ need not be closed (see Problem 8.6.5 for an example). On the other hand, there are some linear operators T that map D_H *into* a compact subset of K. In this case the closure of $T(D_H)$ is compact. We call these the *compact operators* on H.

Definition 8.5.2 (Compact Operators). Let H and K be Hilbert spaces.

(a) A linear operator $T\colon H \to K$ is *compact* if the closure of $T(D_H)$ is a compact subset of K.

(b) The set of all compact operators from H to K is denoted by

$$\mathcal{B}_0(H,K) \;=\; \bigl\{T\colon H \to K \;:\; T \text{ is compact}\bigr\}.$$

When $H = K$ we use the abbreviation $\mathcal{B}_0(H) = \mathcal{B}_0(H,H)$. $\diamondsuit$

As we observed above, if the domain H is finite dimensional, then every linear operator $T\colon H \to K$ is compact (even if K is infinite dimensional). In contrast, Problem 8.5.8 shows that if H is infinite dimensional, then there always exists a linear operator $T\colon H \to K$ that is not compact.

Note that that while Definition 8.5.2 does require a compact operator to be linear, it does not explicitly require continuity. However, we prove next that all compact operators are automatically continuous.

Theorem 8.5.3. *If H and K are Hilbert spaces, then $\mathcal{B}_0(H,K) \subseteq \mathcal{B}(H,K)$, i.e., every compact operator $T\colon H \to K$ is bounded.*

Proof. We will prove the contrapositive statement. Assume that $T\colon H \to K$ is linear but unbounded. Then the operator norm of T is infinite, i.e.,

$$\|T\| \;=\; \sup_{\|x\|=1} \|Tx\| \;=\; \infty.$$

Consequently, for each $n \in \mathbb{N}$ there must exist a vector $x_n \in H$ such that $\|x_n\| = 1$ but $\|Tx_n\| \ge n$. Hence $T(D_H)$ is an unbounded set. As every compact set is bounded, it follows that $T(D_H)$ cannot be contained in any compact set, and therefore T is not compact. $\square$

The next theorem gives some equivalent formulations of compact operators.

Theorem 8.5.4. *Let H and K be Hilbert spaces. If $T\colon H \to K$ is linear, then the following three statements are equivalent.*

(a) *T is compact.*

(b) *$T(D_H)$ is totally bounded.*

(c) *If $\{x_n\}_{n\in\mathbb{N}}$ is a bounded sequence in H, then $\{Tx_n\}_{n\in\mathbb{N}}$ contains a convergent subsequence in K.*

Proof. We will prove one implication, and assign the proof of the remaining implications as Problem 8.5.9.

(b) $\Rightarrow$ (c). Assume that $E = T(D_H)$ is totally bounded. We claim that $\overline{E}$ is also totally bounded. To see this, fix any $r > 0$. Then, since E is totally bounded, there exist finitely many points $x_1, \dots, x_N$ such that

$$E \subseteq B_r(x_1) \cup \cdots \cup B_r(x_N).$$

Suppose that $y \in \overline{E}$. Then there exist points $y_n \in E$ such that $y_n \to y$. Choose n large enough that $\|y - y_n\| < r$. As $y_n \in E$, there is some k such that $y_n \in B_r(x_k)$. Hence $\|y_n - x_k\| < r$, and therefore, by the Triangle Inequality, $\|y - x_k\| < 2r$. Hence

$$\overline{E} \subseteq B_{2r}(x_1) \cup \cdots \cup B_{2r}(x_N).$$

Thus $\overline{E}$ is totally bounded. Since it is also closed and H is complete, Theorem 8.4.9 implies that $\overline{E}$ is compact. Therefore T is a compact operator. $\square$

Part (c) of Theorem 8.5.4 often provides the simplest method of proving that a given operator is compact. To illustrate, we will use that reformulation to prove that $\mathcal{B}_0(H,K)$ is a subspace (and not just a subset) of $\mathcal{B}(H,K)$.

Lemma 8.5.5. *Let H and K be Hilbert spaces. If operators S, $T\colon H \to K$ are compact and a, b are scalars, then $aS + bT$ is compact.*

Proof. Let $\{x_n\}_{n\in\mathbb{N}}$ be a bounded sequence of vectors in H. Since S is compact, there exists a subsequence $\{y_n\}_{n\in\mathbb{N}}$ of $\{x_n\}_{n\in\mathbb{N}}$ such that Sy_n converges. Since $\{y_n\}_{n\in\mathbb{N}}$ is bounded and T is compact, there is a subsequence $\{z_n\}_{n\in\mathbb{N}}$ of $\{y_n\}_{n\in\mathbb{N}}$ such that Tz_n converges. Consequently Sz_n and Tz_n both converge, so $(S+T)z_n = Sz_n + Tz_n$ converges. Therefore $S+T$ is compact by Theorem 8.5.4. A similar argument establishes closure under scalar multiplication. $\square$

Next we prove that the operator norm limit of a sequence of compact operators is compact. Consequently $\mathcal{B}_0(H,K)$ is a *closed* subspace of $\mathcal{B}(H,K)$.

Theorem 8.5.6. *Let H and K be Hilbert spaces, and assume $T_n\colon H \to K$ is compact for each $n \in \mathbb{N}$. If $T \in \mathcal{B}(H,K)$ is such that $\|T - T_n\| \to 0$, then T is compact.*

Proof. We will use a Cantor diagonalization argument (see the proof of Theorem 8.4.9 for another example of this type of argument).

Suppose that $\{x_n\}_{n\in\mathbb{N}}$ is a bounded sequence in H. Since T_1 is compact, there exists a subsequence $\{x_n^{(1)}\}_{n\in\mathbb{N}}$ of $\{x_n\}_{n\in\mathbb{N}}$ such that $\{T_1 x_n^{(1)}\}_{n\in\mathbb{N}}$ converges. Then, since T_2 is compact, there exists a subsequence $\{x_n^{(2)}\}_{n\in\mathbb{N}}$ of $\{x_n^{(1)}\}_{n\in\mathbb{N}}$ such that $\{T_2 x_n^{(2)}\}_{n\in\mathbb{N}}$ converges (note that $\{T_1 x_n^{(2)}\}_{n\in\mathbb{N}}$ also converges). Continue to construct subsequences in this way.

By construction, for every m the vector $x_m^{(m)}$ belongs to the subsequence $\{x_n^{(1)}\}_{n\in\mathbb{N}}$. Likewise, if $m \ge 2$ then $x_m^{(m)}$ is an element of $\{x_n^{(2)}\}_{n\in\mathbb{N}}$. In general, if $k \in \mathbb{N}$ then $\{x_m^{(m)}\}_{m\ge k}$ is a subsequence of $\{x_n^{(k)}\}_{n\in\mathbb{N}}$. Therefore $\{T_k x_m^{(m)}\}_{m\ge k}$ converges. For simplicity of notation, let $y_m = x_m^{(m)}$. Since our original sequence $\{x_n\}_{n\in\mathbb{N}}$ is bounded, $R = \sup \|y_m\| < \infty$.

Fix $\varepsilon > 0$. Then there exists some k such that $\|T - T_k\| < \varepsilon/R$. Since $\{T_k y_m\}_{m\in\mathbb{N}}$ converges, it is Cauchy. Therefore there is some M such that $\|T_k y_m - T_k y_n\| < \varepsilon$ for all $m, n \ge M$. Hence, if $m, n \ge M$ then

$$\begin{aligned}
\|Ty_m - Ty_n\| &\le \|Ty_m - T_k y_m\| + \|T_k y_m - T_k y_n\| + \|T_k y_n - T y_n\| \\
&\le \|T - T_k\|\,\|y_m\| + \varepsilon + \|T_k - T\|\,\|y_n\| \\
&\le R\frac{\varepsilon}{R} + \varepsilon + R\frac{\varepsilon}{R} = 3\varepsilon.
\end{aligned}$$

Thus $\{Ty_m\}_{m\in\mathbb{N}}$ is Cauchy. Since K is a Hilbert space, this sequence must therefore converge. Thus T is compact. □

The next theorem (whose proof is Problem 8.5.9) shows that the composition of a compact operator with a bounded operator, in either order, is compact.

Theorem 8.5.7. *Let H, K, L be Hilbert spaces, and let $A\colon H \to K$ and $B\colon K \to L$ be linear operators. If A is compact and B is bounded, or if A is bounded and B is compact, then BA is compact.* ◊

In particular, if $T \in \mathcal{B}_0(H)$ and $A \in \mathcal{B}(H)$, then AT and TA both belong to $\mathcal{B}_0(H)$. Using the language of abstract algebra, this says that $\mathcal{B}_0(H)$ is a (two-sided) *ideal* in $\mathcal{B}(H)$.

Problems

8.5.8. Let H and K be Hilbert spaces.

(a) Prove that if H is finite dimensional, then every linear mapping $T\colon H \to K$ is both compact and continuous.

(b) Assume H is infinite dimensional. Use Problem 6.3.6 to show that there exists a linear operator $T\colon H \to K$ that is neither compact nor continuous.

8.5.9. (a) Prove the remaining implications in Theorem 8.5.4.

(b) Prove Theorem 8.5.7.

8.5.10. Let H and K be Hilbert spaces, and suppose that $T\colon H \to K$ is compact.

(a) Prove that $T(D_H)$ can be covered by countably many balls $B_{r_n}(x_n)$ such that $\inf r_n = 0$.

(b) Prove that $\overline{\mathrm{range}}(T)$ is a separable subspace of K.

8.5.11. Let H and K be Hilbert spaces, and assume that $T\colon H \to K$ is compact and $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal sequence in H.

(a) Show that any subsequence $\{f_n\}_{n\in\mathbb{N}}$ of $\{e_n\}_{n\in\mathbb{N}}$ has a subsequence $\{g_n\}_{n\in\mathbb{N}}$ such that $\{Tg_n\}_{n\in\mathbb{N}}$ converges to the zero vector in K.

(b) Prove that $Te_n \to 0$ as $n \to \infty$.

8.5.12. Assume H is an infinite-dimensional Hilbert space and $T\colon H \to K$ is compact and injective. Prove that $T^{-1}\colon \mathrm{range}(T) \to H$ is unbounded.

8.6 Finite-Rank Operators

Recall that the *rank* of a linear operator $T\colon H \to K$ is the vector space dimension of its range, and T is a *finite-rank operator* if $\mathrm{range}(T)$ is finite dimensional. In general, a finite-rank operator need not be bounded (see Problem 6.3.6). We will denote the set of *bounded* finite-rank operators by

$$\mathcal{B}_{00}(H,K) \;=\; \bigl\{T \in \mathcal{B}(H,K) \;:\; T \text{ has finite rank}\bigr\},$$

and we let $\mathcal{B}_{00}(H) = \mathcal{B}_{00}(H,H)$.

Not every linear finite-rank operator is compact (see Example 8.6.3). However, we prove next that if a finite-rank linear operator is bounded, then it is compact.

Theorem 8.6.1. *If $T\colon H \to K$ is bounded, linear, and has finite rank, then T is compact. Hence*

$$\mathcal{B}_{00}(H,K) \;\subseteq\; \mathcal{B}_0(H,K). \tag{8.19}$$

Proof. Since T is bounded, $T(D_H)$ is a bounded subset of the finite-dimensional space $\mathrm{range}(T)$. All finite-dimensional normed spaces are complete, so $\mathrm{range}(T)$ is a Hilbert space. Hence $\overline{T(D_H)}$ is a closed and bounded subset of the *finite-dimensional* Hilbert space $\mathrm{range}(T)$. It follows from this that the closure of $T(D_H)$ is a compact set.

Combining Theorem 8.6.1 with Theorem 8.5.6, we obtain the following corollary.

Corollary 8.6.2. *If $T\colon H \to K$ is linear and there exist bounded linear finite-rank operators T_n such that $\|T - T_n\| \to 0$, then T is compact.* $\diamondsuit$

To illustrate, we will use Corollary 8.6.2 to prove that the multiplication operator M_λ is compact when the sequence λ belongs to c_0. This operator will not have finite rank if λ contains infinitely many nonzero components.

Example 8.6.3. Assume that $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for a Hilbert space H, and fix any sequence $\lambda = (\lambda_n)_{n\in\mathbb{N}} \in c_0$ (i.e., $\lambda_n \to 0$ as $n \to \infty$). Let

$$M_\lambda x \;=\; \sum_{n=1}^{\infty} \lambda_n \langle x, e_n\rangle\, e_n, \qquad x \in H, \tag{8.20}$$

be the "multiplication operator" from Theorem 8.1.4. Since $c_0 \subseteq \ell^\infty$, that theorem tells us that M_λ is bounded. We will use Corollary 8.6.2 to show that the extra assumption that λ_n converges to zero implies that M_λ is compact.

Note that M_λ need not have finite rank. We will define finite-rank operators T_N that "approximate" M_λ. We do this by replacing the infinite series in equation (8.20) by a finite sum. Specifically, given $N > 0$, we define $T_N\colon H \to H$ by

$$T_N x \;=\; \sum_{n=1}^{N} \lambda_n \langle x, e_n\rangle\, e_n, \qquad x \in H.$$

Since $T_N x$ is always a finite linear combination of $e_1, \dots, e_N$, we have

$$\mathrm{range}(T_N) \;\subseteq\; \mathrm{span}\{e_1, \dots, e_N\}.$$

Therefore T_N has finite rank. It is also linear, and since $|\lambda_n| \le \|\lambda\|_\infty$,

$$\|T_N x\|^2 \;=\; \sum_{n=1}^{N} |\lambda_n \langle x, e_n\rangle|^2 \;\le\; \|\lambda\|_\infty^2 \sum_{n=1}^{N} |\langle x, e_n\rangle|^2 \;\le\; \|\lambda\|_\infty^2\, \|x\|^2,$$

where the final inequality follows from Bessel's Inequality. This shows that T_N is bounded (in fact, it shows that $\|T_N\| \le \|\lambda\|_\infty$).

If we fix an $x \in H$, then $T_N x \to M_\lambda x$. However, we will prove more—we will prove that T_N converges *in operator norm* to M_λ. To do this, we compute that for any fixed $x \in H$ we have

$$\begin{aligned}
\|M_\lambda x - T_N x\|^2 \;&=\; \biggl\| \sum_{n=1}^{\infty} \lambda_n \langle x, e_n\rangle\, e_n \;-\; \sum_{n=1}^{N} \lambda_n \langle x, e_n\rangle\, e_n \biggr\|^2 \\
&=\; \biggl\| \sum_{n=N+1}^{\infty} \lambda_n \langle x, e_n\rangle\, e_n \biggr\|^2 \qquad \text{(because the series converges!)}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{n=N+1}^{\infty} |\lambda_n|^2 \, |\langle x, e_n\rangle|^2 \qquad \text{(Plancherel's Equality)} \\
&\le \left(\sup_{n>N} |\lambda_n|^2\right) \sum_{n=N+1}^{\infty} |\langle x, e_n\rangle|^2 \\
&\le \left(\sup_{n>N} |\lambda_n|^2\right) \|x\|^2 \qquad \text{(Bessel's Inequality).}
\end{aligned}
$$

Taking the supremum over all unit vectors x, we obtain

$$\|M_\lambda - T_N\| \;=\; \sup_{\|x\|=1} \|(M_\lambda - T_N)x\| \;\le\; \sup_{n>N} |\lambda_n|.$$

The reader should now check that our assumption that $\lambda_n \to 0$ implies that

$$\lim_{N\to\infty} \sup_{n>N} |\lambda_n| \;=\; 0.$$

Therefore $\|M_\lambda - T_N\| \to 0$ as $N \to \infty$. This shows that $T_N \to M_\lambda$ in operator norm, and consequently Corollary 8.6.2 implies that M_λ is compact. $\diamondsuit$

For example, consider $\lambda = (\frac{1}{n})_{n\in\mathbb{N}}$, i.e.,

$$M_\lambda x \;=\; \sum_{n=1}^{\infty} \frac{1}{n}\, \langle x, e_n\rangle\, e_n, \qquad x \in H.$$

Since $\lambda \in c_0$, the operator M_λ is compact. However, it does not have finite rank. In fact, every orthonormal basis vector e_k belongs to the range, because

$$M_\lambda e_k \;=\; \sum_{n=1}^{\infty} \frac{1}{n}\, \langle e_k, e_n\rangle\, e_n, \;=\; \frac{e_k}{k}.$$

Even so, Problem 8.1.9 shows that M_λ not surjective; instead, range(M_λ) is a dense, but proper, subspace of H.

Problems

8.6.4. Let H and K be Hilbert spaces, and let $T \in \mathcal{B}(H,K)$ be given. Show that T has finite rank if and only if there exist vectors $y_1, \dots, y_N \in H$ and $z_1, \dots, z_N \in K$ such that

$$Tx \;=\; \sum_{k=1}^{N} \langle x, y_k\rangle\, z_k, \qquad x \in H.$$

In case this holds, find T^* and show that T^* has finite rank.

8.6.5. Let $\{e_n\}_{n\in\mathbb{N}}$ be an orthonormal basis for a separable Hilbert space H. Let $\lambda = (\lambda_n)_{n\in\mathbb{N}}$ be any sequence in c_0 such that $\lambda_n \neq 0$ for every n. By Example 8.6.3, the corresponding multiplication operator $M_\lambda\colon H \to H$ is compact. Prove that $M_\lambda(D_H)$ is not a closed set in H.

8.6.6. Let $A = [a_{ij}]_{i,j\in\mathbb{N}}$ be an infinite matrix such that

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}|a_{ij}|^2 \;<\; \infty.$$

Given $x \in \ell^2$, let Ax be defined as in equation (6.13). Let A_N be the matrix that has the same entries as A for $i, j \leq N$, but is zero elsewhere. Prove the following statements.

(a) $A_N\colon \ell^2 \to \ell^2$ is bounded and has finite rank.

(b) $A\colon \ell^2 \to \ell^2$ is bounded, and $A_N \to A$ in operator norm.

(c) $A\colon \ell^2 \to \ell^2$ is compact.

8.7 Eigenvalues and Invariant Subspaces

Let X be a vector space, and suppose that $A\colon X \to X$ is linear. We say that a scalar $\lambda \in \mathbb{F}$ is an *eigenvalue* of A if there exists a nonzero vector $x \in X$ such that $Ax = \lambda x$. Any such nonzero vector x is called an *eigenvector* of A. To emphasize the eigenvalue that corresponds to an eigenvector x, we sometimes say that x is a λ-*eigenvector* of A.

Restating the definition,

$$\lambda \text{ is an eigenvalue of } A \quad\iff\quad \ker(A-\lambda I) \;\neq\; \{0\}.$$

In this case we call $\ker(A-\lambda I)$ the *eigenspace* corresponding to λ, or simply the λ-*eigenspace* for short. If $\ker(A-\lambda I) \neq \{0\}$, then every nonzero vector in $\ker(A-\lambda I)$ is a λ-eigenvector for A (in particular, every nonzero vector in $\ker(A)$, if there are any, is a 0-eigenvector of A). The (geometric) *multiplicity* of an eigenvalue λ is the dimension of its λ-eigenspace $\ker(T-\lambda I)$.

If X is a normed space, then we have the following inequality relating the magnitudes of any eigenvalues of A to its operator norm.

Lemma 8.7.1. *Let X be a normed space, and let $A\colon X \to X$ be a bounded linear operator. If λ is an eigenvalue of A, then $|\lambda| \leq \|A\|$.*

Proof. Suppose that $Ax = \lambda x$, where $x \neq 0$. If we set $y = x/\|x\|$, then y is a unit vector, and $Ay = \lambda y$. Therefore

$$\|Ay\| \;=\; \|\lambda y\| \;=\; |\lambda|\,\|y\| \;=\; |\lambda|.$$

Since y is only one of the unit vectors in X, it follows that

$$\|A\| \;=\; \sup_{\|z\|=1} \|Az\| \;\geq\; \|Ay\| \;=\; |\lambda|. \qquad \square$$

We introduce a related type of subspace.

Definition 8.7.2. Let X be a vector space. A subspace M of X is said to be *invariant* under a linear operator $A\colon X \to X$ if $A(M) \subseteq M$, where $A(M) = \{Ax : x \in M\}$ is the direct image of M under A. $\diamondsuit$

If λ is an eigenvalue of A, then its eigenspace $M = \ker(A-\lambda I)$ is invariant under A (why?). Hence eigenspaces are special cases of invariant subspaces.

For Hilbert spaces, we have the following relationship between subspaces that are invariant under a bounded operator and those that are invariant under its adjoint (the proof is assigned as Problem 8.7.6).

Lemma 8.7.3. *Let H be a Hilbert space, and let $A \in \mathcal{B}(H)$ be a bounded linear operator. If M is a closed subspace of H that is invariant under A, then $M^\perp$ is invariant under A^*.* $\diamondsuit$

We will also need the following fact about eigenvalues and eigenvectors of a self-adjoint operator.

Lemma 8.7.4. *If H is a Hilbert space and $A \in \mathcal{B}(H)$ is self-adjoint, then all eigenvalues of A are real, and eigenvectors of A corresponding to distinct eigenvalues are orthogonal.*

Proof. Suppose that λ is an eigenvalue of A, and let x be a corresponding eigenvector. Then, since $A = A^*$,

$$\lambda\,\|x\|^2 \;=\; \langle \lambda x, x\rangle \;=\; \langle Ax, x\rangle \;=\; \langle x, Ax\rangle \;=\; \langle x, \lambda x\rangle \;=\; \overline{\lambda}\,\|x\|^2.$$

Since $x \neq 0$, this implies that $\lambda = \overline{\lambda}$, and therefore λ is real.

Now suppose that $\lambda \neq \mu$ are two different eigenvalues of A, and x is a λ-eigenvector while y is a μ-eigenvector. Since the eigenvalues are real, it follows that

$$\lambda\,\langle x, y\rangle \;=\; \langle \lambda x, y\rangle \;=\; \langle Ax, y\rangle \;=\; \langle x, Ay\rangle \;=\; \langle x, \mu y\rangle \;=\; \mu\,\langle x, y\rangle.$$

But $\lambda \neq \mu$, so we must have $\langle x, y\rangle = 0$. $\square$

Turning to compact operators, the following result shows that the multiplicity of a nonzero eigenvalue of a compact operator must be finite.

Lemma 8.7.5. *Let H be a Hilbert space. If λ is a nonzero eigenvalue of a compact operator $T\colon H \to H$, then* $\dim(\ker(T-\lambda I)) < \infty$.

Proof. Let $M = \ker(T - \lambda I)$. Since T is compact, it is bounded and therefore continuous. This implies that M is a closed subspace of H (see Problem 6.4.2). If M was infinite-dimensional, then would contain an infinite orthonormal sequence $\{e_n\}_{n\in\mathbb{N}}$. By the definition of M we have $Te_n = \lambda e_n$. But then, for any $m \neq n$,

$$\|Te_m - Te_n\| \;=\; \|\lambda e_m - \lambda e_n\| \;=\; |\lambda|\,\|e_m - e_n\| \;=\; |\lambda|\,\sqrt{2},$$

the last equality following from the orthonormality of the e_n. Since $\lambda \neq 0$, this implies that $\{Te_n\}_{n\in\mathbb{N}}$ contains no convergent subsequences. This contradicts the fact that T is compact. □

Although the multiplicity of a nonzero eigenvalue of a compact operator is finite, the multiplicity of a 0-eigenvalue can be infinite. For example, the zero operator on H is compact, yet the zero eigenspace is all of H.

Problems

8.7.6. Prove Lemma 8.7.3.

8.7.7. Let $L,\ R\colon \ell^2 \to \ell^2$ be the left-shift and right-shift operators from Examples 6.6.3 and 6.6.4.

(a) Find all of the eigenvalues and eigenvectors of L and R.

(b) Compute LR and RL and show that $LR \neq RL$. (In contrast, if A and B are $n \times n$ matrices, then $AB = I$ if and only if $BA = I$.)

(c) Show that L and R are not compact operators.

(d) Exhibit two bounded but noncompact operators $A,\ B\colon \ell^2 \to \ell^2$ such that $AB = 0$ (thus it is possible for the product of noncompact operators to be compact).

8.7.8. Let $A \in \mathcal{B}(H)$ be a positive operator on a Hilbert space H. Show that all eigenvalues of A are real and nonnegative, and if A is positive definite then all eigenvalues are real and strictly positive.

8.7.9. Show that if $U\colon H \to H$ is a unitary operator on a Hilbert space H, then $U^* = U^{-1}$, U is normal, and any eigenvalue λ of U satisfies $|\lambda| = 1$.

8.7.10. Assume $A \in \mathcal{B}(H)$ is normal, and prove the following statements.

(a) If λ is an eigenvalue of A, then $\overline{\lambda}$ is an eigenvalue of A^*.

(b) The λ-eigenspace of A is the $\overline{\lambda}$-eigenspace of A^*, i.e., $\ker(A - \lambda I) = \ker(A^* - \overline{\lambda} I)$.

(c) Eigenvectors of A corresponding to distinct eigenvalues are orthogonal.

8.7.11. Let H be a Hilbert space, and assume that $T \in \mathcal{B}(H)$ is both compact and self-adjoint. Suppose that M is a closed subspace of H that is invariant under T. Let $S\colon M \to M$ be the restriction of T to M, i.e., $Sx = Tx$ for $x \in M$. Prove that S is both compact and self-adjoint on M.

Remark: The notation $S = T|_M$ is often used to indicate that S is the restriction of T to the domain M.

8.8 Existence of an Eigenvalue

If H is a finite-dimensional Hilbert space and $A \in \mathcal{B}(H)$, then (by choosing a basis for H) we can identity A with an $n \times n$ matrix. Since every $n \times n$ matrix has at least one complex eigenvalue, we know that A must have at least one eigenvalue if $\mathbb{F} = \mathbb{C}$. In contrast, even if $\mathbb{F} = \mathbb{C}$, a linear operator on an infinite-dimensional Hilbert space need not have any eigenvalues. For example, if we take $H = \ell^2$ and let $Rx = (0, x_1, x_2, \dots)$ be the *right-shift operator* defined in Example 6.6.4, then R has no eigenvalues at all (see Problem 8.7.7).

The right-shift operator is not compact. However, compactness by itself is not enough to ensure that an operator will have an eigenvalue. We will prove that an operator that is both compact *and* self-adjoint must have at least one eigenvalue. In order to do this, we need the following sufficient condition for the existence of an eigenvalue of a compact operator.

Lemma 8.8.1. *If $T\colon H \to H$ is compact and $\lambda \neq 0$, then*

$$\inf_{\|x\|=1} \|Tx - \lambda x\| \;=\; 0 \qquad \Longrightarrow \qquad \lambda \text{ is an eigenvalue of } T.$$

Proof. Suppose that $\inf_{\|x\|=1} \|Tx - \lambda x\| = 0$. Then there exist unit vectors $x_n \in H$ such that $\|Tx_n - \lambda x_n\| \to 0$. Since T is compact, $\{Tx_n\}_{n\in\mathbb{N}}$ has a convergent subsequence. That is, there exist indices n_k and a vector $y \in H$ such that $Tx_{n_k} \to y$ as $k \to \infty$. Therefore

$$\lambda x_{n_k} \;=\; \bigl(\lambda x_{n_k} - Tx_{n_k}\bigr) + Tx_{n_k} \;\to\; 0 + y \;=\; y \quad \text{as } k \to \infty. \tag{8.21}$$

Applying the continuity of T, it follows that

$$\lambda Tx_{n_k} \;\to\; Ty \quad \text{as } k \to \infty.$$

On the other hand, since $Tx_{n_k} \to y$ we also have

$$\lambda Tx_{n_k} \;\to\; \lambda y \quad \text{as } k \to \infty.$$

Therefore $Ty = \lambda y$ by the uniqueness of limits. Further, by applying the continuity of the norm to equation (8.21), we see that

$$\|y\| = \lim_{k\to\infty} \|\lambda x_{n_k}\| = |\lambda| \neq 0.$$

Thus we have both $Ty = \lambda y$ and $y \neq 0$, so y is a λ-eigenvector for T. □

Now we will show that compactness combined with self-adjointness implies the existence of an eigenvalue. By Lemmas 8.7.1 and 8.7.4 we know that the eigenvalues of a self-adjoint operator T are real and are bounded in absolute value by the operator norm of T. Therefore any eigenvalue λ of a self-adjoint operator T must be a real scalar in the range $-\|T\| \leq \lambda \leq \|T\|$. We will show that if T is both compact and self-adjoint then one of these extremes must be an eigenvalue.

Theorem 8.8.2. *If $T\colon H \to H$ is compact and self-adjoint, then at least one of $\|T\|$ or $-\|T\|$ is an eigenvalue of T.*

Proof. Since T is self-adjoint, Theorem 8.3.8 tells us that

$$\|T\| = \sup_{\|x\|=1} |\langle Tx, x\rangle|.$$

Consequently, there exist unit vectors x_n such that $|\langle Tx_n, x_n\rangle| \to \|T\|$. Since T is self-adjoint, each inner product $\langle Tx_n, x_n\rangle$ is real, so we can find a subsequence $\{\langle Ty_n, y_n\rangle\}_{n\in\mathbb{N}}$ that converges to one of $\|T\|$ or $-\|T\|$. Let λ be either $\|T\|$ or $-\|T\|$, as appropriate. Then $\|y_n\| = 1$ for every n and $\langle Ty_n, y_n\rangle \to \lambda$. Since both λ and $\langle Ty_n, y_n\rangle$ are real, we therefore have

$$\begin{aligned}
\|Ty_n - \lambda y_n\|^2 &= \|Ty_n\|^2 - 2\lambda \langle Ty_n, y_n\rangle + \lambda^2 \|y_n\|^2 \\
&\leq \|T\|^2 \|y_n\|^2 - 2\lambda \langle Ty_n, y_n\rangle + \lambda^2 \|y_n\|^2 \\
&= \lambda^2 - 2\lambda \langle Ty_n, y_n\rangle + \lambda^2 \\
&\to \lambda^2 - 2\lambda^2 + \lambda^2 = 0.
\end{aligned}$$

Consequently, Lemma 8.8.1 implies that λ is an eigenvalue of T. □

The *spectrum* of an operator T is the set of all scalars λ such that $T - \lambda I$ is not invertible. In particular, each eigenvalue of T is in the spectrum, and the set of eigenvalues is called the *point spectrum* of T. For more details on the spectrum of operators, see [Con90] or [Rud91].

Problems

8.8.3. Let $T\colon H \to H$ be compact and fix $\lambda \neq 0$. Use Lemma 8.8.1 and Problem 6.2.10 to show that if λ is not an eigenvalue of T and $\overline{\lambda}$ is not an eigenvalue of T^*, then $T - \lambda I$ is a topological isomorphism.

8.9 The Spectral Theorem for Compact Self-Adjoint Operators

The Spectral Theorem provides a fundamental decomposition for self-adjoint operators on a Hilbert space. The version of the Spectral Theorem that we will present asserts that if T is a *compact* self-adjoint operator on H, then there is an orthonormal sequence in H that consists of eigenvectors of T, and T has a simple representation with respect to this orthonormal sequence. This is a basic version of the Spectral Theorem. The theorem generalizes to compact normal operators with little change (essentially, eigenvalues can be complex in the normal case, but must be real for self-adjoint operators). More deeply, there is a Spectral Theorem for bounded self-adjoint operators that are not compact, and even for some that are unbounded. We refer to texts on operator theory for such extensions, e.g., see [Con00], [Rud91].

We will obtain the Spectral Theorem by iterating Theorem 8.8.2. If T is the zero operator, then the index J in the following theorem is $J = \varnothing$ (and in this case we implicitly interpret a series of the form $\sum_{n\in\varnothing}$ as being zero). If $T \neq 0$ has finite rank, then the index set J will be $J = \{1, \dots, N\}$ where $N = \dim\bigl(\mathrm{range}(T)\bigr)$, and if $T \neq 0$ does not have finite rank then we will have $J = \mathbb{N}$. We will need to use the results of several problems in the proof, including Problems 8.5.10, 8.5.11, 8.7.11, and 8.9.5.

Theorem 8.9.1 (The Spectral Theorem for Compact Self-Adjoint Operators). *Let $T\colon H \to H$ be compact and self-adjoint. Then there exist:*

(a) *countably many nonzero real numbers $(\lambda_n)_{n\in J}$, either finitely many or satisfying $\lambda_n \to 0$ if infinitely many, and*

(b) *an orthonormal basis $\{e_n\}_{n\in J}$ of $\overline{\mathrm{range}}(T)$,*

such that

$$Tx \;=\; \sum_{n\in J} \lambda_n \langle x, e_n\rangle\, e_n, \qquad x \in H. \tag{8.22}$$

Each λ_n is an eigenvalue of T, and each e_n is a λ_n-eigenvector for T.

Proof. If $T = 0$ then we can simply take $J = \varnothing$, so assume that T is not the zero operator. Let

$$R \;=\; \overline{\mathrm{range}}(T)$$

be the closure of the range of T. Since R is a closed subspace of H, it is itself a Hilbert space. Further, because T is compact, Problem 8.5.10 implies that R is separable. We proved in Theorem 5.6.10 that every separable Hilbert space, including R in particular, has a countable orthonormal basis.

To begin the iteration process, set $H_1 = H$ and $T_1 = T$. Theorem 8.8.2 implies T_1 has a real eigenvalue λ_1 that satisfies $|\lambda_1| = \|T_1\| > 0$. Let e_1 be a corresponding eigenvector. Since eigenvectors are nonzero, we can rescale e_1 so that $\|e_1\| = 1$ (simply divide e_1 by its length).

Let $E_1 = \operatorname{span}\{e_1\}$. Since e_1 is an eigenvector of T, the one-dimensional subspace E_1 is invariant under T. Therefore, by Lemma 8.7.3, its orthogonal complement

$$H_2 \;=\; E_1^\perp \;=\; \operatorname{span}\{e_1\}^\perp \;=\; \{e_1\}^\perp$$

is invariant under $T^* = T$. Therefore we can create a new linear operator $T_2 \colon H_2 \to H_2$ by simply restricting the domain of T to the space H_2. That is, T_2 is defined by $T_2x = Tx$, but only for $x \in H_2$. By doing this we obtain an operator T_2 that maps H_2 into itself.

It is possible that T_2 might be the zero operator. If that is the case, then we set $J = \{1\}$ and stop the iteration. Otherwise, we apply Problem 8.7.11, which shows that $T_2 \colon H_2 \to H_2$ is both compact and self-adjoint as an operator on the Hilbert space H_2. Applying Theorem 8.8.2 to T_2, we see that T_2 has an eigenvalue λ_2 such that $|\lambda_2| = \|T_2\|$. Let $e_2 \in H_2$ be a corresponding eigenvector, normalized so that $\|e_2\| = 1$. Because $H_2 = \{e_1\}^\perp$, we have $e_2 \perp e_1$. Also, because $e_2 \in H_2$, we have $Te_2 = T_2e_2 = \lambda_2 e_2$, and therefore λ_2 is an eigenvalue of T (not just T_2). Further, because $T_2x = T_1x$ for all $x \in H_2$ (which is a subset of H_1), we also have

$$\begin{aligned}
|\lambda_2| \;=\; \|T_2\| \;&=\; \sup\bigl\{\|T_2x\| : \|x\| = 1, x \in H_2\bigr\} \\
&\le\; \sup\bigl\{\|T_1x\| : \|x\| = 1, x \in H_1\bigr\} \;=\; \|T_1\| \;=\; |\lambda_1|.
\end{aligned}$$

Now we set $E_2 = \operatorname{span}\{e_1, e_2\}$ and $H_3 = E_2^\perp = \{e_1, e_2\}^\perp$. This space is invariant under T (why?), so we can define an operator $T_3 \colon H_3 \to H_3$ by restricting T to H_3. If $T_3 = 0$, then we stop at this point. Otherwise, we continue as before to construct an eigenvalue λ_3 (which will satisfy $|\lambda_3| \le |\lambda_2|$) and a unit eigenvector e_3 (which will be orthogonal to both e_1 and e_2). As we continue in this process, there are two possibilities.

Case 1: $T_{N+1} = 0$ for some N. In this case the iteration has stopped after a finite number of steps. We have obtained eigenvalues $\lambda_1, \dots, \lambda_N$ and corresponding orthonormal eigenvectors $e_1, \dots, e_N$. We have defined the closed subspaces

$$E_N \;=\; \operatorname{span}\{e_1, \dots, e_N\} \qquad \text{and} \qquad H_{N+1} \;=\; E_N^\perp \;=\; \{e_1, \dots, e_N\}^\perp.$$

Since E_N and H_{N+1} are orthogonal complements in H, if we fix $x \in H$ then there exist unique vectors $p_x \in E_N$ and $q_x \in H_{N+1}$ such that $x = p_x + q_x$ (in fact, p_x is the orthogonal projection of x onto E_N, and $q_x = x - p_x$). Since $e_1, \dots, e_N$ is an orthonormal basis for E_N we know that p_x is given by the formula

$$p_x \;=\; \sum_{n=1}^{N} \langle x, e_n\rangle\, e_n.$$

Also, since $q_x \in H_{N+1}$,

$$Tq_x \;=\; T_{N+1}\, q_x \;=\; 0.$$

On the other hand, $Te_n = \lambda_n e_n$, so

$$Tx \;=\; Tp_x + Tq_x \;=\; \sum_{n=1}^{N} \langle x, e_n\rangle\, Te_n \;+\; 0 \;=\; \sum_{n=1}^{N} \lambda_n \langle x, e_n\rangle\, e_n. \tag{8.23}$$

This is true for every $x \in H$. Setting $J = \{1, \dots, N\}$, it follows that equation (8.22) holds.

By equation (8.23), we always have $Tx \in \operatorname{span}\{e_1, \dots, e_N\}$, so $\operatorname{range}(T) \subseteq \operatorname{span}\{e_1, \dots, e_N\}$. On the other hand, the fact that $\lambda_1, \dots, \lambda_N$ are all nonzero implies that $e_1, \dots, e_N$ all belong to the range of T. Consequently $\operatorname{range}(T) = \operatorname{span}\{e_1, \dots, e_N\}$. Therefore $\operatorname{range}(T)$ is closed (because it is finite dimensional) and $\{e_1, \dots, e_N\}$ is an orthonormal basis for $\operatorname{range}(T)$. Therefore the proof is complete for this case.

Case 2: $T_N \neq 0$ for any N. In this case the iteration never ends. We obtain eigenvalues $\lambda_1, \lambda_2, \dots$ (with $|\lambda_1| \ge |\lambda_2| \ge \cdots$), and corresponding orthonormal eigenvectors $e_1, e_2, \dots$. Because T is compact, Problem 8.5.11 implies that $Te_n \to 0$. Using the fact that e_n is an eigenvector and $\|e_n\| = 1$, we see that

$$|\lambda_n| \;=\; |\lambda_n|\, \|e_n\| \;=\; \|\lambda_n e_n\| \;=\; \|Te_n\| \;\to\; 0 \qquad \text{as } n \to \infty.$$

Let

$$M \;=\; \overline{\operatorname{span}}\{e_n\}_{n\in\mathbb{N}}.$$

If we choose $x \in H$, then there exist unique vectors $p_x \in M$ and $q_x \in M^\perp$ such that $x = p_x + q_x$. In fact, p_x is the orthogonal projection of x on to M. Therefore, since $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for M, p_x has the form

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

If we fix $N > 1$, then $E_{N-1} = \operatorname{span}\{e_1, \dots, e_{N-1}\} \subseteq M$. Therefore

$$M^\perp \;\subseteq\; E_{N-1}^\perp \;=\; H_N.$$

Since T_N is the restriction of T to T_N, it follows that

$$\|Tq_x\| \;=\; \|T_N q_x\| \;\le\; \|T_N\|\, \|q_x\| \;=\; |\lambda_N|\, \|q_x\| \;\to\; 0 \qquad \text{as } N \to \infty.$$

Therefore $Tq_x = 0$. Consequently,

$$Tx \;=\; Tp_x + Tq_x \;=\; \sum_{n=1}^{\infty} \langle x, e_n\rangle\, Te_n \;+\; 0 \;=\; \sum_{n=1}^{\infty} \lambda_n \langle x, e_n\rangle\, e_n.$$

This is true for every $x \in H$. Setting $J = \mathbb{N}$, it follows that equation (8.22) holds. It only remains to show that $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for $\overline{\operatorname{range}}(T)$, and we assign this proof as Problem 8.9.5. $\square$

We stated Theorem 8.9.1 in terms of the *nonzero* eigenvalues of T. The zero eigenspace is $\ker(T)$, and the vectors in this space are eigenvectors of T for the eigenvalue $\lambda = 0$. If H is separable, then $\ker(T)$ is separable. By combining an orthonormal basis for $\ker(T)$ with the orthonormal basis for $\overline{\operatorname{range}}(T)$ given by Theorem 8.9.1, we obtain the following reformulation and converse of the Spectral Theorem. This form states that every compact, self-adjoint operator on a separable Hilbert space is actually a multiplication operator M_λ of the form given in equation (8.3). We assign the proof of this result as Problem 8.9.6.

Corollary 8.9.2. *Let H be a separable Hilbert space. If H is finite dimensional then set $J = \{1, \dots, N\}$ where $N = \dim(H)$, otherwise let $J = \mathbb{N}$. If $T \in \mathcal{B}(H)$, then the following two statements are equivalent.*

(a) *T is compact and self-adjoint.*

(b) *There exists a real sequence $\lambda = (\lambda_n)_{n\in J}$, with $\lambda \in c_0$ if $J = \mathbb{N}$, and an orthonormal basis $\{e_n\}_{n\in J}$ for H such that*

$$Tx \;=\; M_\lambda x \;=\; \sum_{n\in J} \lambda_n \,\langle x, e_n\rangle\, e_n, \qquad x \in H. \qquad \diamondsuit$$

Thus, if T is a compact, self-adjoint operator on a *separable* Hilbert space, then there exists an orthonormal basis for H that consists entirely of eigenvectors of T. Problem 8.9.8 shows that if $\mathbb{F} = \mathbb{C}$ and $H = \mathbb{C}^n$, so T is identified with an $n \times n$ Hermitian matrix, then the existence of this orthonormal basis of eigenvectors implies that T is *diagonalizable*, i.e., we can write $T = U\Lambda U^{-1}$ where U is invertible (in fact, unitary) and Λ is diagonal.

8.9.1 The Square Root and Absolute Value Operators

As an application of the Spectral Theorem, the next exercise constructs a square root of a compact positive operator.

Exercise 8.9.3. Assume that $T\colon H \to H$ is both compact and positive, and let equation (8.22) be the representation of T given by the Spectral Theorem. Define

$$T^{1/2} f \;=\; \sum_{n\in J} \lambda_n^{1/2} \,\langle f, e_n\rangle\, e_n, \qquad f \in H.$$

Show that $T^{1/2}$ is a well-defined, compact, positive operator that satisfies $(T^{1/2})^2 = T$. $\diamondsuit$

More generally, it can be shown that every positive operator on a Hilbert space space has a square root (a proof can be found in [Heil11, Thm. 2.18]).

An arbitrary compact operator T need not have a square root, but the composition T^*T is both compact and positive, so Exercise 8.9.3 implies that T^*T has a square root. We have a special name for this operator.

Definition 8.9.4 (Absolute Value). If $T\colon H \to H$ is compact, then the positive compact operator

$$|T| \;=\; (T^*T)^{1/2}$$

is called the *absolute value* of T. $\diamondsuit$

It can be shown [Con90] that every bounded operator $A \in \mathcal{B}(H)$ can be written in the form $A = UP$ where P is positive and U is a *partial isometry* (meaning that $\|Uf\| = \|f\|$ for all $f \in (\ker U)^\perp$), and this *polar decomposition* is unique if we impose the condition that $\ker(U) = \ker(P)$. In particular, if A is compact then P coincides with the operator $|A|$ defined above. Thus, it is possible to define the absolute value of any bounded operator on H, not just the compact operators considered in Definition 8.9.4.

Problems

8.9.5. Complete the proof of Case 2 of Theorem 8.9.1 by showing that $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for $\overline{\mathrm{range}}(T)$.

8.9.6. Prove Corollary 8.9.2.

8.9.7. Let H be a Hilbert space, and assume that $T\colon H \to H$ is both compact and positive. Let equation (8.22) be the representation of T given by the Spectral Theorem. Define

$$T^{1/2}f \;=\; \sum_{n\in J} \lambda_n^{1/2}\, \langle f, e_n\rangle\, e_n, \qquad f \in H.$$

Show that $T^{1/2}$ is well-defined, compact, positive, and is a square root of T in the sense that $(T^{1/2})^2 = T$.

8.9.8. This problem presents the Spectral Theorem for operators on $\mathbb{F}^n$. Let A be an $n \times n$ matrix with scalar entries. Show that A is Hermitian if and only if $A = U\Lambda U^{-1}$ where U is unitary and Λ is a real diagonal matrix. In this case, prove that the diagonal entries of Λ are the eigenvalues of A, and the columns of U form an orthonormal basis for $\mathbb{F}^n$ consisting of eigenvectors of A.

8.9.9. The *spectral radius* $\rho(T)$ of a square $d \times d$ matrix T is

$$\rho(T) \;=\; \max\{|\lambda| : \lambda \text{ is an eigenvalue of } T\}.$$

Let A be an $m \times n$ matrix (note that A need not be square). Show directly that if we place the ℓ^2 norm on both $\mathbb{F}^n$ and $\mathbb{F}^m$, then the operator norm of A as a mapping of $\mathbb{F}^n$ into $\mathbb{F}^m$ is $\|A\| = \rho(A^*A)^{1/2}$. Compare this to the operator norms derived in Problem 6.3.5.

8.9.10. Let vectors $v_1, \dots, v_N \in \mathbb{F}^d$ be given. The *analysis operator* for these vectors is the matrix R that has $v_1, \dots, v_N$ as rows, and the *synthesis operator* is $C = R^{\mathrm{H}}$. The *frame operator* is $S = C^{\mathrm{H}}C \;=\; RR^{\mathrm{H}}$.

(a) Describe Rx, Cx, and Sx for $x \in \mathbb{F}^d$.

(b) Prove that S has an orthonormal basis of eigenvectors $\{w_1, \dots, w_d\}$ and corresponding nonnegative eigenvalues $0 \le \lambda_1 \le \cdots \le \lambda_d$.

(c) Given $x \in \mathbb{F}^d$, prove that

$$\sum_{k=1}^{N} |\langle x, v_k\rangle|^2 \;=\; \langle Sx, x\rangle \;=\; \sum_{j=1}^{d} \lambda_j\, |\langle x, w_j\rangle|^2.$$

(d) Prove that the following three conditions are equivalent:

i. $\operatorname{span}\{v_1, \dots, v_N\} = \mathbb{F}^d$,

ii. $\lambda_1 > 0$,

iii. there exist constants A, $B > 0$ such that

$$A\,\|x\|^2 \;\le\; \sum_{k=1}^{N} |\langle x, v_k\rangle|^2 \;\le\; B\,\|x\|^2, \qquad \text{all } x \in \mathbb{F}^d. \tag{8.24}$$

Remark: A set of vectors $\{v_1, \dots, v_N\}$ that satisfies equation (8.24) is called a *frame* for $\mathbb{F}^d$.

8.10 Hilbert–Schmidt Operators

Throughout this section, H will denote an **infinite-dimensional, separable** *Hilbert space.*

A Hilbert–Schmidt operator is a special type of compact operator on a Hilbert space. Although the definition of a Hilbert–Schmidt operator makes sense when H is nonseparable (by using a complete orthonormal system instead of a countable orthonormal basis), we will restrict our discussion to Hilbert–Schmidt operators on separable spaces. For simplicity of presentation we will further assume that H is infinite dimensional, and therefore has an orthonormal basis of the form $\{e_n\}_{n=1}^{\infty}$. All of the results we derive have analogues for finite-dimensional Hilbert space, but they are usually not as interesting since every linear operator that maps a finite-dimensional Hilbert space into itself is Hilbert–Schmidt.

8.10.1 Definition and Basic Properties

Suppose that H is an infinite-dimensional Hilbert space and $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for H. In this case $\sum \|e_n\|^2 = \infty$. A Hilbert–Schmidt operator will be required to perform some "compression" on H, in the sense that the image $\{Te_n\}_{n\in\mathbb{N}}$ of the orthonormal basis will need to satisfy the condition $\sum \|Te_n\|^2 < \infty$. Here is the formal definition.

Definition 8.10.1. An operator $T \in \mathcal{B}(H)$ is a *Hilbert–Schmidt operator* if there exists an orthonormal basis $\{e_n\}_{n\in\mathbb{N}}$ for H such that

$$\sum_{n=1}^{\infty} \|Te_n\|^2 < \infty. \qquad \diamondsuit$$

Here are some trivial examples.

Example 8.10.2. (a) The identity operator on an infinite-dimensional Hilbert space is not Hilbert–Schmidt.

(b) If we let H be finite-dimensional, then it has an orthonormal basis of the form $\{e_n\}_{n=1}^{d}$ where d is finite. In this case, a Hilbert–Schmidt operator is an operator T on H that satisfies $\sum_{n=1}^{d} \|Te_n\|^2 < \infty$. Since d is finite, every linear operator $T\colon H \to H$ satisfies this requirement, so every linear operator that maps H into itself is Hilbert–Schmidt (and also bounded and compact). $\diamondsuit$

The next exercise shows that the choice of orthonormal basis in Definition 8.10.1 is irrelevant.

Exercise 8.10.3. Fix $T \in \mathcal{B}(H)$. Given an orthonormal basis $\mathcal{E} = \{e_n\}_{n\in\mathbb{N}}$ for H, set

$$S(\mathcal{E}) = \left(\sum_{n=1}^{\infty} \|Te_n\|^2\right)^{1/2}.$$

Prove that $S(\mathcal{E})$ is independent of the choice of orthonormal basis $\mathcal{E}$. That is, if $S(\mathcal{E})$ is finite for one orthonormal basis then it is finite for all and always takes the same value, and if $S(\mathcal{E})$ is infinite for one orthonormal basis then it is infinite for all. $\diamondsuit$

We use the following notation to denote the space of Hilbert–Schmidt operators on H.

Definition 8.10.4. The space of *Hilbert–Schmidt operators* on H is

$$\mathcal{B}_2(H) \;=\; \bigl\{T \in \mathcal{B}(H) \;:\; T \text{ is Hilbert–Schmidt}\bigr\}.$$

The *Hilbert–Schmidt norm* of $T \in \mathcal{B}_2(H)$ is

$$\|T\|_{\mathcal{B}_2} \;=\; \biggl(\sum_{n=1}^{\infty} \|Te_n\|^2\biggr)^{1/2},$$

where $\{e_n\}_{n\in\mathbb{N}}$ is any orthonormal basis for H. $\diamondsuit$

Remark 8.10.5. Writing

$$\|T\|_{\mathcal{B}_2}^2 \;=\; \sum_{n=1}^{\infty} \|Te_n\|^2 \;=\; \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} |\langle Te_n, e_m\rangle|^2, \tag{8.25}$$

we see that the Hilbert–Schmidt norm can be viewed as the ℓ^2-norm of the matrix representation of T with respect to the orthonormal basis $\{e_n\}$. $\diamondsuit$

We have called $\|\cdot\|_{\mathcal{B}_2}$ a norm on $\mathcal{B}_2(H)$. The following exercise justifies this terminology, and also presents some of the other basic properties of Hilbert–Schmidt operators.

Exercise 8.10.6. Prove the following statements.

(a) The Hilbert–Schmidt norm dominates the operator norm, i.e.,

$$\|T\| \;\le\; \|T\|_{\mathcal{B}_2} \quad \text{for all } T \in \mathcal{B}_2(H).$$

(b) $\|\cdot\|_{\mathcal{B}_2}$ is a norm on $\mathcal{B}_2(H)$, and $\mathcal{B}_2(H)$ is complete with respect to this norm.

(c) $\mathcal{B}_2(H)$ is closed under adjoints, and $\|T^*\|_{\mathcal{B}_2} = \|T\|_{\mathcal{B}_2}$ for all $T \in \mathcal{B}_2(H)$.

(d) If $T \in \mathcal{B}_2(H)$ and $A \in \mathcal{B}(H)$, then AT, $TA \in \mathcal{B}_2(H)$, and their operator norms satisfy

$$\|AT\|_{\mathcal{B}_2} \;\le\; \|A\|\,\|T\|_{\mathcal{B}_2}, \qquad \|TA\|_{\mathcal{B}_2} \;\le\; \|A\|\,\|T\|_{\mathcal{B}_2}.$$

Consequently, $\mathcal{B}_2(H)$ is a two-sided ideal in $\mathcal{B}(H)$.

(e) Every bounded, finite-rank linear operator on H is Hilbert–Schmidt.

(f) Every Hilbert–Schmidt operator on H is compact.

(g) The space $\mathcal{B}_{00}(H)$ of finite-rank operators is dense in $\mathcal{B}_2(H)$ with respect to both the operator norm $\|\cdot\|$ and the Hilbert–Schmidt norm $\|\cdot\|_{\mathcal{B}_2}$. $\diamondsuit$

Combining statements (e) and (f) from the preceding exercise, we see that

$$\mathcal{B}_{00}(H) \;\subseteq\; \mathcal{B}_2(H) \;\subseteq\; \mathcal{B}_0(H) \;\subseteq\; \mathcal{B}(H). \tag{8.26}$$

For a finite dimensional space, all of these spaces are equal. However, for an infinite-dimensional space, each of the inclusions in equation (8.26) is proper (see Problem 8.10.15).

8.10.2 Singular Numbers and the Hilbert–Schmidt Norm

In general, a compact operator T on H need not have any eigenvalues. However, if we compose T with its adjoint, we obtain the self-adjoint operator T^*T. Since T^*T is also compact, the Spectral Theorem implies that T^*T has eigenvalues. In fact, Theorem 8.9.1 implies that there exist a sequence of eigenvalues $(\mu_n)_{n\in\mathbb{N}}$ and an orthonormal sequence of eigenvectors $\{e_n\}_{n\in\mathbb{N}}$ such that T^*T has the form

$$T^*Tf \;=\; \sum_{n\in J} \mu_n \langle f, e_n\rangle\, e_n, \qquad f \in H.$$

The eigenvalues μ_n are real, nonzero, and converge to zero. In fact, since T^*T is a positive operator, we must have $\mu_n \geq 0$. Since μ_n is nonzero, this implies that $\mu_n > 0$ for every n. We introduce some terminology for these eigenvalues.

Definition 8.10.7 (Singular Numbers). Let $T\colon H \to H$ be compact, and let $(\mu_n)_{n\in\mathbb{N}}$ and $\{e_n\}_{n\in J}$ be as constructed above. The *singular numbers* or *singular values* of T are

$$s_n(T) \;=\; \mu_n^{1/2}, \qquad n \in \mathbb{N},$$

taken in decreasing order:

$$s_1(T) \;\geq\; s_2(T) \;\geq\; \cdots \;>\; 0.$$

The vectors e_n are corresponding *singular vectors* of T. $\diamondsuit$

The next exercise shows that the singular numbers of a self-adjoint operator can be written directly in terms of the eigenvalues of the operator.

If T is self-adjoint, then $T^*T = T^2$. In this case the eigenvalues of T^*T are simply the squares of the eigenvalues of T. The singular numbers are the square roots of these eigenvalues, so this gives us the following relation between the eigenvalues and singular numbers of a compact, self-adjoint operator.

Lemma 8.10.8. *Let T be a compact, self-adjoint operator on H, and let $(\lambda_n(T))_{n\in\mathbb{N}}$ be its nonzero eigenvalues taken Then the singular numbers of T are*

$$s_n(T) \;=\; |\lambda_n(T)|, \qquad n \in \mathbb{N}.$$

In particular, if T is positive, then $s_n(T) = \lambda_n(T)$ for every n. $\diamondsuit$

Now we reformulate the definition of Hilbert–Schmidt operators in terms of singular numbers.

Exercise 8.10.9. Let $T\colon H \to H$ be compact, and let $s = (s_n(T))_{n\in\mathbb{N}}$ be the sequence of singular numbers of T. Prove that T is Hilbert–Schmidt if and only if $s \in \ell^2$. Furthermore, in this case the Hilbert–Schmidt norm of T is precisely the ℓ^2-norm of the singular numbers:

$$\|T\|_{\mathcal{B}_2} \;=\; \|s\|_2 \;=\; \biggl(\sum_{n=1}^{\infty} s_n(T)^2\biggr)^{1/2}.$$

In particular, if T is a self-adjoint Hilbert–Schmidt operator, then

$$\|T\|_{\mathcal{B}_2} \;=\; \biggl(\sum_{n=1}^{\infty} \lambda_n(T)^2\biggr)^{1/2}$$

where $(\lambda_n(T))_{n\in\mathbb{N}}$ is the sequence of nonzero eigenvalues of T. $\diamondsuit$

8.10.3 Integral Operators with Square-Integrable Kernels

Theorem 6.8.7 showed that integral operators whose kernels are square integrable are bounded on $L^2(\mathbb{R})$. Our next goal is to show that all such integral operators are compact. In order to do this, we first need to construct a convenient orthonormal basis for $L^2(\mathbb{R}^2)$. This construction will also be used later in the proof of Theorem 8.10.12.

Exercise 8.10.10. Let $\{e_n\}_{n\in\mathbb{N}}$ be any orthonormal basis for $L^2(\mathbb{R})$, and define

$$e_{mn}(x,y) \;=\; (e_m \otimes e_n)(x,y) \;=\; e_m(x)\,\overline{e_n(y)}, \qquad x,y \in \mathbb{R}.$$

Prove that $\{e_{mn}\}_{m,n\in\mathbb{N}}$ is an orthonormal basis for $L^2(\mathbb{R}^2)$. $\diamondsuit$

Now we give an exercise that will prove that if $k \in L^2(\mathbb{R}^2)$, then the corresponding integral operator L_k is compact. In Section 8.10, we will introduce a class of compact operators that we call *Hilbert–Schmidt operators*, and in Theorem 8.10.12 we will prove that an integral operator is Hilbert–Schmidt if and only if it has a square-integrable kernel. Therefore we will refer to the integrable operators considered in the following exercise as *Hilbert–Schmidt integral operators*.

Exercise 8.10.11 (Hilbert–Schmidt Integral Operators). Fix a function $k \in L^2(\mathbb{R}^2)$, and let L_k be the integral operator with kernel k.

(a) Let $\{e_{mn}\}_{m,n\in\mathbb{N}}$ be an orthonormal basis for $L^2(\mathbb{R}^2)$ of the type constructed in Exercise 8.10.10. Define

$$k_N = \sum_{m=1}^{N}\sum_{n=1}^{N} \langle k, e_{mn}\rangle\, e_{mn},$$

and show that the corresponding integral operator L_{k_N} is bounded and has finite rank.

(b) Show that $L_k - L_{k_N}$ is the integral operator whose kernel is $k - k_N$, and its operator norm satisfies

$$\|L_k - L_{k_N}\| \le \|k - k_N\|_2.$$

Use this to prove that L_k is a compact mapping on $L^2(\mathbb{R})$. $\diamondsuit$

8.10.4 Hilbert–Schmidt Integral Operators

Now we specialize to integral operators on $L^2(\mathbb{R})$. Theorem 6.8.7 showed that if $k \in L^2(\mathbb{R}^2)$, then the corresponding integral operator L_k is a bounded operator on $L^2(\mathbb{R})$, and Exercise 8.10.11 showed further that L_k is compact. According to the next result, L_k is Hilbert–Schmidt when $k \in L^2(\mathbb{R}^2)$, and conversely each Hilbert–Schmidt operator on $L^2(\mathbb{R})$ can be written as an integral operators whose kernel is square integrable. As a consequence, this theorem shows that the space of Hilbert–Schmidt operators on $L^2(\mathbb{R})$ is isometrically isomorphic to $L^2(\mathbb{R}^2)$, i.e.,

$$\mathcal{B}_2(L^2(\mathbb{R})) \cong L^2(\mathbb{R}^2).$$

Theorem 8.10.12 (Hilbert–Schmidt Kernel Theorem).

(a) *If $k \in L^2(\mathbb{R}^2)$, then the integral operator L_k with kernel k is Hilbert–Schmidt, and its operator norm is*

$$\|L_k\|_{\mathcal{B}_2} = \|k\|_2.$$

(b) *If T is a Hilbert–Schmidt operator on $L^2(\mathbb{R})$, then there exists a function $k \in L^2(\mathbb{R}^2)$ such that $T = L_k$.*

(c) *The mapping $k \mapsto L_k$ is an isometric isomorphism of $L^2(\mathbb{R}^2)$ onto $\mathcal{B}_2(L^2(\mathbb{R}))$.*

Proof. (a) Assume that $k \in L^2(\mathbb{R}^2)$. We already know that L_k is a compact operator, so our task is to show that L_k is Hilbert–Schmidt. Let $\{e_n\}_{n\in\mathbb{N}}$ be any orthonormal basis for $L^2(\mathbb{R})$. If we set

$$e_{mn}(x,y) \;=\; (e_m \otimes e_n)(x,y) \;=\; e_m(x)\,\overline{e_n(y)}, \tag{8.27}$$

then $\{e_{mn}\}_{m,n\in\mathbb{N}}$ is an orthonormal basis for $L^2(\mathbb{R}^2)$.

Fix any integers m, $n \in \mathbb{N}$. Since both k and e_{mn} are square-integrable, their product $k \cdot \overline{e_{mn}}$ belongs to $L^1(\mathbb{R}^2)$. Consequently, Fubini's Theorem allows us to interchange integrals in the following calculation of the inner product of k with e_{mn}:

$$\begin{aligned}
\langle k, e_{mn}\rangle \;&=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} k(x,y)\,\overline{e_m(x)}\,e_n(y)\,dx\,dy \\
&=\; \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} k(x,y)\,e_n(y)\,dy\right)\overline{e_m(x)}\,dx \\
&=\; \int_{-\infty}^{\infty} L_k e_n(x)\,\overline{e_m(x)}\,dx \\
&=\; \langle L_k e_n, e_m\rangle.
\end{aligned}$$

Consequently, the Hilbert–Schmidt norm of L_k is

$$\begin{aligned}
\|L_k\|_{\mathcal{B}_2}^2 \;&=\; \sum_{n=1}^{\infty} \|L_k e_n\|^2 \\
&=\; \sum_{n=1}^{\infty}\left(\sum_{m=1}^{\infty} |\langle L_k e_n, e_m\rangle|^2\right) \\
&=\; \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} |\langle k, e_{mn}\rangle|^2 \\
&=\; \|k\|_2^2 \;<\; \infty.
\end{aligned}$$

Hence L_k is Hilbert–Schmidt, and furthermore $\|L_k\|_{\mathcal{B}_2} = \|k\|_2$.

(b) Let T be any Hilbert–Schmidt operator on $L^2(\mathbb{R})$. Choose an orthonormal basis $\{e_n\}_{n\in\mathbb{N}}$ for $L^2(\mathbb{R})$, and let e_{mn} be defined as in equation (8.27). Equation (8.25) shows that

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} |\langle Te_m, e_n\rangle|^2 \;<\; \infty.$$

Since $\{e_{mn}\}_{m,n\in\mathbb{N}}$ is an orthonormal basis for $L^2(\mathbb{R}^2)$, we can therefore define a function $k \in L^2(\mathbb{R}^2)$ by

$$k \;=\; \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \langle Te_n, e_m \rangle \, e_{mn}. \tag{8.28}$$

The series on the preceding line converges unconditionally in $L^2(\mathbb{R}^2)$, i.e., it converges regardless of what ordering we impose on the index set $\mathbb{N} \times \mathbb{N}$. Let L_k be the integral operator whose kernel is k. By part (a), L_k is Hilbert–Schmidt and we have $\|L_k\|_{\mathcal{B}_2} = \|k\|_2$. Our goal is to show that $L_k = T$.

Part (a) actually tells us that the mapping $k \mapsto L_k$ is an isometric mapping of $L^2(\mathbb{R})$ into $\mathcal{B}_2(H)$. Continuous mappings preserve convergence of series, so by applying this fact to equation (8.29) we see that

$$L_k \;=\; \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \langle Te_n, e_m \rangle \, L_{e_{mn}}, \tag{8.29}$$

where this series converges in the norm of $\mathcal{B}_2(H)$. Since the Hilbert–Schmidt norm dominates the operator norm, the series in equation (8.29) also converges with respect to the operator norm. As a consequence, for each function $f \in L^2(\mathbb{R})$ we have

$$L_k f \;=\; \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \langle Te_n, e_m \rangle \, L_{e_{mn}} f, \tag{8.30}$$

where this series converges in L^2-norm.

Actually, we can say more about the convergence of the series in equations (8.29) and (8.30). The series for k given in equation (8.28) converges *unconditionally*, i.e., it converges regardless of how we order the index set $\mathbb{N} \times \mathbb{N}$. Since convergence is preserved by continuous maps, it follows that the series in equation (8.29) also converges unconditionally (see Problem 6.5.9). From this it follows that the series in equation (8.30) converges unconditionally in L^2-norm for each individual function $f \in L^2(\mathbb{R})$.

Now, from Example 6.8.4 we know that $L_{e_{mn}}$ is the rank-one operator whose rule is $L_{e_{mn}} f = \langle f, e_n \rangle \, e_m$. Substituting this into equation (8.30) and using the unconditionality of the convergence to reorder the summations, we compute that

$$\begin{aligned}
L_k f \;&=\; \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \langle Te_n, e_m \rangle \, L_{e_{mn}} f, \\
&=\; \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \langle Te_n, e_m \rangle \, \langle f, e_n \rangle \, e_m \\
&=\; \sum_{n=1}^{\infty} \langle f, e_n \rangle \left(\sum_{m=1}^{\infty} \langle Te_n, e_m \rangle \, e_m \right)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{n=1}^{\infty} \langle f, e_n \rangle \, Te_n \\
&= T\biggl(\sum_{n=1}^{\infty} \langle f, e_n \rangle \, e_n\biggr) \qquad \text{(since } T \text{ is continuous)} \\
&= Tf.
\end{aligned}$$

Thus $L_k f = Tf$ for every f, so $T = L_k$.

(c) This is an immediate consequence of parts (a) and (b). □

In summary, Theorem 8.10.12 says that the space $\mathcal{B}_2(L^2(\mathbb{R}))$ is isometrically isomorphic to the Hilbert space $L^2(\mathbb{R}^2)$. Therefore the space of Hilbert–Schmidt operators inherits a Hilbert space structure from $L^2(\mathbb{R}^2)$. In terms of kernels, the inner product on $\mathcal{B}_2(L^2(\mathbb{R}))$ is

$$\langle L_k, L_h \rangle \;=\; \langle k, h \rangle,$$

With respect to this inner product, the mapping $k \mapsto L_k$ is a unitary. Since $L^2(\mathbb{R}^2)$ is separable, it follows that $\mathcal{B}_2(L^2(\mathbb{R}))$ is also separable.

As in the proof of Theorem 8.10.12, let $\{e_{mn}\}_{m,n\in\mathbb{N}}$ be an orthonormal basis for $L^2(\mathbb{R}^2)$ that has the form $e_{mn} = e_m \otimes e_n$. Since a unitary mapping sends an orthonormal basis to an orthonormal basis, it follows that

$$\bigl\{L_{e_{mn}}\bigr\}_{m,n\in\mathbb{N}}.$$

is an orthonormal basis for $\mathcal{B}_2(L^2(\mathbb{R}))$. By Example 6.8.4, $L_{e_{mn}}$ is a bounded, rank-one operator. Following Notation 6.8.5, we often identify $L_{e_{mn}}$ with the function $e_m \otimes e_n$ and write $L_{e_{mn}} = e_m \otimes e_n$. In this notation, $\{e_m \otimes e_n\}_{m,n\in\mathbb{N}}$ is an orthonormal basis for $\mathcal{B}_2(L^2(\mathbb{R}))$ consisting of bounded rank-one operators. The orthonormal basis representation of a Hilbert–Schmidt operator T as a superposition of these rank-one operators is

$$T \;=\; \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \langle Te_n, e_m \rangle \, (e_m \otimes e_n).$$

This series converges unconditionally with respect to the Hilbert–Schmidt norm, and hence also converges unconditionally with respect to the operator norm.

8.10.5 The Singular Value Decomposition

As in the discussion leading up to Corollary 8.9.2, it is often convenient to extend the sequence of singular vectors so that we obtain an orthonormal basis for all of H. The singular vectors $\{e_n\}_{n\in J}$, as defined above, form an orthonormal basis for $\overline{\text{range}}(T)$. Since T^*T is self-adjoint,

$$\overline{\text{range}}(T)^\perp \;=\; \ker(T^*T).$$

Hence we can form an orthonormal basis for all of H by combining $\{e_n\}_{n\in J}$ with an orthonormal basis $\{f_m\}_{m\in I}$ for $\ker(T^*T)$. We can regard the vectors f_m as being singular vectors of T corresponding to zero singular numbers. Each vector $f \in H$ can be written

$$f \;=\; \sum_{n\in J} \langle f, e_n\rangle\, e_n \;+\; \sum_{n\in I} \langle f, f_m\rangle\, f_m,$$

and we have

$$\begin{aligned} T^*Tf \;&=\; \sum_{n\in J} \langle f, e_n\rangle\, T^*Te_n \;+\; \sum_{n\in I} \langle f, f_m\rangle\, T^*Tf_m \\ &=\; \sum_{n\in J} s_n(T)^2\, \langle f, e_n\rangle\, e_n. \end{aligned}$$

Problems

8.10.13. Given $g, h \in H$, show that the Hilbert–Schmidt norm of the tensor-product operator $g \otimes h$ is $\|g \otimes h\|_{\mathcal{B}_2} = \|g\|\,\|h\|$.

8.10.14. Prove that the Hilbert–Schmidt norm is invariant under multiplication by unitary operators, i.e., if $T \in \mathcal{B}_2(H)$ is Hilbert–Schmidt and $U \in \mathcal{B}(H)$ is unitary, then $\|TU\|_{\mathcal{B}_2} = \|T\|_{\mathcal{B}_2} = \|UT\|_{\mathcal{B}_2}$.

8.10.15. Show that if H is infinite-dimensional, then $\mathcal{B}_{00}(H) \ne \mathcal{B}_2(H)$ and $\mathcal{B}_2(H) \ne \mathcal{B}_0(H)$.

8.10.16. Define a kernel k on the domain $[0,1]^2$ by $k(x,y) = |x-y|$. Show that the corresponding integral operator L_k on $L^2[0,1]$ is Hilbert–Schmidt, and find $\sum \lambda_n^2$, where $\{\lambda_n\}_{n\in\mathbb{N}}$ is the set of nonzero eigenvalues of L_k. FINISH SOLUTION.

8.10.17. Let T be a compact operator on H, and let $|T|$ be the absolute value operator constructed in Definition 8.9.4. Show that the singular values of T are the eigenvalues of $|T|$, i.e., $s_n(T) = \lambda_n(|T|)$ for every n.

8.10.18. We say that a compact operator $T \in \mathcal{B}_0(H)$ is *trace-class* if its sequence of singular numbers $(s_n(T))_{n\in J}$ is summable, i.e., $\sum s_n(T) < \infty$. The space of *trace-class operators* on H is denoted by $\mathcal{B}_1(H)$. Since $\ell^1 \subseteq \ell^2$, we have $\mathcal{B}_1(H) \subseteq \mathcal{B}_2(H)$.

Let T be an arbitrary compact operator on H, and let $|T|$ be its absolute value, as constructed in Definition 8.9.4. Show that if $\{f_k\}$ is any orthonormal basis for H, then

$$\sum_k \bigl\langle |T|f_k,\, f_k \bigr\rangle \;=\; \sum_{n \in J} s_n(T).$$

Note that these quantities might be infinite. Prove that the following three statements are equivalent.

(a) T is trace-class.

(b) $\sum \bigl\langle |T|f_k,\, f_k \bigr\rangle < \infty$ for some orthonormal basis $\{f_k\}$ for H.

(c) $|T|^{1/2} \in \mathcal{B}_2(H)$.

Chapter 9
Functional Analysis

Throughout this chapter the symbol $\mathbb{F}$ will denote a choice of either the real line $\mathbb{R}$ or the complex plane $\mathbb{C}$.

We will often use the symbols f, g, h (instead of x, y, z) to denote a generic element of a normed space.

9.1 Notation for Linear Functionals

The Riesz Representation Theorem for Hilbert spaces tells us that every vector in H is uniquely identified with a continuous linear functional in H^*. However, because the identification $g \mapsto \mu_g$ is antilinear, the standard notation for denoting functions can be confusing in this context. If μ is a functional on H, then the standard way to denote the value of μ at f is to write $\mu(f)$. Unfortunately, if $\mu = \mu_g$ and we define $c\mu$ by $(c\mu)(f) = c\mu(f)$, then $c\mu$ is *not* the functional identified with cg. Instead, Theorem 6.9.2(c) tells us that the functional identified with cg is $\mu_{cg} = \bar{c}\mu_g$. If we want $c\mu$ to denote *this* functional then we must declare that $(c\mu)(f) = \bar{c}\,\mu(f)$. This is inconsistent with the standard interpretation of the meaning of $c\mu$. This unpleasant inconsistency is a consequence of the fact that an inner product is antilinear in the second variable.

To alleviate this problem, we introduce an alternative notation for functionals. Let X be a normed space. Given a functional μ on X, instead of writing $\mu(f)$ we will denote the value that μ takes at $f \in X$ by

$$\langle f, \mu \rangle.$$

If $X = H$ is a Hilbert space, then by the Riesz Representation we know that μ is canonically identified with some element $g \in H$, and the value that μ takes at f is the inner product of f with g. Hence in this case we have $\langle f, \mu \rangle = \langle f, g \rangle$, so our notation is consistent with our identification of μ

with g. On the other hand, if X is a generic normed space then there need not be any inner product on X, and the symbols $\langle f, \mu \rangle$ simply denote the value that μ takes at f. To maintain consistency, we *declare* that the notation $\langle f, \mu \rangle$ is linear as a function of f but antilinear as a function of μ. That is, when we use this notation, the functional $c\mu$ is defined to be the functional whose rule is

$$\langle f, c\mu \rangle \;=\; \bar{c}\,\langle f, \mu \rangle, \qquad f \in X.$$

We will see below some concrete illustrations of why this notation is convenient, but first we give a precise formalization of both the standard notation $\mu(f)$ and our alternative notation $\langle f, \mu \rangle$.

Notation 9.1.1 (Notation for Linear Functionals). Let X be a normed linear space. Given a fixed linear functional $\mu\colon X \to \mathbb{C}$, we use the following two notations to denote the image of f under μ.

(a) We write $\mu(f)$ to denote the image of f under μ, with the understanding that this notation is linear in both f and μ, i.e.,

$$\mu(\alpha f + \beta g) \;=\; \alpha\mu(f) + \beta\mu(g)$$

and

$$(\alpha\mu + \beta\nu)(f) \;=\; \alpha\mu(f) + \beta\nu(f).$$

(b) We write $\langle f, \mu \rangle$ to denote the image of f under μ, with the understanding that this notation is linear in f but antilinear in μ, i.e.,

$$\langle \alpha f + \beta g, \mu \rangle \;=\; \alpha\langle f, \mu \rangle + \beta\langle g, \mu \rangle$$

while

$$\langle f, \alpha\mu + \beta\nu \rangle \;=\; \overline{\alpha}\langle f, \mu \rangle + \overline{\beta}\langle f, \nu \rangle. \tag{9.1}$$

This will be our preferred notation for linear functionals throughout the remainder of this volume. $\diamondsuit$

In summary, the notation $\langle f, \mu \rangle$ simply denotes the image of f under the linear functional μ. However, using this notation we consider the pairing of a vector f with a linear functional μ to be a generalization of the inner product on a Hilbert space, which means that it is a *sesquilinear form* on $X \times X^*$ that is linear as a function of f but antilinear as a function of μ. The notation $\mu(f)$ also denotes the image of f under μ, but this notation is a *bilinear form* that is linear both as a function of f and as a function of μ.

We give some examples that illustrate the use of this notation.

Example 9.1.2. Let $\delta\colon C_b(\mathbb{R}) \to \mathbb{C}$ be the functional constructed in Example 6.3.2. Using the standard notation, the rule for this functional is

$$\delta(f) \;=\; f(0), \qquad f \in C_b(\mathbb{R}).$$

Using our alternative notation, we write the rule for this functional as

$$\langle f, \delta \rangle \;=\; f(0), \qquad f \in C_b(\mathbb{R}).$$

The symbols $\delta(f)$ and $\langle f, \delta \rangle$ both denote the scalar $f(0)$.

When we use the standard notation, $i\delta$ denotes the functional whose rule is

$$(i\delta)(f) \;=\; i\,\delta(f) \;=\; i f(0).$$

When we use our alternative notation, $i\delta$ denotes the functional whose rule is

$$\langle f, i\delta \rangle \;=\; -i\,\delta(f) \;=\; -i f(0).$$

The relationship between δ and $i\delta$ depends on which notation we choose to use. This choice is purely a matter of convenience. For our purposes, the alternative notation is more convenient and will be used henceforth. $\diamondsuit$

Example 9.1.3. Let $\mu\colon X \to \mathbb{C}$ be a linear functional on a normed space X. According to Notation 9.1.1, $\mu(f)$ and $\langle f, \mu \rangle$ both denote the image of $f \in X$ under μ. Since μ is a functional, $\mu(f) = \langle f, \mu \rangle$ is a *scalar* for each $f \in X$. Therefore the operator norm of the functional μ expressed in these two notations is

$$\|\mu\| \;=\; \sup_{\|f\|=1} |\mu(f)| \;=\; \sup_{\|f\|=1} |\langle f, \mu \rangle|.$$

Using the sesquilinear form notation, we have

$$|\langle f, \mu \rangle| \;\le\; \|f\|\,\|\mu\|, \qquad f \in X.$$

In this notation, to prove that a bounded linear functional is continuous, we would assume that $f_n \to f$ in X, and then write

$$\lim_{n\to\infty} \bigl|\langle f, \mu \rangle - \langle f_n, \mu \rangle\bigr| \;=\; \lim_{n\to\infty} \bigl|\langle f - f_n, \mu \rangle\bigr| \;\le\; \lim_{n\to\infty} \|f - f_n\|\,\|\mu\| \;=\; 0.$$

This shows that $\langle f_n, \mu \rangle \to \langle f, \mu \rangle$ as $n \to \infty$, which implies that μ is continuous. $\diamondsuit$

9.2 The Hahn–Banach Theorem

Let H and K be Hilbert spaces. Suppose that M is a closed subspace of H and $A\colon M \to K$ is a linear operator. Even though A is only defined on a subspace of H, we can easily find extensions of A to the entire space H by making use of the fact that H has an inner product. In particular, we know that every vector $f \in H$ can be written as $f = p + e$ for a unique choice of vectors $p \in M$ and $e \in M^{\perp}$. Setting $Af = Ap$, we obtain a linear extension of A to the space H. If A is bounded on M then this extension is also be bounded on H, and in fact the operator norm of the extension is precisely the same as the operator norm of the original operator A. This construction is essentially trivial because of the fact that $M^\perp$ has an orthogonal complement $M^\perp$.

As soon as we move from Hilbert spaces to Banach spaces, the analogous question becomes far more difficult to consider. In a generic Banach space, we no longer have the machinery of orthogonal complements. The Axiom of Choice can be used to show that if M is a subspace of a vector space X, then there exists a subspace N such that $X = M + N$ and $M \cap N = \{0\}$. However, this is typically not much help to us, because even if M is closed, the subspace N need not be closed. The situation that we encounter more often is that we have a *closed* subspace M of a Banach space X and we want to know whether there exists a *closed* subspace N such that every vector $f \in X$ can be written as $f = m + n$ for a unique choice of $m \in M$ and $n \in N$. If this is possible, then we say that M is a *complemented subspace* of X. Unfortunately, there exist closed subspaces of Banach spaces that are not complemented! We do not even need to consider "exotic" spaces in order to find examples, e.g., it is known that c_0 is not a complemented subspace of ℓ^∞ [Phi40], and if $1 < p \leq \infty$, $p \neq 2$, then ℓ^p contains closed subspaces that are not complemented [Mur37].

Consequently, if X is a Banach space that is not a Hilbert space, it is not at all clear whether we will be able to extend operators from subspaces to the entire space, especially if we are interested in preserving properties such as continuity or operator norm. Even for linear functionals, we face a difficult question: Can we extend a linear functional from a subspace of a Banach space to the entire space? If we can, and if the functional is bounded, will the extension also be bounded? Can we also ask that the extension have other special properties?

The Hahn–Banach Theorem is a powerful result that deals with these questions. We will state the Hahn–Banach Theorem and explore a variety of its implications in this section.

9.2.1 Abstract Statement of the Hahn–Banach Theorem

When it comes to linear functionals, most of our interest in this volume centers on bounded linear functionals whose domain is a normed space or a Banach space. That is the setting we will focus on when we consider the applications of the Hahn–Banach Theorem, but the theorem itself can be formulated in a much more general context. Here is the abstract statement of the Hahn–Banach Theorem for functionals on subspaces of complex vector spaces.

Theorem 9.2.1 (Hahn–Banach Theorem). *Let X be a (complex) vector space and let ρ be a seminorm on X. If M is a subspace of X and $\lambda\colon M \to \mathbb{C}$ is a linear functional on M that satisfies*

$$|\langle f, \lambda\rangle| \;\le\; \rho(f), \qquad f \in M, \tag{9.2}$$

then there exists a linear functional $\Lambda\colon X \to \mathbb{C}$ such that

$$\Lambda|_M \;=\; \lambda \qquad \textit{and} \qquad |\langle f, \Lambda\rangle| \;\le\; \rho(f), \quad f \in X. \qquad \diamondsuit \tag{9.3}$$

Equation (9.2) is a kind of boundedness condition on the functional λ. If X is a normed space, then λ is bounded on the subspace M if and only if there is a constant C such that

$$|\langle f, \lambda\rangle| \;\le\; C\,\|f\|, \qquad f \in M.$$

In Theorem 9.2.1, we only require that we have a dominating *seminorm* ρ instead of a dominating *norm*. The power of the Hahn–Banach Theorem is that even though we may not have a complement of the subspace M, we can still extend λ from M to X. Furthermore, this extension Λ will satisfy the *same* boundedness condition that λ satisfies—and it satisfies this on the *entire* space X.

Theorem 9.2.1 also holds for real vector spaces. In fact, in that setting the hypotheses can be made even weaker: The dominating seminorm ρ can be replaced by a *sublinear function* $q\colon X \to \mathbb{R}$ that satisfies

$$q(f+g) \;\le\; q(f) + q(g) \qquad \text{and} \qquad q(cf) \;=\; q(f)$$

for all f, $g \in X$ and scalars $c \ge 0$.

The proof of the Hahn–Banach Theorem takes some preparation, and will be presented in Chapter 10 (or see [Con90], [Fol99], or [Rud91]). In general, the proof is not constructive, as it relies on the Axiom of Choice in the form of Zorn's Lemma. The Hahn–Banach Theorem is intimately tied to convexity, and there are generalizations to more abstract settings, such as locally convex topological vector spaces.

9.2.2 Corollaries of the Hahn–Banach Theorem

In practice, it is usually not the Hahn–Banach Theorem itself but rather its corollaries that are applied. Since these corollaries are so important, when invoking any one of them it is customary to write "by the Hahn–Banach Theorem" instead of "by a corollary to the Hahn–Banach Theorem."

Our first corollary states that any bounded linear functional λ on a (not necessarily closed) subspace M of a normed space X has an extension to the entire space whose operator norm on X equals the operator norm of λ on M. This is easy to prove constructively when the domain is a Hilbert space (indeed, this is precisely the content of Problem 9.2.19). However, it is far from obvious that such an extension is possible if X is not an inner product space.

Corollary 9.2.2 (Hahn–Banach). *Let X be a normed space and let M be a subspace of X. If $\lambda \in M^*$ (i.e., λ is a bounded linear functional on X), then there exists a bounded linear functional $\Lambda \in X^*$ such that*

$$\Lambda|_M \;=\; \lambda \qquad \textit{and} \qquad \|\Lambda\| \;=\; \|\lambda\|.$$

Proof. Set $\rho(f) = \|\lambda\|\,\|f\|$ for $f \in X$. Note that ρ is defined on all of X, and it is a seminorm on X (in fact, it is a norm if $\lambda \neq 0$). Further,

$$|\langle f, \lambda\rangle| \;\le\; \|f\|\,\|\lambda\| \;=\; \rho(f), \qquad \text{all } f \in M.$$

Hence, Theorem 9.2.1 implies that there exists a linear functional $\Lambda\colon X \to \mathbb{C}$ such that $\Lambda|_M = \lambda$ and

$$|\langle f, \Lambda\rangle| \;\le\; \rho(f) \;=\; \|\lambda\|\,\|f\|. \qquad \text{all } f \in X. \tag{9.4}$$

The fact that λ is a restriction of Λ implies that $\|\lambda\| \le \|\Lambda\|$. Conversely, equation (9.4) tells us that

$$\|\Lambda\| \;=\; \sup_{\|f\|=1} |\langle f, \Lambda\rangle| \;\le\; \|\lambda\|. \qquad \diamondsuit$$

To motivate our next corollary, let μ be a bounded linear functional on a normed space X. By definition, the operator norm of μ is

$$\|\mu\| \;=\; \sup_{f \in X,\, \|f\|=1} |\langle f, \mu\rangle|.$$

Thus, we obtain the operator norm of μ by "looking back" at its action on elements of X. The next corollary provides a complementary viewpoint: The norm of a vector $f \in X$ can be obtained by "looking forward" to its action on elements of X^*. If X is an inner product space, then the proof of Corollary 9.2.3 is trivial (see Problem 9.2.17), but for arbitrary normed spaces we need to invoke the Hahn–Banach Theorem.

Corollary 9.2.3 (Hahn–Banach). *If X is a normed space and $f \in X$, then*

$$\|f\| \;=\; \sup_{\mu \in X^*,\, \|\mu\|=1} |\langle f, \mu\rangle|. \tag{9.5}$$

Further, the supremum is achieved.

Proof. Fix $f \in X$, and set

$$\alpha \;=\; \sup_{\mu \in X^*,\, \|\mu\|=1} |\langle f, \mu\rangle|.$$

If $f = 0$ then $\alpha = 0$ and the proof is trivial, so we may assume that $f \neq 0$. By definition of the operator norm we have

$$|\langle f, \mu\rangle| \le \|f\|\, \|\mu\|,$$

so $\alpha \le \|f\|$.

To prove the opposite inequality, let M be the one-dimensional subspace of X spanned by f:

$$M \;=\; \operatorname{span}\{f\} \;=\; \{cf : c \in \mathbb{C}\}.$$

Define $\lambda\colon M \to \mathbb{C}$ by

$$\langle cf, \lambda\rangle \;=\; c\,\|f\|, \qquad c \in \mathbb{C}.$$

Then $\lambda \in M^*$ and

$$\|\lambda\| \;=\; \sup_{m \in M,\, \|m\|=1} |\langle m, \lambda\rangle| \;=\; \sup_{cf \in M,\, \|cf\|=1} |\langle cf, \lambda\rangle| \;=\; 1.$$

Corollary 9.2.2 therefore implies that there exists a functional $\Lambda \in X^*$ such that $\Lambda|_M = \lambda$ and $\|\Lambda\| = \|\lambda\| = 1$. In particular, since $f \in M$ we have $\alpha \ge |\langle f, \Lambda\rangle| = |\langle f, \lambda\rangle| = \|f\|$, and therefore the supremum in equation (9.5) is achieved. $\square$

Now we give one of the most powerful and often-used implications of the Hahn–Banach Theorem. It states that we can find a bounded linear functional that separates a point from a closed subspace of a normed space. Again, this is easy to prove constructively for Hilbert spaces (see Problem 9.2.20), but it is quite amazing that we can do this in arbitrary normed spaces.

Corollary 9.2.4 (Hahn–Banach). *Let X be a normed linear space. Suppose that:*

(a) *M is a closed subspace of X,*

(b) *$f_0 \in X \backslash M$, and*

(c) $d = \operatorname{dist}(f_0, M) = \inf\big\{\|f_0 - m\| : m \in M\big\}$.

Then there exists a functional $\mu \in X^*$ such that

$$\langle f_0, \mu\rangle = 1, \qquad \mu|_M = 0, \qquad \text{and} \qquad \|\mu\| = \frac{1}{d}.$$

Proof. Note that $d > 0$ since M is closed. Define $M_1 = \operatorname{span}\{M, f_0\}$. Since $f_0 \notin M$, each vector $f \in M_1$ can be written as

$$f \;=\; m_f + t_f f_0 \quad \text{for some unique } m_f \in M,\ t_f \in \mathbb{C}.$$

Define $\lambda\colon M_1 \to \mathbb{C}$ by

$$\langle f, \lambda \rangle \;=\; t_f, \qquad \text{all } f \in M_1.$$

Then λ is linear, $\lambda|_M = 0$, and $\langle f_0, \lambda \rangle = 1$.

If $f \in M_1$ and $t_f \neq 0$ then we have $m_f/t_f \in M$, so

$$\|f\| \;=\; \|t_f f_0 + m_f\| \;=\; |t_f| \left\| f_0 - \left(\frac{-m_f}{t_f}\right) \right\| \;\geq\; |t_f|\, d.$$

If $t_f = 0$ (so $f \in M$), the inequality $\|f\| \geq |t_f|\, d$ still holds. Therefore

$$|\langle f, \lambda \rangle| \;=\; |t_f| \;\leq\; \frac{1}{d}\, \|f\|, \qquad \text{all } f \in M_1,$$

so $\|\lambda\| \leq 1/d$.

On the other hand, since d is the distance from f_0 to M, there exist vectors $m_n \in M$ such that $\|f_0 - m_n\| \to d$. Since λ vanishes on M we have $\langle m_n, \lambda \rangle = 0$ for every n, and therefore

$$1 \;=\; \langle f_0, \lambda \rangle \;=\; \langle f_0 - m_n, \lambda \rangle \;\leq\; \|f_0 - m_n\|\, \|\lambda\| \;\to\; d\, \|\lambda\| \quad \text{as } n \to \infty.$$

Consequently, $\|\lambda\| \geq 1/d$.

Finally, applying Corollary 9.2.2, there exists some functional $\mu \in X^*$ such that $\mu|_{M_1} = \lambda$ and $\|\mu\| = \|\lambda\|$. This functional μ has all of the required properties. □

In Section 7.2, and Theorem 6.10.3 in particular, we saw that each sequence $y \in \ell^1$ can be identified with the bounded linear functional $\mu_y \in (\ell^\infty)^*$ defined by

$$\langle x, \mu_y \rangle \;=\; \langle x, y \rangle \;=\; \sum_{k=1}^{\infty} x_k \overline{y_k}, \qquad x \in \ell^\infty.$$

As an application of Corollary 9.2.4 we will prove that there are functionals in $(\ell^\infty)^*$ that cannot be identified with elements of ℓ^1 in this way.

Example 9.2.5. Recall that c_0 is a closed subspace of ℓ^∞. Consequently, the Hahn–Banach Theorem in the form of Corollary 9.2.4 implies that there exists a functional $\mu \in (\ell^\infty)^*$ such that $\mu|_{c_0} = 0$, but μ is not zero on all of ℓ^∞. Suppose that there was some sequence $y \in \ell^1$ such that $\mu = \mu_y$. Then, since the kth standard basis vector δ_k belongs to c_0, we would have

$$y_k \;=\; \langle \delta_k, \mu_y \rangle \;=\; \langle \delta_k, \mu \rangle \;=\; 0.$$

This is true for every k, so the vector y is the zero vector, $y = 0$. But then $\mu = \mu_y = 0$, which is a contradiction since μ is not the zero functional on ℓ^∞. Therefore μ is a functional on ℓ^∞ cannot be identified with μ_y for any $y \in \ell^1$. □

In summary, Theorem 6.10.3 showed us that $T\colon y \mapsto \mu_y$ is an isometric map of ℓ^1 into $(\ell^\infty)^*$, and Example 9.2.5 shows that this mapping T is not surjective. Problem 9.2.28 gives a related construction. In Theorem 6.10.3 we proved that T is a surjective map of ℓ^p onto $\bigl(\ell^{p'}\bigr)^*$ for indices in the range $1 < p < \infty$.

9.2.3 Orthogonal Complements in Normed Spaces

The orthogonal complement of a set S in a Hilbert space H is the set of all vectors in H that are orthogonal to every element of S. That is, $g \in S^\perp$ if and only if $\langle f, g\rangle = 0$ for all $f \in S$. If we instead consider a set S that is a subset of a normed space X, then we no longer have a notion of orthogonality. On the other hand, we still have the concept of linear functionals on X. The pairing $\langle f, \mu\rangle$ with $f \in X$ and $\mu \in X^*$ is our substitute for the inner product, and we can consider the condition $\langle f, \mu\rangle = 0$ to be a generalization of the concept of orthogonality. However, we must keep in mind that f and μ no longer lie in the same space. Instead, f is an element of X while μ is an element of X^*. Still, this gives us a way of generalizing the notion of orthogonal complements from subsets of Hilbert spaces to subsets of normed spaces. The drawback is that the orthogonal complement of a set $S \subseteq X$ will be a subset of X^*, not a subset of X. Here is the formal definition.

Definition 9.2.6 (Orthogonal Complement). If S is a subset of a normed space X, then its *orthogonal complement* $S^\perp$ is the subset of X^* defined by

$$S^\perp \;=\; \bigl\{\mu \in X^* : \langle f, \mu\rangle = 0 \text{ for all } f \in S\bigr\}. \qquad \diamondsuit$$

Exercise 9.2.7. Given a subset S of a normed space X, show that $S^\perp$ is a closed subspace of X^*. $\diamondsuit$

Unfortunately, S and $S^\perp$ are contained in different spaces. Still, the next exercise gives a useful characterization of complete sequences in a normed space in terms of orthogonal complements. For the special case of Hilbert spaces, this characterization appeared earlier in Exercise 5.5.23. Recall that a set S is *complete* if its finite linear span $\mathrm{span}(S)$ is dense in X, i.e., if $\overline{\mathrm{span}}(S) = X$.

Exercise 9.2.8. Given a normed space X, prove the following statements.

(a) If S is a subset of X, then

$$S \text{ is complete in } X \iff S^\perp = \{0\}.$$

(b) If M is a subspace of X, then

$$M \text{ is dense in } X \iff M^\perp = \{0\}. \qquad \diamondsuit$$

In other words, Exercise 9.2.8(a) says that a subset S is complete (i.e., its finite linear span is dense) if and only if the only bounded linear functional that vanishes on S is the zero functional.

Specializing to Hilbert spaces, we recover the fact that a set S is complete if and only if the only vector orthogonal to every element of S is the zero vector.

9.2.4 X^{**} and Reflexivity

Let X be a normed linear space. Following our notational conventions, we write $\langle f, \mu \rangle$ to denote the action of a functional μ on a vector f. This notation is linear in f and antilinear in μ. Therefore, if we hold μ fixed then the mapping $f \mapsto \langle f, \mu \rangle$ is a bounded linear functional on X. On the other hand, if we hold f fixed then the mapping $\mu \mapsto \langle f, \mu \rangle$ is an antilinear functional on X^*. Keeping f fixed, we can create a linear function of μ by instead considering the mapping

$$\mu \mapsto \overline{\langle f, \mu \rangle}, \qquad \mu \in X^*.$$

Thus, each vector $f \in X$ determines a linear functional $\widetilde{f} \colon X^* \to \mathbb{C}$ by the rule

$$\langle \mu, \widetilde{f} \,\rangle = \overline{\langle f, \mu \rangle}, \qquad \mu \in X^*. \tag{9.6}$$

If this linear functional $\widetilde{f}$ is bounded, then it is a bounded linear functional on X^* and hence belongs to the dual space of X^*, which we denote by X^{**}. The next result shows that $\widetilde{f}$ is indeed bounded and its operator norm $\|\widetilde{f}\,\|$ is precisely equal to $\|f\|$. This gives us a natural identification of X with a subset of X^{**}.

Theorem 9.2.9. *Let X be a normed space. For each vector $f \in X$, define $\widetilde{f} : X^* \to \mathbb{C}$ by equation* (9.6)*. Then the following statements hold.*

(a) *$\widetilde{f}$ is a bounded linear functional on X^*, and its operator norm satisfies*

$$\|\widetilde{f}\,\| = \|f\|.$$

(b) *The mapping $\pi \colon X \to X^{**}$ defined by*

$$\pi(f) = \widetilde{f}, \qquad f \in X,$$

is a linear isometry.

Proof. By definition, the operator norm of $\widetilde{f}$ is

$$\|\widetilde{f}\,\| = \sup_{\mu \in X^*,\, \|\mu\|=1} |\langle \mu, \widetilde{f}\,\rangle|.$$

On the other hand, the Hahn–Banach Theorem in the form of Corollary 9.2.3 implies that

$$\|f\| \;=\; \sup_{\mu \in X^*,\, \|\mu\|=1} |\langle f, \mu\rangle|.$$

Since $|\langle f, \mu\rangle| = |\langle \mu, \widetilde{f}\,\rangle|$, statement (a) follows, and statement (b) is an immediate consequence of statement (a). □

The mapping π that is constructed in Theorem 9.2.9 is quite important, so we give it the following name.

Definition 9.2.10 (Natural Embedding of X into X^{**}). Let X be a normed space.

(a) The isometry $\pi\colon f \mapsto \widetilde{f}$ defined in Theorem 9.2.9 is called the *natural embedding* or the *canonical embedding* of X into X^{**}.

(b) If the natural embedding of X into X^{**} is surjective, then we say that X is *reflexive*. ♢

We emphasize that in order for X to be called reflexive, the natural embedding must be a surjective isometry. There exist Banach spaces X such that $X \cong X^{**}$ even though X is not reflexive [Jam51].

Here is the form of the natural embedding on the space $L^p(E)$.

Example 9.2.11. Fix $1 \le p < \infty$, and let E be a measurable subset of $\mathbb{R}^d$. Choose a function $f \in L^p(\mathbb{R}^d)$, and let $\widetilde{f}$ be the image of f under the natural embedding of $L^p(E)$ into $L^p(E)^{**}$. By definition, $\widetilde{f}$ is a bounded linear functional on $L^p(\mathbb{R}^d)^*$. By Theorem 7.2.4, each element of $L^p(E)^*$ is canonically identified with an element of $L^{p'}(E)$. In other words, every element of $L^p(E)^*$ has the form μ_g for some unique $g \in L^{p'}(E)$, where μ_g is the functional on $L^p(E)$ given by

$$\langle f, \mu_g\rangle \;=\; \langle f, g\rangle \;=\; \int_E f(x)\,\overline{g(x)}\,dx, \qquad f \in L^p(E).$$

Therefore the action of $\widetilde{f}$ on μ_g is simply the "inner product" of g with f (this is not a true inner product because f and g belong to different spaces):

$$\langle \mu_g, \widetilde{f}\,\rangle \;=\; \overline{\langle f, \mu_g\rangle} \;=\; \overline{\langle f, g\rangle} \;=\; \langle g, f\rangle \;=\; \int_E g(x)\,\overline{f(x)}\,dx. \qquad ♢$$

To summarize, each function $f \in L^p(E)$ determines a unique functional $\widetilde{f} \in L^p(E)^{**}$, and the action of $\widetilde{f}$ on $\mu_g \in L^p(E)^*$ is precisely the action of $g \in L^{p'}(E)$ on f via the "inner product"

$$\langle \mu_g, \widetilde{f}\,\rangle \;=\; \langle g, f\rangle, \qquad g \in L^{p'}(E),\ f \in L^p(E).$$

We usually identify f with $\widetilde{f}$, and in this sense of identification we write

$$L^p(E) \subseteq L^p(E)^{**}.$$

If equality holds on the line above (i.e., if the natural embedding is surjective), then $L^p(E)$ is reflexive. Exercise 9.4.4 will show that this happens precisely for indices p that lie in the range $1 < p < \infty$. Similarly, ℓ^p is reflexive if and only if $1 < p < \infty$.

Exercise 9.2.12. Recall that ${c_0}^* \cong \ell^1$ and $(\ell^1)^* \cong \ell^\infty$, so we have ${c_0}^{**} \cong \ell^\infty$. Show that the natural embedding $T\colon c_0 \to \ell^\infty$ is the inclusion mapping ($Tx = x$ for $x \in c_0$), and prove that c_0 is not reflexive. $\diamondsuit$

9.2.5 Adjoints of Operators on Banach Spaces

We saw in Section 8.2 that if H, K are Hilbert spaces and $A \in \mathcal{B}(H,K)$, then there exists an adjoint operator $A^* \in \mathcal{B}(K,H)$ that is uniquely defined by the requirement that

$$\forall\, f \in H, \quad \forall\, g \in H, \quad \langle Af, g\rangle_K \;=\; \langle f, A^*g\rangle_H. \tag{9.7}$$

Now we let X, Y be Banach spaces, and consider an operator $A \in \mathcal{B}(X,Y)$. We will see that, as a consequence of the Hahn–Banach Theorem, there exists a unique adjoint $A^* \in \mathcal{B}(Y^*, X^*)$, and this adjoint is defined by a condition very similar to the one that appears in equation (9.7). However, while A maps X into Y, the adjoint A^* maps Y^* into X^*. Therefore, in contrast to the situation for Hilbert spaces, we usually cannot consider compositions of A with A^*.

Exercise 9.2.13. Let X and Y be Banach spaces, and fix a bounded linear operator $A \in \mathcal{B}(X,Y)$.

(a) Given $\mu \in Y^*$, define a functional $A^*\mu\colon X \to \mathbb{C}$ by

$$\langle f, A^*\mu\rangle \;=\; \langle Af, \mu\rangle, \qquad f \in X.$$

Show that $A^*\mu$ is linear and bounded, and therefore $A^*\mu \in X^*$.

(b) Show that the mapping $A^*\colon \mu \mapsto A^*\mu$ is a bounded linear mapping of Y^* into X^*, and the operator norm of this mapping is $\|A^*\| = \|A\|$.

(c) Prove that A^* is the unique operator that maps Y^* into X^* and satisfies

$$\forall\, f \in X, \quad \forall\, \mu \in Y^*, \quad \langle Af, \mu\rangle \;=\; \langle f, A^*\mu\rangle. \qquad \diamondsuit \tag{9.8}$$

We formalize this construction as follows.

Definition 9.2.14 (Adjoint). Let X, Y be Banach spaces. The *adjoint* of an operator $A \in \mathcal{B}(X,Y)$ is the unique operator $A^*\colon Y^* \to X^*$ that satisfies equation (9.8). $\diamondsuit$

Here are some examples.

Example 9.2.15. Choose $1 \le p < \infty$, and fix $\phi \in L^\infty(\mathbb{R})$. Define an operator $M_\phi\colon L^p(\mathbb{R}) \to L^p(\mathbb{R})$ by

$$M_\phi f \;=\; f\phi, \qquad f \in L^p(\mathbb{R}).$$

Exercise 6.6.13 shows that M_ϕ is bounded and $\|M_\phi\| = \|\phi\|_\infty$.

Theorem 7.2.4 tells us that the dual space of $L^p(\mathbb{R})$ can be naturally identified with $L^{p'}(\mathbb{R})$. Using this identification, the adjoint of M_ϕ is the unique operator $M_\phi^*\colon L^{p'}(\mathbb{R}) \to L^{p'}(\mathbb{R})$ that satisfies $\langle f, M_\phi^* g\rangle = \langle M_\phi f, g\rangle$ for all $f \in L^p(\mathbb{R})$ and $g \in L^p(\mathbb{R})^* = L^{p'}(\mathbb{R})$. Since

$$\begin{aligned} \langle f, M_\phi^* g\rangle \;=\; \langle M_\phi f, g\rangle \;=\; \langle f\phi, g\rangle \;&=\; \int_{-\infty}^{\infty} f(t)\,\phi(t)\,\overline{g(t)}\,dt \\ &=\; \int_{-\infty}^{\infty} f(t)\,\overline{g(t)\,\overline{\phi(t)}}\,dt \\ &=\; \langle f, g\overline{\phi}\rangle, \end{aligned}$$

it follows that $M_\phi^* g = g\overline{\phi}$ for $g \in L^{p'}(\mathbb{R})$. Thus M_ϕ^* is multiplication by the function $\overline{\phi}$. Therefore

$$M_\phi^* \;=\; M_{\overline{\phi}},$$

although we should be careful to note that the operator M_ϕ maps $L^p(\mathbb{R})$ into itself while $M_\phi^* = M_{\overline{\phi}}$ maps $L^{p'}(\mathbb{R})$ into itself. $\diamondsuit$

Exercise 9.2.16. Assume that k satisfies the conditions of Schur's Test, as given in equation (6.26). We know then that the corresponding integral operator L_k defined by

$$L_k f(x) \;=\; \int_{-\infty}^{\infty} k(x,y)\,f(y)\,dy$$

is a bounded mapping of $L^p(\mathbb{R})$ into itself for each index $1 \le p \le \infty$. Prove that the adjoint of L_k is also an integral operator, and find the corresponding kernel function k^* such that $(L_k)^* = L_{k^*}$. $\diamondsuit$

Problems

9.2.17. Given an inner product space H, prove the following statements.

(a) If $\langle x, y\rangle = \langle x, z\rangle$ for every $x \in H$, then $y = z$.

(b) If $x \in H$, then $\|x\| = \sup_{\|y\|=1} |\langle x, y\rangle|$.

9.2.18. Given $y = (y_k) \in \ell^1$, define $\mu_y \colon c_0 \to \mathbb{F}$ by $\langle x, \mu_y \rangle = \sum x_k y_k$. Show that $y \mapsto \mu_y$ is an isometric isomorphism of ℓ^1 onto ${c_0}^*$. Thus ${c_0}^* \cong \ell^1$.

9.2.19. Suppose that M is a subspace of a Hilbert space H and fix a vector $\lambda \in M$. Show directly that there exists a vector $\Lambda \in H$ such that $\langle f, \Lambda \rangle = \langle f, \lambda \rangle$ for $f \in M$ and $\|\Lambda\| = \|\lambda\|$.

9.2.20. Suppose that M is a closed subspace of a Hilbert space H, $f_0 \in H \backslash M$, and $d = \operatorname{dist}(f_0, M)$. Show directly that there exists a vector $\mu \in H$ such that $\langle f_0, \mu \rangle = 1$, $\langle f, \mu \rangle = 0$ for all $f \in M$, and $\|\mu\| = 1/d$.

9.2.21. Let S be a subspace of a normed space X. Prove that the closure of S satisfies

$$\overline{S} \;=\; \bigcap \bigl\{ \ker(\mu) : \mu \in X^* \text{ and } S \subseteq \ker(\mu) \bigr\}.$$

9.2.22. Let g be an element of a Banach space X. Show that there exists a functional $\mu \in X^*$ such that $\|\mu\| = 1$ and $\langle g, \mu \rangle = \|g\|$.

9.2.23. (a) Let X be a normed space. Show that if X^* is separable then X is separable, but the converse can fail.

(b) Show that ℓ^1 and $L^1(E)$ are not reflexive.

9.2.24. Given Banach spaces X, Y and operators A, $B \in \mathcal{B}(X, Y)$, prove that $(\alpha A + \beta B)^* = \overline{\alpha}\, A^* + \overline{\beta}\, B^*$ for all α, $\beta \in \mathbb{C}$.

9.2.25. Given a closed subspace M of a Banach space X, let $\epsilon \colon M \to X$ be the embedding map, i.e., $\epsilon(x) = x$ for $x \in M$. Prove that $\epsilon^* \colon X^* \to M^*$ is the restriction map defined by $\epsilon^* \mu = \mu|_M$ for $\mu \in X^*$.

9.2.26. (a) Let M be a closed subspace of a Banach space X. Let $\rho_X \colon X \to X^{**}$ and $\rho_M \colon M \to M^{**}$ be the natural embeddings. Let $i \colon M \to X$ be the inclusion map, $i(x) = x$ for $x \in M$. Show that there exists an isometry $\phi \colon M^{**} \to X^{**}$ such that

$$\rho_X \circ i \;=\; \phi \circ \rho_M.$$

Prove further that $\phi(M^{**}) = (M^\perp)^\perp$.

Remark: It may be helpful to draw a commutative diagram showing the four spaces X, X^{**}, M, M^{**} and the four mappings ρ_X, ρ_M, i, and ϕ.

(b) Use part (a) to show that if X is reflexive, then any closed subspace of X is reflexive.

9.2.27. Let X, Y be Banach spaces. Show that if $T \colon X \to Y$ is a topological isomorphism, then its adjoint $T^* \colon Y^* \to X^*$ is a topological isomorphism, and if T is an isometric isomorphism then so is T^*.

9.2.28. This problem will use the Hahn–Banach Theorem to construct an element of $(\ell^\infty)^*$ that does not belong to ℓ^1. Recall from Problem 6.10.8 that the subspace c of ℓ^∞ that consists of the convergent sequences is a closed subspace of ℓ^∞.

(a) Show that there exists a bounded linear functional μ on ℓ^∞ that satisfies

$$\langle x, \mu \rangle \;=\; \lim_{k\to\infty} x_k, \qquad x = (x_k)_{k\in\mathbb{N}} \in c.$$

(b) Show that $\mu \neq \widetilde{x}$ for any $x \in \ell^1$, where $\widetilde{x}$ is the image of x under the natural embedding of ℓ^1 into $(\ell^1)^{**} = (\ell^\infty)^*$.

(c) Is the functional μ constructed in part (a) unique?

9.3 The Baire Category Theorem

It is not possible to write the Euclidean plane $\mathbb{R}^2$ as the union of *countably many* straight lines. This statement is a special case of the Baire Category Theorem, which we will prove in this section. The Baire Category Theorem says that a complete metric space cannot be written as a countable union of sets that are "nowhere dense" in the following sense.

Definition 9.3.1 (Nowhere Dense Sets). Let X be a metric space, and fix $E \subseteq X$.

(a) E is *nowhere dense* or *rare* if $X\setminus\overline{E}$ is dense in X.

(b) E is *meager* or *first category* if it can be written as a countable union of nowhere dense sets.

(c) E is *nonmeager* or *second category* if it is not meager. $\diamondsuit$

Although the terms "first category" and "second category" are in common use, they are easily confusable, so we will prefer the more descriptive terms "meager" and "nonmeager."

Here is a useful reformulation of the meaning of "nowhere dense."

Exercise 9.3.2. Given a nonempty subset E of a metric space X, prove that the following two statements are equivalent.

(a) E is nowhere dense.

(b) $\overline{E}$ has an empty interior, i.e., $\overline{E}$ contains no nonempty open subsets. $\diamondsuit$

In particular, a nonempty *closed* set is nowhere dense if and only if it contains no open balls. For example, the Cantor set is a nowhere dense subset of $[0,1]$ because it is closed and contains no intervals. The set of rationals $\mathbb{Q}$ also contains no intervals, but it is not a nowhere dense subset of $\mathbb{R}$ because its closure is $\mathbb{R}$, which has a nonempty interior. On the other hand, $\mathbb{Q}$ is meager in $\mathbb{R}$ since it is the countable union of the singletons $\{x\}$ with x ranging through $\mathbb{Q}$, and each singleton $\{x\}$ is nowhere dense.

Now we prove the Baire Category Theorem. In contrast to the Hahn–Banach Theorem, the driving property here is completeness rather than convexity.

Theorem 9.3.3 (Baire Category Theorem). *If X is a nonempty complete metric space, then X is a nonmeager subset of X. That is, if*

$$X = \bigcup_{n=1}^{\infty} E_n, \tag{9.9}$$

then at least one of the sets E_n cannot be nowhere dense.

Proof. Suppose that $X = \cup E_n$ where each set E_n is nowhere dense. Then each of the sets

$$U_n = X \setminus \overline{E_n}, \qquad n \in \mathbb{N},$$

is a dense, open subset of X.

Since U_1 is dense, it is nonempty. Hence there exists some point $x_1 \in U_1$. Since U_1 is open, there is some radius $r_1 > 0$ such that

$$B_1 = B_{r_1}(x_1) \subseteq U_1.$$

Since U_2 is dense, there is a point $x_2 \in U_2 \cap B_1$. Since U_2 and B_1 are both open, there exists a radius $r_2 > 0$ such that

$$B_2 = B_{r_2}(x_2) \subseteq U_2 \cap B_1.$$

By making r_2 smaller we can ensure that we also have $\overline{B_2} \subseteq B_1$. And, if needed, by taking r_2 to be even smaller yet we can further assume that $r_2 < r_1/2$.

Continuing in this way, we obtain points $x_n \in U_n$ and balls B_n such that

$$B_n = B_{r_n}(x_n) \subseteq U_n, \qquad \overline{B_n} \subseteq B_{n-1}, \qquad \text{and} \qquad r_n < \frac{r_{n-1}}{2}.$$

In particular, $r_n \to 0$, and the balls B_n form a nested decreasing sequence.

Fix $\varepsilon > 0$, and let N be large enough that $r_N < \varepsilon$. If $m, n > N$, then we have $x_m, x_n \in B_N$, so $\mathrm{d}(x_m, x_n) < 2r_N < 2\varepsilon$. Thus the sequence $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy. Since X is complete, this implies that there is some point $x \in X$ such that $x_n \to x$ as $n \to \infty$.

Now let $M > 0$ be any fixed integer. Since the balls B_n form a nested decreasing sequence, we have $x_n \in B_{M+1}$ for all $n > M$. Since $x_n \to x$, this implies that $x \in \overline{B_{M+1}} \subseteq B_M$. This is true for every M, so

$$x \in \bigcap_{n=1}^{\infty} B_n \subseteq \bigcap_{n=1}^{\infty} U_n = \bigcap_{n=1}^{\infty} (X \setminus \overline{E_n}).$$

But this implies that $x \notin \cup E_n$, which is a contradiction. $\square$

The hypothesis of completeness in the Baire Category Theorem is necessary, i.e., the theorem can fail if X is not complete. For example, the set of rationals $\mathbb{Q}$ is a metric space with respect to $\mathrm{d}(x,y) = |x-y|$. However, $\mathbb{Q}$ is not complete, yet $\mathbb{Q}$ is a meager set in $\mathbb{Q}$.

Specializing to the case where all of the sets E_n in Theorem 9.3.3 are closed gives us the following corollary. This is the form of the Baire Category Theorem that we typically apply in practice.

Corollary 9.3.4. *Assume that X is a nonempty complete metric space. If*

$$X = \bigcup_{n=1}^{\infty} E_n \tag{9.10}$$

where each E_n is a closed *subset of X, then at least one set E_n must contain a nonempty open subset.*

Proof. If none of the E_n contain a nonempty open subset then they are all nowhere dense. Consequently, if equation (9.10) holds then X is a meager set, which contradicts the Baire Category Theorem. $\square$

Various applications of the Baire Category Theorem are given in the Problems. We present one implication here. Before stating the next corollary, note that if U is an open set that is dense in X, then its complement $E = X \setminus U$ is closed and contains no balls (why?). Therefore $E = X \setminus U$ is a nowhere dense subset of X.

Corollary 9.3.5. *Let X be a complete metric space. If U_n is an open, dense subset of X for each $n \in \mathbb{N}$, then $\cap U_n \neq \varnothing$. In fact, $\cap U_n$ is dense in X.*

Proof. Each set $E_n = U_n^{\mathrm{C}} = X \setminus U_n$ is closed and nowhere dense.

Suppose that $\cap U_n$ is not dense in X. Then there is some $x \in X$ such that there is an open ball $B_r(x)$ entirely contained in the complement of $\cap U_n$. That is,

$$B_r(x) \subseteq \left(\bigcap_{n=1}^{\infty} U_n\right)^{\mathrm{C}} = \bigcup_{n=1}^{\infty} E_n.$$

The closed ball

$$Y = \overline{B_{r/2}(x)}$$

is a closed subset of X and therefore, by Problem 8.4.11, is itself a complete metric space. We have $Y = \bigcup_{n=1}^{\infty} (E_n \cap Y)$. Each set $E_n \cap Y$ is closed in Y, and its complement in Y is

$$Y \setminus (E_n \cap Y) = Y \cap U_n.$$

Since U_n is dense in X, the set $Y \cap U_n$ is dense in Y. Hence $E_n \cap Y$ is closed and its complement is open and dense in Y, Therefore we have written Y as a countable union of the nowhere dense sets $E_n \cap Y$, which contradicts the Baire Category Theorem. $\square$

For an exercise that uses Corollary 9.3.5, see Problem 9.3.16.

Problems

9.3.6. Determine whether the following subsets of ℓ^1 are nowhere dense.

(a) The set S consisting of sequences whose components past the 99th entry are all zero:

$$S = \{x = (x_1, \dots, x_{99}, 0, 0, \dots) : x_1, \dots, x_{99} \in \mathbb{F}\}.$$

(b) c_{00} (the space of all "finite sequences").

(c) $Q = \{x = (x_k)_{k \in \mathbb{N}} \in \ell^1 : x_k \geq 0 \text{ for every } k\}$.

9.3.7. Which of the sets listed in Problem 9.3.6 are meager subsets of ℓ^1?

9.3.8. Prove the following statements.

(a) If $x \in \mathbb{R}$, then $\mathbb{R} \setminus \{x\}$ is an open, dense subset of $\mathbb{R}$.

(b) The set of irrationals, $\mathbb{R} \setminus \mathbb{Q}$, is a dense subset of $\mathbb{R}$, but it is neither open nor closed.

9.3.9. Let $\mathcal{P}$ be the set of all polynomials with complex coefficients. Prove that there is no norm $\|\cdot\|$ on $\mathcal{P}$ such that $\mathcal{P}$ is complete with respect to $\|\cdot\|$.

9.3.10. Suppose that f is an infinitely differentiable function on $\mathbb{R}$ such that for each $x \in \mathbb{R}$ there exists some integer $n_x \geq 0$ so that $f^{(n_x)}(x) = 0$. Prove that there exists some open interval (a, b) and some polynomial p such that $f(x) = p(x)$ for all $x \in (a, b)$.

9.3.11. Show that $\mathrm{Lip}[0,1]$ is a meager subset of $C[0,1]$. Conclude that $C^1[0,1]$ is a meager subset of $C[0,1]$.

9.3.12. Let D be the subset of $C[0,1]$ consisting of all functions $f \in C[0,1]$ that have a right-hand derivative at at least one point in $[0,1]$. Show that D is meager in $C[0,1]$, and conclude that there are functions in $C[0,1]$ that are not differentiable at any point.

9.3.13. Prove that the set of rationals $\mathbb{Q}$ is an F_σ-set in $\mathbb{R}$, but is not a G_δ-set.

9.3.14. For each $n \in \mathbb{N}$, let $f_n \colon \mathbb{R} \to \mathbb{R}$ be a continuous function, and suppose that $f_n(x)$ converges to $f(x)$ for each $x \in \mathbb{R}$. Prove that

$$D = \{x \in \mathbb{R} : f \text{ is discontinuous at } x\}$$

is a meager subset of $\mathbb{R}$.

Hint: Consider $E_k = \{x \in \mathbb{R} : \operatorname{osc}_f(x) \ge \frac{1}{k}\}$, where

$$\operatorname{osc}_f(x) \;=\; \inf_{\delta>0} \sup\bigl\{|f(y) - f(z)| : y, z \in B_\delta(x)\bigr\}.$$

Note that f is continuous at x if and only if $\operatorname{osc}_f(x) = 0$.

9.3.15. Given an infinite-dimensional Banach space X, prove the following statements.

(a) Every Hamel basis for X is uncountable.

(b) If M is an infinite-dimensional subspace of X that has a countable Hamel basis, then M is a meager subset of X and cannot be closed.

(c) $C_c(\mathbb{R})$ is a meager, dense subspace of $C_0(\mathbb{R})$ that does not have a countable Hamel basis.

9.3.16. Suppose that U is an unbounded open subset of $\mathbb{R}$, and let

$$A \;=\; \bigl\{x \in \mathbb{R} : kx \in U \text{ for infinitely many } k \in \mathbb{Z}\bigr\}.$$

(a) Given $n \in \mathbb{N}$, set $A_n = \bigcup_{|k|>n} U/k$, where $U/k = \{x/k : x \in U\}$. Prove that A_n is dense in $\mathbb{R}$.

Hint: Show that $A_n \cap (a,b) \ne \varnothing$ for every open interval (a,b).

(b) Show that $A = \bigcap_{n=1}^\infty A_n$, and prove that A is dense in $\mathbb{R}$.

9.4 The Uniform Boundedness Principle

The Uniform Boundedness Principle states that a family of bounded linear operators on a Banach space that are uniformly bounded *at each individual point* must actually be *uniformly bounded in operator norm*! We will use the Baire Category Theorem to prove this result.

Theorem 9.4.1 (Uniform Boundedness Principle). *Let X be a Banach space and let Y be a normed space. If $\{A_i\}_{i\in I}$ is a collection of operators in $\mathcal{B}(X,Y)$ that satisfies*

$$\forall\, f \in X, \quad \sup_{i\in I} \|A_i f\| \;<\; \infty,$$

then

$$\sup_{i\in I} \|A_i\| \;=\; \sup_{i\in I} \sup_{\|f\|=1} \|A_i f\| \;<\; \infty.$$

Proof. Set

$$E_n \;=\; \biggl\{f \in X \,:\, \sup_{i\in I} \|A_i f\| \le n\biggr\}.$$

By hypothesis, we have $X = \cup E_n$. Exercise: Show that E_n is closed (use the fact that each operator A_i is continuous).

Applying the Baire Category Theorem, we see that some set E_n must contain an open ball, say

$$B_r(f_0) \subseteq E_n.$$

Choose any vector $f \in X$, and let

$$g = f_0 + sf \quad \text{where } s = \frac{r}{2\|f\|}.$$

Then $g \in B_r(f_0) \subseteq E_n$, so

$$\|A_i f\| = \left\|A_i\Big(\frac{g-f_0}{s}\Big)\right\| \le \frac{1}{s}\big(\|A_i g\| + \|A_i f_0\|\big) \le \frac{2\,\|f\|}{r}\,2n = \frac{4n}{r}\,\|f\|.$$

Consequently $\|A_i\| \le 4n/r$, and therefore

$$\sup_{i\in I} \|A_i\| \le \frac{4n}{r} < \infty. \qquad \diamondsuit$$

The following special case of the Uniform Boundedness Principle can be very useful. Sometimes the names "Uniform Boundedness Principle" and "Banach–Steinhaus Theorem" are used interchangeably.

Exercise 9.4.2 (Banach–Steinhaus Theorem). Let X and Y be Banach spaces. Suppose that $\{A_n\}_{n\in\mathbb{N}}$ is a countable sequence of operators in $\mathcal{B}(X,Y)$, and

$$Af = \lim_{n\to\infty} A_n f \quad \text{exists for each } f \in X.$$

Prove that $A \in \mathcal{B}(X,Y)$ and $\|A\| \le \sup \|A_n\| < \infty$. $\diamondsuit$

The hypotheses of the Banach–Steinhaus Theorem do *not* imply that $A_n \to A$ in operator norm (see Problem 8.1.8). Also, it is important in Exercise 9.4.2 that we have a *countable sequence* of operators $\{A_n\}_{n\in\mathbb{N}}$ rather than a generic net of operators.

The next result is a typical application of the Banach–Steinhaus Theorem.

Theorem 9.4.3. *Fix $1 \le p \le \infty$ and let $y = (y_k)_{k\in\mathbb{N}}$ be a sequence of scalars. Then*

$$\sum_{k=1}^{\infty} x_k y_k \text{ converges for every } x = (x_k)_{k\in\mathbb{N}} \in \ell^p \iff y \in \ell^{p'}.$$

Proof. $\Leftarrow$. This follows from Hölder's Inequality (Theorem 5.2.2).

$\Rightarrow$. We will prove the case $1 < p < \infty$ (the cases $p = 1$ and $p = \infty$ are assigned as Problem 9.4.6).

Assume that the series $\sum x_k y_k$ converges for every sequence $x \in \ell^p$. Define functionals μ_N, μ on ℓ^p by

$$\langle x, \mu \rangle \;=\; \sum_{k=1}^{\infty} x_k \overline{y_k} \qquad \text{and} \qquad \langle x, \mu_N \rangle \;=\; \sum_{k=1}^{N} x_k \overline{y_k}.$$

Both μ and μ_N are linear. Given $x \in \ell^p$, we have

$$|\langle x, \mu_N \rangle| \;\le\; \biggl(\sum_{k=1}^{N} |x_k|^p\biggr)^{1/p} \biggl(\sum_{k=1}^{N} |y_k|^{p'}\biggr)^{1/p'} \;\le\; C_N \, \|x\|_p,$$

where $C_N = \bigl(\sum_{k=1}^{N} |y_k|^{p'}\bigr)^{1/p'}$ is a finite constant that is independent of x (though it can depend on N). Therefore $\mu_N \in (\ell^p)^*$ for each integer N. By hypothesis,

$$\lim_{N\to\infty} \langle x, \mu_N \rangle \;=\; \langle x, \mu \rangle, \qquad x \in \ell^p.$$

Applying the Banach–Steinhaus Theorem, it follows that $\mu \in (\ell^p)^*$ and

$$\|\mu\| \;\le\; C \;=\; \sup_N \|\mu_N\| \;<\; \infty.$$

It remains to show that $y \in \ell^{p'}$. We proved in Theorem 6.10.3 that $(\ell^p)^*$ is isomorphic to $\ell^{p'}$. Since $\mu \in (\ell^p)^* \cong \ell^{p'}$, there exists some sequence $z \in \ell^{p'}$ such that $\langle x, \mu \rangle = \langle x, z \rangle = \sum x_k \overline{z_k}$ for all $x \in \ell^p$. Letting δ_k denote the kth standard basis vector, we have $\overline{y_k} = \langle \delta_k, \mu \rangle = \overline{z_k}$, so $y = z \in \ell^{p'}$.

Here is an alternative direct proof that y belongs to $\ell^{p'}$. Set

$$x_N \;=\; \bigl(\alpha_1 \, |y_1|^{p'-1}, \dots, \alpha_N \, |y_N|^{p'-1}, 0, 0, \dots\bigr) \;\in\; \ell^p,$$

where α_k is a scalar of unit modulus such that $\alpha_k \overline{y_k} = |y_k|$. Since $1 < p' < \infty$ and $(p'-1)p = p'$,

$$\begin{aligned}
\sum_{k=1}^{N} |y_k|^{p'} \;&=\; \sum_{k=1}^{N} \alpha_k \, |y_k|^{p'-1} \, \overline{y_k} \\
&=\; |\langle x_N, \mu \rangle| \\
&\le\; \|x_N\|_p \, \|\mu\| \\
&\le\; C \biggl(\sum_{k=1}^{N} |y_k|^{(p'-1)p}\biggr)^{1/p} \\
&=\; C \biggl(\sum_{k=1}^{N} |y_k|^{p'}\biggr)^{1/p}.
\end{aligned}$$

Dividing through by $\left(\sum_{k=1}^N |y_k|^{p'}\right)^{1/p}$ and noting that $1 - \frac{1}{p} = \frac{1}{p'}$, we obtain

$$\left(\sum_{k=1}^N |y_k|^{p'}\right)^{1/p'} = \left(\sum_{k=1}^N |y_k|^{p'}\right)^{1-\frac{1}{p}} \le C.$$

Letting $N \to \infty$, it follows that $\|y\|_{p'} \le C < \infty$. $\square$

Combining the preceding results with the statements made in Section 7.2, we see that $L^p(E)^* \cong L^{p'}(E)$ and $(\ell^p)^* \cong \ell^{p'}$ for $1 \le p < \infty$. Consequently,

$$L^p(E)^{**} \cong L^p(E) \quad \text{and} \quad (\ell^p)^{**} \cong \ell^p, \qquad 1 < p < \infty.$$

This is not quite enough to show that $L^p(E)$ and ℓ^p are reflexive when $1 < p < \infty$, because to conclude reflexivity we must show that the *natural embedding* is the isomorphism. The next exercise asks for a proof of this fact.

Exercise 9.4.4. Let $E \subseteq \mathbb{R}^d$ be Lebesgue measurable. Prove that ℓ^p and $L^p(E)$ are reflexive for each index $1 < p < \infty$. $\diamondsuit$

Problem 9.2.23 told us that if a dual space X^* is separable then X must be separable as well. Since ℓ^1 is separable but $\ell^\infty = (\ell^1)^*$ is not, it follows that ℓ^1 is not reflexive, and similarly $L^1(E)$ is not reflexive. Problem 9.2.28 gives an example of a bounded linear functional on ℓ^∞ that cannot be identified with an element of ℓ^1. The spaces ℓ^∞ and $L^\infty(E)$ are not reflexive either, although it is more difficult to show this.

Here is one more example of a nonreflexive space.

Example 9.4.5. By Problem 9.2.18,

$$(c_0)^{**} \cong (\ell^1)^* \cong \ell^\infty.$$

Since c_0 is separable while ℓ^∞ is not, c_0 cannot be isomorphic to ℓ^∞, and therefore c_0 cannot be isomorphic to $(c_0)^{**}$. Therefore c_0 is not reflexive. The space c_0 is one of the few easily exhibited nonreflexive separable spaces whose dual is separable. $\diamondsuit$

Problems

9.4.6. Prove Theorem 9.4.3 for the cases $p = 1$ and $p = \infty$.

9.4.7. Let X, Y be Banach spaces, and let S be a dense subspace of X. Suppose that $A_n \in \mathcal{B}(X,Y)$ for $n \in \mathbb{N}$, and $Ax = \lim_{n\to\infty} A_n x$ exists for each vector $x \in S$.

(a) Show that if $\sup \|A_n\| < \infty$, then A extends to a bounded map that is defined on all of X, and $Ax = \lim_{n\to\infty} A_n x$ for all $x \in X$.

(b) Give an example that shows that the hypothesis $\sup \|A_n\| < \infty$ in part (a) is necessary.

9.4.8. (a) Let X be a Banach space. Prove that $S \subseteq X^*$ is bounded if and only if $\sup\{|\langle x, \mu\rangle| : \mu \in S\} < \infty$ for each $x \in X$.

(b) Let X be a normed linear space. Prove that $S \subseteq X$ is bounded if and only if $\sup\{|\langle x, \mu\rangle| : x \in S\} < \infty$ for each $\mu \in X^*$.

9.4.9. Fix $1 \le p,\ q \le \infty$. Let $A = [a_{ij}]_{i,j\in\mathbb{N}}$ be an infinite matrix and for each $i \in \mathbb{N}$ define $a_i = (a_{ij})_{j\in\mathbb{N}}$. Suppose that:

(a) $(Ax)_i = \langle x, \overline{a_i}\rangle = \sum_j a_{ij} x_j$ converges for each $x \in \ell^p$ and $i \in \mathbb{N}$, and

(b) $Ax = \big((Ax)_i\big)_{i\in\mathbb{N}} = \big(\langle x, \overline{a_i}\rangle\big)_{i\in\mathbb{N}} \in \ell^q$ for each $x \in \ell^p$.

Identifying the matrix A with the map $x \mapsto Ax$, prove that $A \in \mathcal{B}(\ell^p, \ell^q)$.

9.4.10. Let X, Y, Z be Banach spaces. Suppose that $B\colon X \times Y \to Z$ is bilinear, i.e., $B_f(h) = B(f,h)$ and $B^g(h) = B(h,g)$ are linear functions of h for each $f \in X$ and $g \in Y$. Prove that the following three statements are equivalent.

(a) $B_f\colon Y \to Z$ and $B^g\colon X \to Z$ are continuous for each $f \in X$ and $g \in Y$.

(b) There is a constant $C > 0$ such that

$$\|B(f,g)\| \ \le\ C\,\|f\|\,\|g\|, \qquad f \in X,\ g \in Y.$$

(c) B is a continuous mapping of $X \times Y$ into Z (the norm on $X \times Y$ is $\|(x,y)\| = \|x\|_X + \|y\|_Y$). Note that B need not be linear *on the domain* $X \times Y$.

9.5 The Open Mapping Theorem

By definition, a function that maps one topological space into another is continuous if the inverse image of any open set is open. The *direct image* of an open set under a continuous map need not be open. For example, $f(x) = \sin x$ is a continuous mapping on the real line, but f maps the open interval $(0, 2\pi)$ to the closed interval $[-1, 1]$. On the other hand, some functions do map open sets to open sets. We give those functions the following name.

Definition 9.5.1 (Open Mapping). Let X and Y be topological spaces. We say that a function $A\colon X \to Y$ is an *open mapping* if

$$U \text{ is open in } X \implies A(U) \text{ is open in } Y. \qquad \diamondsuit$$

Although a continuous function need not be an open mapping, in this section we will prove the Open Mapping Theorem, which asserts that a continuous linear surjection of one Banach space onto another must be an open map.

The key to the proof of the Open Mapping Theorem is the following lemma. For clarity, we will write $B_r^X(f)$ and $B_r^Y(g)$ to distinguish open balls in X from open balls in Y.

Lemma 9.5.2. *Let X be a Banach space, let Y be a normed space, and fix an operator $A \in \mathcal{B}(X,Y)$. If $\overline{A(B_1^X(0))}$ contains an open ball in Y, then $A(B_1^X(0))$ contains an open ball of the form $B_r^Y(0)$.*

Proof. Suppose that $\overline{A(B_1^X(0))}$ contains an open ball, say $B_s^Y(g)$. Set $r = s/2$.

Step 1. We will show that

$$B_r^Y(0) \ \subseteq \ \overline{A(B_1^X(0))}. \tag{9.11}$$

To see this, choose a vector $h \in B_r^Y(0)$. Then $\|h\|_X < r = s/2$, so

$$2h + g \ \in \ B_s^Y(g) \ \subseteq \ \overline{A(B_1^X(0))}.$$

Therefore, there exist vectors $f_n \in X$ with $\|f_n\|_X < 1$ such that $Af_n \to 2h+g$. Also,

$$g \ \in \ B_s^Y(g) \ \subseteq \ \overline{A(B_1^X(0))},$$

so there exist vectors $g_n \in X$ with $\|g_n\|_X < 1$ such that $Ag_n \to g$. Since f_n and g_n both have norm strictly less than 1, we have

$$w_n \ = \ \frac{f_n - g_n}{2} \ \in \ B_1^X(0),$$

and

$$Aw_n \ = \ \frac{Af_n - Ag_n}{2} \ \to \ \frac{(2h+g) - g}{2} \ = \ h \qquad \text{as } n \to \infty.$$

Therefore $h \in \overline{A(B_1^X(0))}$, so we have shown that equation (9.11) holds.

Now we will show that $B_{r/2}^Y(0) \subseteq A(B_1^X(0))$. To see this, choose any vector $h \in B_{r/2}^Y(0)$. Then $\|h\| < r/2$, so by rescaling equation (9.11) it follows that

$$h \ \in \ \overline{A(B_{1/2}^X(0))}.$$

Hence there must exist some vector $f_1 \in X$ such that $\|f_1\| < 1/2$ and

$$\|h - Af_1\| \ < \ \frac{r}{4}.$$

Consequently

$$h - Af_1 \ \in \ B_{r/4}^Y(0) \ \subseteq \ \overline{A(B_{1/4}^X(0))},$$

so we can can find some vector $f_2 \in X$ such that $\|f_2\| < 1/4$ and

$$\|(h - Af_1) - Af_2\| \ < \ \frac{r}{8}.$$

Continuing in this way, we obtain vectors $f_n \in X$ such that $\|f_n\| < 2^{-n}$ and

$$\|h - Ag_n\| \;<\; \frac{r}{2^{n+1}}, \qquad \text{where} \;\; g_n \;=\; \sum_{k=1}^{n} f_k. \tag{9.12}$$

It follows from equation (9.12) that $Ag_n \to h$. On the other hand, since $\|f_k\| < 2^{-k}$, the series

$$g \;=\; \sum_{k=1}^{\infty} f_k$$

converges absolutely. Since X is a Banach space, this series converges in X. As g_n is the nth partial sum of the series, we conclude that $g_n \to g$ as $n \to \infty$. Since A is continuous, it follows that $Ag_n \to Ag$, and therefore $h = Ag$. As $\|g\| < 1$, this implies that $h \in A(B_1^X(0))$, so we have shown that

$$B_{r/2}^Y(0) \;\subseteq\; A(B_1^X(0)). \qquad \square$$

Now we prove the Open Mapping Theorem.

Theorem 9.5.3 (Open Mapping Theorem). *If X, Y are Banach spaces and $A\colon X \to Y$ is a continuous linear surjection, then A is an open mapping.*

Proof. Since $X = \cup_{k=1}^{\infty} B_k^X(0)$ and A is a surjective map, we have $Y = \cup_{k=1}^{\infty} A(B_k^X(0))$, and therefore

$$Y \;=\; \bigcup_{k=1}^{\infty} \overline{A(B_k^X(0))}.$$

Since Y is a Banach space, the Baire Category Theorem implies that at least one set $\overline{A(B_k^X(0))}$ must contain an open ball. Since X is a Banach space, we can apply Lemma 9.5.2 and conclude that there is some $r > 0$ such that

$$B_r^Y(0) \;\subseteq\; A(B_1^X(0)). \tag{9.13}$$

Now let U be an open subset of X and fix $g \in A(U)$. Then $g = Af$ for some $f \in U$, so there must be some $s > 0$ such that $B_s^X(f) \subseteq U$. Rescaling equation (9.13), it follows that there is some $t > 0$ such that

$$B_t^Y(0) \;\subseteq\; A(B_s^X(0)).$$

Using the linearity of A, we compute that

$$B_t^Y(g) \;=\; B_t^Y(0) + Af \;\subseteq\; A(B_s^X(0) + f) \;=\; A(B_s^X(f)) \;\subseteq\; A(U).$$

Hence $A(U)$ is an open set. $\square$

The hypotheses in the Open Mapping Theorem that X and Y are both complete is necessary; see [Con90].

Often, the operator A that we wish to apply the Open Mapping Theorem to is a bijection and not merely surjective. In this case, the direct image of a set under A is the inverse image of the same set under the inverse mapping A^{-1}. Restating the Open Mapping Theorem for the case of bijections gives us the following useful result.

Theorem 9.5.4 (Inverse Mapping Theorem). *If X, Y are Banach spaces and $A\colon X \to Y$ is a continuous linear bijection, then $A^{-1}\colon Y \to X$ is continuous. Consequently A is a topological isomorphism.*

Proof. Let U be any open subset of X. The inverse image of U under A^{-1} is $(A^{-1})^{-1}(U) = A(U)$. This is an open set because the Open Mapping Theorem implies that A is an open mapping. $\square$

The next exercise is a typical application of the Inverse Mapping Theorem. The bijection in this exercise is the identity mapping I. However, there are two different Banach spaces with two different norms in the exercise, namely, $(X, \|\cdot\|)$ and $(X, |||\cdot|||)$.

Exercise 9.5.5. Suppose X is a vector space that is complete with respect to two norms $\|\cdot\|$ and $|||\cdot|||$. Suppose that there exists a constant $C > 0$ such that $\|f\| \le C\,|||f|||$ for all $f \in X$. Show that $\|\cdot\|$ and $|||\cdot|||$ are equivalent norms on X. $\diamondsuit$

We will use the Inverse Mapping Theorem to derive two results for operators on Hilbert spaces. The following theorem shows that a bounded linear operator has closed range if and only if its adjoint has closed range (this also holds for operators on Banach space, see [Rud91, Thm. 4.14]).

Theorem 9.5.6. *Fix $A \in \mathcal{B}(H,K)$, where H and K are Hilbert spaces. Then*

$$\mathrm{range}(A) \text{ is closed} \quad \Longleftrightarrow \quad \mathrm{range}(A^*) \text{ is closed.}$$

Proof. $\Leftarrow$. Suppose that $\mathrm{range}(A^*)$ is closed, and let $M = \overline{\mathrm{range}}(A)$. Define $T \in \mathcal{B}(H,M)$ by $Tx = Ax$ for $x \in H$. Since $\mathrm{range}(T)$ is dense in M, Theorem 8.2.6 implies that $T^*\colon M \to H$ is injective. Given $y \in K$, write $y = m + e$ where $m \in M$ and $e \in M^\perp$. Since $\ker(A^*) = \mathrm{range}(A)^\perp = M^\perp$, for any $x \in H$ we have

$$\langle x, A^*y\rangle \;=\; \langle x, A^*m\rangle \;=\; \langle Ax, m\rangle \;=\; \langle Tx, m\rangle \;=\; \langle x, T^*m\rangle.$$

Hence $A^*y = A^*m = T^*m$, and it follows from this that $\mathrm{range}(T^*) = \mathrm{range}(A^*)$, which is closed. Now set $N = \mathrm{range}(T^*)$ and define $U \in \mathcal{B}(M,N)$ by $Uy = T^*y$ for $y \in M$. Then U is a continuous bijection, so it is a topological isomorphism by the Inverse Mapping Theorem. Problem 9.2.27 therefore implies that $U^* \in \mathcal{B}(N,M)$ is a topological isomorphism. In particular, $\mathrm{range}(U^*) = M$ is closed.

Fix $y \in M$, so $y = U^*x$ for some $x \in N$. Let z be any vector in K, and let p be its orthogonal projection onto M. Then, since Ax, U^*x, and y all belong to M,

$$\begin{aligned}
\langle y, z\rangle &= \langle U^*x, z\rangle \\
&= \langle U^*x, p\rangle \\
&= \langle x, Up\rangle \\
&= \langle x, T^*p\rangle \\
&= \langle Tx, p\rangle \\
&= \langle Ax, p\rangle \;=\; \langle Ax, z\rangle.
\end{aligned}$$

Therefore $y = Ax$, so $M \subseteq \mathrm{range}(A)$ and hence $\mathrm{range}(A) = M$ is closed.

$\Rightarrow$. Since $(A^*)^* = A$, this follows from the previous case. $\square$

Our next application of the Inverse Mapping Theorem constructs a "pseudoinverse" of a bounded operator A that has closed range. Although A need not be injective, the pseudoinverse $A^\dagger$ acts as a right-inverse of A, at least when we restrict the domain of $A^\dagger$ to $\mathrm{range}(A)$.

Theorem 9.5.7. *Let H and K be Hilbert spaces. Assume that $A \in \mathcal{B}(H,K)$ has closed range, and let P be the orthogonal projection of K onto* $\mathrm{range}(A)$. *Then the mapping $B\colon \ker(A)^\perp \to \mathrm{range}(A)$ defined by $Bx = Ax$ for $x \in \ker(A)^\perp$ is a topological isomorphism, and $A^\dagger = B^{-1}P \in \mathcal{B}(K,H)$ satisfies the following:*

(a) $AA^\dagger y = y$ *for every* $y \in \mathrm{range}(A)$,

(b) $AA^\dagger$ *is the orthogonal projection of K onto* $\mathrm{range}(A)$, *and*

(c) $A^\dagger A$ *is the orthogonal projection of H onto* $\mathrm{range}(A^*)$.

Proof. The mapping B is bounded and linear since it is a restriction of the bounded mapping A. Further, the fact that $H = \ker(A) \oplus \ker(A)^\perp$ implies that B is a bijection of $\ker(A)^\perp$ onto $\mathrm{range}(A)$. Applying the Inverse Mapping Theorem, we conclude that $B\colon \ker(A)^\perp \to \mathrm{range}(A)$ is a topological isomorphism. Hence $B^{-1}\colon \mathrm{range}(A) \to \ker(A)^\perp$ is a topological isomorphism, and therefore $A^\dagger = B^{-1}P$ is bounded.

Exercise: Finish the proof of statements (a)–(c). $\square$

Definition 9.5.8 (Pseudoinverse). Given $A \in \mathcal{B}(H,K)$, the operator $A^\dagger$ constructed in Theorem 9.5.7 is called the *Moore–Penrose pseudoinverse*, or simply the *pseudoinverse*, of A. $\diamondsuit$

Exercise 9.5.12 gives an equivalent characterization of the pseudoinverse.

Problems

9.5.9. Let X and Y be Banach spaces, and suppose that $A \in \mathcal{B}(X,Y)$ is injective and range(A) is a closed subset of Y. Prove that there exists a constant $c > 0$ such that $\|Af\| \ge c\,\|f\|$ for all $f \in X$ (see Problem 6.2.10 for the converse result).

9.5.10. Let X and Y be Banach spaces. Prove that $A \in \mathcal{B}(X,Y)$ is surjective if and only if range(A) is not a meager subset of Y.

9.5.11. Let $A \in \mathcal{B}(H)$ be a fixed positive definite operator on H. Prove the following statements.

(a) A is injective and has dense range.

(b) A is a topological isomorphism that maps H onto itself if and only if A is surjective. Show by example that a positive definite operator need not be a topological isomorphism.

(c) If A is a surjective positive definite operator, then $(f,g) = \langle Af, g\rangle$ defines an inner product on H that is equivalent to the original inner product $\langle\cdot,\cdot\rangle$.

9.5.12. Assume that $A \in \mathcal{B}(H,K)$ has closed range, and let $A^\dagger$ be its pseudoinverse. Prove the following statements.

(a) $\ker(A^\dagger) = \mathrm{range}(A)^\perp$.

(b) $\mathrm{range}(A^\dagger) = \ker(A)^\perp$.

(c) $AA^\dagger y = y$ for all $y \in \mathrm{range}(A)$.

(d) $A^\dagger$ is the unique operator in $\mathcal{B}(K,H)$ that satisfies statements (a)–(c) above.

9.6 The Closed Graph Theorem

The Closed Graph Theorem, which we prove next, gives us some equivalent reformulations for continuity of linear operators that map one Banach space into another. Look carefully at the statement of this result, especially comparing statements (a) and (c) of Theorem 9.6.1. Obviously, statement (a) implies statement (c), but it is rather surprising that statement (c) implies statement (a). Normally, to prove that $A\colon X \to Y$ is continuous, we would assume that $f_n \to f$ and prove that Af_n converges to Af. Statement (c) tells us that we can instead assume both that $f_n \to f$ *and* that Af_n converges to some vector g, and simply show that g equals Af. The caveat is that in order to apply the Closed Graph Theorem, we must know that X and Y are complete and A is linear.

Theorem 9.6.1 (Closed Graph Theorem). *Let X and Y be Banach spaces. If $A\colon X \to Y$ is linear, then the following three statements are equivalent.*

(a) *A is continuous.*

(b) $\operatorname{graph}(A) = \bigl\{(f, Af) : f \in X\bigr\}$ *is a closed subset of $X \times Y$.*

(c) *If $f_n \to f$ in X and $Af_n \to g$ in Y, then $g = Af$.*

Proof. (c) $\Rightarrow$ (a). Assume that statement (c) holds. For clarity, write $\|\cdot\|_X$ and $\|\cdot\|_Y$ for the norms on X and Y, and define

$$|||f||| \;=\; \|f\|_X + \|Af\|_Y, \qquad f \in X.$$

Exercise: Show that $|||\cdot|||$ is a norm on X, and X is complete with respect to this norm.

Since we have $\|f\|_X \le |||f|||$ for $f \in X$ and since X is complete with respect to both norms, it follows from Exercise 9.5.5 that there exists a constant $C > 0$ such that $|||f||| \le C\,\|f\|_X$ for $f \in X$. Consequently, $\|Af\|_Y \le C\,\|f\|_X$, so A is bounded.

Exercise: Prove the remaining implications. □

Problem 9.6.3 shows that the hypotheses in the Closed Graph Theorem that X is complete is necessary, and it can be shown that it is also necessary that Y be complete.

Problems

9.6.2. Use the Closed Graph Theorem to give another proof of Theorem 9.4.3.

9.6.3. Consider the spaces $C_b(\mathbb{R})$ and $C_b^1(\mathbb{R})$, and assume that the norm of both of these spaces is the uniform norm. With respect to this norm, $C_b(\mathbb{R})$ is complete while $C_b^1(\mathbb{R})$ is not. Prove that the differentiation operator $D\colon C_b^1(\mathbb{R}) \to C_b(\mathbb{R})$ given by $Df = f'$ is unbounded, but it has a closed graph, i.e., if $f_n \to f$ uniformly and $f_n' \to g$ uniformly then $f' = g$.

9.6.4. Let H be a Hilbert space. Suppose that a function $A\colon H \to H$ satisfies $\langle Af, g\rangle = \langle f, Ag\rangle$ for all $f, g \in H$. Prove that A is linear and continuous, and hence is a bounded self-adjoint operator on H.

9.7 Schauder Bases

Schauder bases were introduced earlier (see Definition 5.4.8). Essentially, a Schauder basis for an infinite-dimensional Banach space is a countable sequence that has the property that every vector in the space can be expressed as a "unique infinite linear combination" of the basis elements. More precisely, we recall from Definition 5.4.8 that a countable sequence $\{f_n\}_{n\in\mathbb{N}}$ is a *Schauder basis* for a Banach space X if every vector $f \in X$ can be written as

$$f \;=\; \sum_{n=1}^{\infty} \alpha_n(f)\, f_n \tag{9.14}$$

for a unique choice of scalars $\alpha_n(f)$, where the series converges in the norm of X. An analogous definition holds for finite-dimensional spaces using finite sums instead of infinite sums; we will focus on the infinite-dimensional setting here.

Some basic facts about Schauder bases were derived in Chapter 5. For example, Theorem 5.4.3 states that if a Banach space X has a Schauder basis, then X must be separable (however, not every separable Banach space has a Schauder basis!). Now we will apply the functional analysis theorems that we proved in the last few sections to derive additional properties of Schauder bases and related systems.

9.7.1 Continuity of the Coefficient Functionals

Assume that $\{f_n\}_{n\in\mathbb{N}}$ is a Schauder basis for a Banach space X. Then, by equation (9.14), for each $f \in X$ there exist unique scalars $\alpha_n(f)$ such that $f = \sum_{n=1}^{\infty} \alpha_n(f)\, f_n$. If we fix a particular n, then the mapping α_n that sends a vector f to the scalar $\alpha_n(f)$ is a functional on X. Furthermore, it follows directly from the uniqueness requirement of a Schauder basis that α_n is linear. We call $\{\alpha_n\}_{n\in\mathbb{N}}$ the *associated family of coefficient functionals.*

Each functional α_n is linear, but the definition of a Schauder basis does not explicitly require that α_n be bounded. Our first goal is to present an extended exercise that will prove (among other things) that each functional α_n *must be* continuous, and hence belongs to the dual space X^*.

Before stating this exercise, we make a few observations. If we fix an integer $m \in \mathbb{N}$, then we can express the vector f_m in the following two ways:

$$\sum_{n=1}^{\infty} \delta_{mn}\, f_n \;=\; f_m \;=\; \sum_{n=1}^{\infty} \alpha_n(f_m)\, f_n.$$

Since we have assumed that $\{f_n\}$ is a Schauder basis, the uniqueness requirement implies that

$$\alpha_n(f_m) \;=\; \delta_{mn}. \tag{9.15}$$

Sequences that satisfy equation (9.15) are said to be *biorthogonal* (we will formalize this terminology in Definition 9.7.2), so we say that $\{f_n\}_{n\in\mathbb{N}}$ and $\{\alpha_n\}_{n\in\mathbb{N}}$ are *biorthogonal sequences.*

For each integer $N \in \mathbb{N}$, define an operator $S_N \colon X \to X$ by

$$S_N f \;=\; \sum_{n=1}^{N} \alpha_n(f)\, f_n, \qquad f \in X.$$

We call these operators S_N the *partial sum operators* associated with the Schauder basis $\{f_n\}_{n\in\mathbb{N}}$. Since S_N is defined by a finite sum, it is well-defined on all of X, and it is linear since each functional α_n is linear. By construction, the range of S_N is contained in $\operatorname{span}\{f_1,\dots,f_N\}$, so S_N is linear and has finite rank. The biorthogonality stated in equation (9.15) implies that $S_N f_n = f_n$ for $n = 1,\dots,N$, so the range of S_N is precisely

$$\operatorname{range}(S_N) \;=\; \operatorname{span}\{f_1,\dots,f_N\}.$$

Thus S_N is a linear, finite-rank operator on X, and $\operatorname{rank}(S_N) = N$. Moreover, for each $f \in X$ it follows from equation (9.14) that

$$S_N f \to f \quad \text{ as } N \to \infty.$$

This seems like an invitation to apply the Banach–Steinhaus Theorem (Exercise 9.4.2). Unfortunately, in order to apply Banach–Steinhaus, we would need to know that S_N is bounded, and we do not know this because we do not know whether the functionals α_n are bounded (indeed, this is exactly what we hope to prove).

The next exercise lays out one approach to proving that S_N is continuous (from which we also obtain the the continuity of the functionals α_n). The key ingredient is to show that

$$|\!|\!|f|\!|\!| \;=\; \sup_N \|S_N f\|, \qquad f \in X,$$

defines a norm on X that is equivalent to the original norm on X. In solving this exercise, the reader should be careful not to assume that α_n or S_N is bounded.

Exercise 9.7.1. Let $\{f_n\}_{n\in\mathbb{N}}$ be a Schauder basis for a Banach space X.

(a) Show that $|\!|\!|\cdot|\!|\!|$ is a norm on X, and $\|f\| \le |\!|\!|f|\!|\!|$ for all $f \in X$.

(b) Suppose that $\{g_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in X with respect to $|\!|\!|\cdot|\!|\!|$. With N fixed, show that $\{S_N g_n\}_{n\in\mathbb{N}}$ is Cauchy with respect to $\|\cdot\|$. Let h_N be such that $\|h_N - S_N g_n\| \to 0$ as $n \to \infty$, and observe that $h_N \in \operatorname{span}\{f_1,\dots,f_N\}$.

(c) Show that

$$\lim_{n\to\infty} \Bigl(\sup_N \|h_N - S_N g_n\|\Bigr) \;=\; 0,$$

and use this to show that $\{h_N\}_{N\in\mathbb{N}}$ is Cauchy with respect to $\|\cdot\|$. Let $g \in X$ be the element such that $\|g - h_N\| \to 0$.

(d) Show that $S_N(h_{N+1}) = h_N$, and use this to show that $h_N = \sum_{n=1}^N c_n f_n$ where the scalars c_n are independent of N.

(e) Show that $g = \sum_{n=1}^\infty c_n f_n$, and hence $h_N = S_N g$. Use this to show that $|||g - g_n||| \to 0$, and conclude that X is complete with respect to $|||\cdot|||$.

(f) Show that $\|\cdot\|$ and $|||\cdot|||$ are equivalent norms on X (the Inverse Mapping Theorem or the Uniform Boundedness Principle may be helpful at this point). Use this to show that

$$C = \sup_N \|S_N\| < \infty.$$

The number C is called the *basis constant* for $\{f_n\}_{n\in\mathbb{N}}$.

(g) Show that $1 \le \|\alpha_n\| \, \|f_n\| \le 2C$ for each $n \in \mathbb{N}$. $\diamondsuit$

Now that we know that the coefficient functionals associated with a Schauder basis are continuous, we adopt our standard functional notation and write $\langle f, \alpha_n \rangle$ instead of $\alpha_n(f)$, with the understanding that this functional notation is linear in the first variable but antilinear in the second (compare Notation 9.1.1).

9.7.2 Minimal Sequences

As we have seen, if $\{f_n\}$ is a Schauder basis with coefficient functionals $\{\alpha_n\}$, then we have the biorthogonality condition $\langle f_m, \alpha_n \rangle = \delta_{mn}$. Thus, a sequence $\{f_n\} \subseteq X$ that is a Schauder basis has a biorthogonal sequence $\{\alpha_n\} \subseteq X^*$. Unfortunately, in the converse direction, a sequence $\{f_n\}$ that has a biorthogonal sequence need not be a Schauder basis (see Problem 9.7.7). Therefore we will examine biorthogonal sequences in more detail. We give next the formal definition of biorthogonal sequences, and we also introduces a type of "generalized linear independence" property called *minimality*.

Definition 9.7.2 (Minimal and Biorthogonal Sequences). Let X be a Banach space, and let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence in X.

(a) $\{f_n\}_{n\in\mathbb{N}}$ is *minimal* if no vector f_m lies in the closed span of the other f_n:

$$\forall\, m \in \mathbb{N}, \quad f_m \notin \overline{\operatorname{span}}\big(\{f_n\}_{n\neq m}\big).$$

(b) A sequence $\{\mu_n\}_{n\in\mathbb{N}}$ in X^* is *biorthogonal* to $\{f_n\}_{n\in\mathbb{N}}$ if $\langle f_m, \mu_n \rangle = \delta_{mn}$ for all m, $n \in \mathbb{N}$.

(c) A sequence that is both minimal and complete is said to be *exact*. $\diamondsuit$

The following exercise will show that a sequence $\{f_n\}$ has a biorthogonal sequence if and only if $\{f_n\}$ is minimal. This exercise takes place in a generic

Banach space. However, the special case for Hilbert spaces actually appeared earlier (see Problem 5.6.17). In the setting of Hilbert spaces the proof is not difficult, but it relies on the fact that a closed subspace of a Hilbert space has an orthogonal complement in the space. This fact does not hold in general Banach spaces. Instead, the correct tool to use in the Banach space setting is the Hahn–Banach Theorem.

Exercise 9.7.3. Given a sequence $\{f_n\}_{n\in\mathbb{N}}$ in a Banach space X, prove that the following two statements are equivalent.

(a) $\{f_n\}_{n\in\mathbb{N}}$ is minimal.

(b) There exists a sequence $\{\mu_n\}_{n\in\mathbb{N}}$ in X^* that is biorthogonal to $\{f_n\}_{n\in\mathbb{N}}$.

Additionally, prove that the following two statements are equivalent.

(a') $\{f_n\}_{n\in\mathbb{N}}$ is exact (both minimal and complete).

(b') There exists a unique sequence $\{\mu_n\}_{n\in\mathbb{N}}$ in X^* that is biorthogonal to $\{f_n\}_{n\in\mathbb{N}}$. $\diamondsuit$

In particular, every Schauder basis is exact since it is complete and possesses a biorthogonal sequence in X^*. However, we will see below that an exact sequence need not be a Schauder basis.

In finite dimensions, a sequence $\{x_1, \dots, x_n\}$ is minimal if and only if it is linearly independent, and in this case it is a basis for its span. These simple facts do not extend to infinite sequences. Instead, there are several "shades of grey" in the meaning of independence in infinite-dimensional spaces, as illustrated in the following exercise and examples (see Problem 5.6.16 for the definition of an *ω-independent* sequence).

Exercise 9.7.4. Let $\{f_n\}$ be a countable sequence in a Banach space X, and prove the following implications:

$$\text{Schauder basis} \implies \text{minimal} \implies \omega\text{-independent} \implies \begin{array}{c}\text{finitely}\\ \text{independent.}\end{array} \quad \diamondsuit$$

The following examples show that none of the converse implications hold in general, even in a Hilbert space, and and even if we add the assumption that the sequence is complete.

Example 9.7.5. (a) The trigonometric system $\{e^{2\pi inx}\}_{n\in\mathbb{Z}}$ is an orthonormal basis for $L^2[0,1]$. Problem 9.7.7 shows that the sequence $\{xe^{2\pi inx}\}_{n\neq 0}$ is both minimal and complete in $L^2[0,1]$, but is not a Schauder basis for $L^2[0,1]$.

(b) Let $\{e_n\}_{n\in\mathbb{N}}$ be an orthonormal basis for a Hilbert space H. Set $f_1 = e_1$, and for $n > 2$ define $f_n = e_1 + (e_n/n)$. According to Problem 9.7.8, the sequence $\{f_n\}_{n\in\mathbb{N}}$ is ω-independent and complete in H, but it is not minimal.

(c) Problem 5.6.16 constructed a sequence in ℓ^2 that is finitely linearly independent and complete in ℓ^2, but is not ω-independent. $\diamondsuit$

9.7.3 A Characterization of Schauder Bases

We have seen that an arbitrary exact sequence need not be a Schauder basis. Now we will determine exactly what extra property an exact system must possess in order to be a Schauder basis.

Suppose that $\{f_n\}_{n\in\mathbb{N}}$ is a minimal sequence in a Banach space X. Then, by Exercise 9.7.3, there exists a sequence $\{\alpha_n\}$ in X^* that is biorthogonal to $\{f_n\}$ Even though we do not know whether $\{f_n\}$ is a Schauder basis, we can still define partial sum operators $S_N\colon X \to X$ by

$$S_N f \;=\; \sum_{n=1}^{N} \langle f, \alpha_n\rangle\, f_n, \qquad f \in X.$$

Since each α_n is continuous (by definition of minimal sequence), each operator S_N is a bounded on X. *If* $S_N f \to f$, then $f = \sum_{n=1}^{\infty} \langle f, \alpha_n\rangle\, f_n$, and in this case the minimality ensures that this representation is unique. The question is: When will the partial sums converge? This is answered in the next exercise; note that the proof of the implication (c) $\Rightarrow$ (d) is an application of the Uniform Boundedness Principle.

Exercise 9.7.6. Let $\{f_n\}_{n\in\mathbb{N}}$ be a minimal sequence in a Banach space X, and let $\{\alpha_n\}_{n\in\mathbb{N}}$ be its biorthogonal sequence in X^*. Prove that the following four statements are equivalent.

(a) $\{f_n\}_{n\in\mathbb{N}}$ is a Schauder basis for X.

(b) $S_N f \to f$ for each $f \in X$.

(c) $\{f_n\}_{n\in\mathbb{N}}$ is complete, and $\sup \|S_N f\| < \infty$ for each $f \in X$.

(d) $\{f_n\}_{n\in\mathbb{N}}$ is complete, and $\sup \|S_N\| < \infty$. $\diamondsuit$

Consequently, a Schauder basis is precisely an exact sequence whose *basis constant*

$$C \;=\; \sup_N \|S_N\|$$

is finite.

9.7.4 Unconditional Bases

An *unconditional basis* for a Banach space X is a Schauder basis $\{f_n\}$ which has the property that the representations $f = \sum \langle f, \alpha_n\rangle\, f_n$ converge unconditionally for each $f \in X$.

Not every Schauder basis is unconditional. For example, the trigonometric system $\{e^{2\pi inx}\}_{n\in\mathbb{Z}}$ is an orthonormal and hence unconditional basis for $L^2[0,1]$. It is also a Schauder basis for $L^p[0,1]$ for each index $1 < p < \infty$, but it is not an unconditional basis when $p \neq 2$ (for a proof, see [Heil11, Chap. 14]). In contrast, the *Haar system*

$$\{\chi_{[0,1]}\} \;\cup\; \{2^{n/2}\psi(2^n x - k)\}_{n\ge 0,\, k=0,\dots,2^n-1},$$

where $\psi = \chi_{[0,1/2)} - \chi_{[1/2,1)}$, is an unconditional basis for $L^p[0,1]$ for each $1 < p < \infty$.

A Hilbert space example of a Schauder basis that is not unconditional is the sequence

$$\bigl\{\bigl|x - \tfrac{1}{2}\bigr|^{1/4} e^{2\pi i n x}\bigr\}_{n\in\mathbb{Z}}$$

in $L^2[0,1]$ (see [Bab48] and the discussion in [Sin70, pp. 351–354]).

Problems

9.7.7. (a) Show that the sequence $\{xe^{2\pi i n x}\}_{n\neq 0}$ is exact in $L^2[0,1]$, and find its biorthogonal system $\{g_n\}_{n\neq 0}$ in $L^2[0,1]$.

(b) Show that $\sup \|g_n\|_2 = \infty$, and use this to prove that $\{xe^{2\pi i n x}\}_{n\neq 0}$ is not a Schauder basis for $L^2[0,1]$, no matter what ordering of $\mathbb{Z}\backslash\{0\}$ that we choose.

9.7.8. Let $\{e_n\}_{n\in\mathbb{N}}$ be an orthonormal basis for a Hilbert space H. Set $f_1 = e_1$ and $f_n = e_1 + (e_n/n)$ for $n > 2$. Prove that $\{f_n\}$ is ω-independent and complete, but not minimal in H.

9.7.9. (a) Prove that the standard basis $\{\delta_n\}_{n\in\mathbb{N}}$ is an unconditional basis for ℓ^p for each $1 \le p < \infty$, and is also an unconditional basis for c_0.

(b) Recall from Problem 6.10.8 that

$$c \;=\; \bigl\{x = (x_k)_{k\in\mathbb{N}} \in \ell^\infty : \lim_{k\to\infty} x_k \text{ exists}\bigr\}$$

is a closed subspace of ℓ^∞. Find a vector $\delta_0 \in c$ such that $\{\delta_n\}_{n\ge 0}$ is an unconditional basis for c.

(c) Prove that ${c_0}^*$ and c^* are each isometrically isomorphic to ℓ^1.

(d) Prove that c and c_0 are topologically isomorphic.

(e) Show that c and c_0 are not isometrically isomorphic.

9.7.10. For each $n \in \mathbb{N}$, define $y_n = (1, \dots, 1, 0, 0, \dots)$, where the 1 is repeated n times. Show that $\{y_n\}_{n\in\mathbb{N}}$ is a conditional basis for c_0.

9.8 Weak and Weak* Convergence

In this section we will discuss weak and weak* convergence of sequences. A complete discussion of the weak and weak* topologies requires consideration of convergence of *nets*. For the special case of *sequences*, the definition is as follows.

Definition 9.8.1. Let X be a normed linear space.

(a) A sequence $\{f_n\}_{n\in\mathbb{N}}$ in X *converges weakly* to $f \in X$, denoted $f_n \xrightarrow{\text{w}} f$, if

$$\forall\, \mu \in X^*, \quad \lim_{n\to\infty} \langle f_n, \mu\rangle \;=\; \langle f, \mu\rangle.$$

(b) A sequence $\{\mu_n\}_{n\in\mathbb{N}}$ in X^* *converges weak** to $\mu \in X^*$, denoted $\mu_n \xrightarrow{\text{w}^*} \mu$, if

$$\forall\, f \in X, \quad \lim_{n\to\infty} \langle f, \mu_n\rangle \;=\; \langle f, \mu\rangle. \qquad \diamondsuit$$

Thus, weak* convergence is just "pointwise convergence" of the operators μ_n. Weak* convergence is only defined for sequences that lie in a dual space X^*. Given a sequence $\{\mu_n\}_{n\in\mathbb{N}}$ in X^*, we can consider three types of convergence: strong (norm), weak, and weak* convergence. By definition, these are:

$$\begin{aligned}
\mu_n \to \mu \quad &\iff \quad \lim_{n\to\infty} \|\mu - \mu_n\| \;=\; 0,\\
\mu_n \xrightarrow{\text{w}} \mu \quad &\iff \quad \forall\, T \in X^{**}, \quad \lim_{n\to\infty} \langle \mu_n, T\rangle \;=\; \langle \mu, T\rangle,\\
\mu_n \xrightarrow{\text{w}^*} \mu \quad &\iff \quad \forall\, x \in X, \quad \lim_{n\to\infty} \langle x, \mu_n\rangle \;=\; \langle x, \mu\rangle.
\end{aligned}$$

The next exercise shows that strong convergence implies weak convergence, and (if applicable) weak convergence implies weak* convergence.

Exercise 9.8.2. Let X be a normed space and suppose that $f_n,\, f \in X$ and $\mu_n,\, \mu \in X^*$. Show that

$$f_n \to f \implies f_n \xrightarrow{\text{w}} f$$

and

$$\mu_n \to \mu \quad\implies\quad \mu_n \xrightarrow{\text{w}} \mu \quad\implies\quad \mu_n \xrightarrow{\text{w}^*} \mu. \qquad \diamondsuit$$

If X is reflexive, then weak and weak* convergence coincide. In particular, this is the case for Hilbert spaces. In any finite-dimensional space, strong and weak convergence coincide, but they are distinct in any infinite-dimensional space. For example, if $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal sequence in a Hilbert space H, then it follows from Bessel's Inequality that $\langle e_n, f\rangle \to 0$ for each $f \in H$ and hence $e_n \xrightarrow{\text{w}} 0$, but $\|e_n - 0\| = 1 \not\to 0$.

It is trivial to show that any strongly convergent sequence is bounded. The corresponding fact for weak and weak* convergent sequences is more subtle and is most easily proved by applying the Uniform Boundedness Principle.

Exercise 9.8.3. Let X be a Banach space.

(a) Show that if $\mu_n \xrightarrow{\text{w}^*} \mu$ in X^*, then μ is unique and $\sup \|\mu_n\| < \infty$.

(b) Show that if $f_n \xrightarrow{\text{w}} f$ in X, then f is unique and $\sup \|f_n\| < \infty$. $\diamondsuit$

Problems

9.8.4. Show that if $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal sequence in a Hilbert space H, then $e_n \xrightarrow{\text{w}} 0$ as $n \to \infty$.

9.8.5. (a) Given $1 < p < \infty$ and $x_n, y \in \ell^p$, show that $x_n \xrightarrow{\text{w}} y$ in ℓ^p if and only if $x_n(k) \to y(k)$ for each k (componentwise convergence) and $\sup \|x_n\|_p < \infty$. Does either implication remain valid if $p = 1$?

(b) Given $x_n, y \in \ell^1$, show that $x_n \xrightarrow{\text{w}^*} y$ in ℓ^1 if and only if $x_n(k) \to y(k)$ for each k and $\sup \|x_n\|_1 < \infty$.

(c) Given $x_n, y \in \ell^\infty$, show that $x_n \xrightarrow{\text{w}^*} y$ in ℓ^∞ if and only if $x_n(k) \to y(k)$ for each k and $\sup \|x_n\|_\infty < \infty$.

9.8.6. (a) Given $1 < p < \infty$ and $f_n, g \in L^p(\mathbb{R})$, show that $f_n \xrightarrow{\text{w}} g$ if and only if $\sup \|f_n\|_p < \infty$ and $\int_E f_n \to \int_E g$ for every $E \subseteq \mathbb{R}$ with $|E| < \infty$. What happens if $p = 1$ or $p = \infty$?

(b) Given $f_n, g \in L^\infty(\mathbb{R})$, show that $f_n \xrightarrow{\text{w}^*} g$ if and only if $f_n \to g$ pointwise a.e. and $\sup \|f_n\|_\infty < \infty$.

Hints for Selected Exercises and Problems

1.1.16 If Σ contains infinitely many *disjoint* sets $E_1, E_2, \dots$ then Σ must be uncountable. One method of showing the existence of such sets is to define a relation on X by declaring $x \sim y$ if and only if

$$\forall A \in \Sigma, \quad x \in A \Leftrightarrow y \in A.$$

Prove that $\sim$ is an equivalence relation, and show that if Σ is countable then the equivalence classes $[x] = \cap\{A \in \Sigma : x \in A\}$ all belong to Σ.

1.1.18 Let $X = \{x_n\}_{n\in\mathbb{N}}$. Let Σ_N consist of every set $A \in \mathcal{P}(\{x_1, \dots, x_N\})$ together with the complement of each such set A.

1.3.11 Write $E = \cup E_n$ where $E_n = \{x \in X : \mu\{x\} > \frac{1}{n}\}$. How many points can be in E_n?

1.3.13 Define

$$C = \sup\{\mu(F) : F \in \Sigma, F \subseteq E, \mu(F) < \infty\}.$$

If $C < \infty$ then there exist measurable sets $F_k \subseteq E$ with finite measure such that $\mu(F_k) \to C$. Consider the sets $F = \cup F_k$ and $A = E \backslash F$.

1.6.9 Hint: Apply Problem 1.6.8 to both E and E^{C}.

2.1.3 Let $\mathcal{F} = \{F \subseteq Y : f^{-1}(F) \in \Sigma_X\}$, and show that $\mathcal{F}$ is a σ-algebra on Y.

2.1.22 (a) $\Leftrightarrow$ (b). Show that $\Sigma = \big\{E \subseteq \overline{\mathbb{R}} : E \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}\big\}$ is a σ-algebra on $\overline{\mathbb{R}}$.

2.1.28 Given an arbitrary $a \in \mathbb{R}$, find $a_k \in A$ such that $\{f > a\} = \cup\{f > a_k\}$.

2.1.29 For a counterexample, let N be a nonmeasurable subset of $\mathbb{R}$ and define $f \colon \mathbb{R} \to \mathbb{R}$ by $f(x) = x$ if $x \in N$, and $f(x) = -x$ if $x \notin N$.

2.1.31 Consider $E_k = \{f > k\}$.

2.2.12 Prove the results for nonnegative functions first, and then write $f = f^+ - f^-$, $g = g^+ - g^-$.

2.4.10 (c) Proceed through cases. Case 1: $f = g = 0$ a.e. Case 2: $g_k = g$ for every k. Case 3: Arbtrary f_k, g_k.

2.4.12 To prove the Triangle Inequality, first show that if a, b, $c \geq 0$ and $a \leq b + c$, then

$$\frac{a}{1+a} \ \leq \ \frac{b}{1+b} + \frac{c}{1+c}.$$

References

[Bab48] K. I. Babenko, On conjugate functions (Russian), *Doklady Akad. Nauk SSSR (N. S.)*, **62** (1948), pp. 157–160.

[Ben97] J. J. Benedetto, *Harmonic Analysis and Applications*, CRC Press, Boca Raton, FL, 1997.

[BC09] J. J. Benedetto and W. Czaja, *Integration and Modern Analysis*, Birkhäuser, Boston, 2009.

[BBT97] A. M. Bruckner, J. B. Bruckner, and B. S. Thomson, *Real Analysis*, Prentice-Hall, Upper Saddle River, NJ, 1997.

[Cho83] M. N. Choi, Tricks or treats with the Hilbert matrix. *Amer. Math. Monthly*, **90** (1983), pp. 301–312.

[Con90] J. B. Conway, *A Course in Functional Analysis*, Second Edition, Springer-Verlag, New York, 1990.

[Con00] J. B. Conway, *A Course in Operator Theory*, American Mathematical Society, Providence, RI, 2000.

[DM72] H. Dym and H. P. McKean, *Fourier Series and Integrals*, Academic Press, New York–London, 1972.

[Fol99] G. B. Folland, *Real Analysis*, Second Edition, Wiley, New York, 1999.

[Heil11] C. Heil, *A Basis Theory Primer*, Expanded Edition, Birkhäuser, Boston, 2011.

[Heil17] C. Heil, *Metrics, Norms, Inner Products, and Operator Theory*, Birkhäuser, Boston, 2017 (to appear).

[Heil18] C. Heil, *Introduction to Real Analysis*, Birkhäuser, Boston, 2018? (to appear).

[Her85] D. A. Herrero, From my chest of examples of Fourier transforms, *Rev. Un. Mat. Argentina*, **32** (1985), pp. 41–47.

[Jam51] R. C. James, A non-reflexive Banach space isometric with its second conjugate space, *Proc. Nat. Acad. Sci. U. S. A.*, **37** (1951), pp. 174–177.

[Kat04] Y. Katznelson, *An Introduction to Harmonic Analysis*, Third Edition, Cambridge University Press, Cambridge, UK, 2004.

[Mur37] F. J. Murray, On complementary manifolds and projections in spaces L_p and l_p, *Trans. Amer. Math. Soc.*, **41** (1937), pp. 138–152.

[Phi40] R. S. Phillips, On linear transformations, *Trans. Amer. Math. Soc.*, **48** (1940), pp. 516–541.

[Rud76] W. Rudin, *Principles of Mathematical Analysis*, Third Edition. McGraw-Hill, New York, 1976.

[Rud87] W. Rudin, *Real and Complex Analysis*, Third Edition. McGraw-Hill, New York, 1987.

[Rud91] W. Rudin, *Functional Analysis*, Second Edition. McGraw-Hill, New York, 1991.

[Sch11] I. Schur, Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlichvielen Veränderlichen [Remarks on the theory of bounded bilinear forms with infinitely many variables], *J. für Reine und Angew. Math.*, **140** (1911), pp. 1–28.

[Sin70] I. Singer, *Bases in Banach Spaces* I, Springer-Verlag, New York, 1970.

[SS05] E. M. Stein and R. Shakarchi, *Real Analysis*, Princeton University Press, Princeton, NJ, 2005.

[WZ77] R. L. Wheeden and A. Zygmund, *Measure and Integral*, Marcel Dekker, New York-Basel, 1977.

Index