Real Analysis

Jesse Peterson

February 1, 2017

Contents

Chapter 1

Preliminaries

1.1 Sets

We assume that the reader is familiar with the basic language and concepts of set theory. We use the notation $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ to denote respectively the non-negative integers (including zero), the integers, the rational numbers, the real numbers, and the complex numbers. If $\mathcal{A}$ is a collection of sets then we denote their **union** by $\cup_{A \in \mathcal{A}} A = \{a \mid a \in A \text{ for some } A \in \mathcal{A}\}$, and their **intersection** by $\cap_{A \in \mathcal{A}} A = \{a \mid a \in A \text{ for all } A \in \mathcal{A}\}$. If the family of sets is indexed $\mathcal{A} = \{A_i\}_{i \in I}$ then we also denote the union and intersection respectively by $\cup_{i \in I} A_i$ and $\cap_{i \in I} A_i$. The **difference** of two sets A and B is $A \setminus B = \{a \mid a \in A \text{ and } a \notin B\}$, and their **symmetric difference** is $A \Delta B = (A \setminus B) \cup (B \setminus A)$. If A is a subset of a set X, and we write A^c for the **complement** of A in X, i.e., $A^c = X \setminus A$.

The **power set** of a set X is denoted by 2^X and is the collection of all subsets of X, i.e., $2^X = \{A \mid A \subset X\}$. The **Cartesian product** $X \times Y$ of two sets X and Y consists of all ordered pairs (x, y) such that $x \in X$ and $y \in Y$. A **function** (or **mapping**) $f : X \to Y$ from X to Y is a a subset of $X \times Y$ which has the property that for each $x \in X$ there exists a unique $y \in Y$ such that the pair (x, y) is contained in this subset. In this case we write $y = f(x)$ (or sometimes $y = f_x$) for each $x \in X$. If $A \subset X$ and $B \subset Y$, then the **image** of A is denoted by $f(A) = \{f(a) \mid a \in A\}$, and the **inverse image** of B is denoted by $f^{-1}(B) = \{x \in X \mid f(x) \in B\}$. If $f : X \to Y$ and $g : Y \to Z$, the **composition** of f and g is denoted by $g \circ f$, and is defined by the formula $(g \circ f)(x) = g(f(x))$.

A function f is **injective** (or **1-1**) if $f(x) = f(y)$ only when $x = y$, and f is **surjective** (or **onto**) if $f(X) = Y$. f is **bijective** if it is both injective and surjective, and in this case f has a unique **inverse map** $f^{-1} : Y \to X$ such that $f^{-1} \circ f$ and $f \circ f^{-1}$ are the identity maps on X and Y respectively.

A **sequence** in a set X is a function from $\mathbb{N}$ to X. If $f : \mathbb{N} \to X$ is a sequence and $g : \mathbb{N} \to \mathbb{N}$ is such that $g(n) < g(m)$ whenever $n < m$, then we say that $f \circ g$ is a **subsequence** of f. Through abuse of notation, we will often identify a sequence with its range, for instance, we may say "let $\{a_n\}_{n\in\mathbb{N}} \subset X$ be a sequence".

If $\mathcal{A}$ is a family of sets, their **Cartesian product** $\prod_{A\in\mathcal{A}} A$ consists of all functions $f : \mathcal{A} \to \cup_{a\in\mathcal{A}} A$ such that $f(A) \in A$ for each $A \in \mathcal{A}$. Similar to unions and intersections, if $\mathcal{A}$ is an indexed family $\mathcal{A} = \{A_i\}_{i\in I}$ then the Cartesian product is written $\prod_{i\in I} A_i$. If $X = \prod_{i\in I} A_i$, and $i \in I$, then the coordinate map $\pi_i : X \to A_i$ is given by $\pi_i(x) = x_i$, and we call x_i the ith coordinate of x. If each A_i is a fixed set A, then we denote $\pi_{i\in I} A_i$ by A^I. If $I = \{1, 2, \ldots, n\}$, then we denote A^I by A^n and identify this with the set of ordered n-tuples of elements of A.

1.1.1 Countability

If X and Y are sets we write $|X| \leq |Y|$ (resp. $|X| = |Y|$) if there exists an injective (resp. bijective) map $f : X \to Y$. We also write $|X| < |Y|$ if $|X| \leq |Y|$, and there is no bijection from X to Y.

Theorem 1.1.1 (The Cantor-Schröder-Bernstein Theorem)**.** *If* $|X| \leq |Y|$ *and* $|Y| \leq |X|$ *then* $|X| = |Y|$.

Proof. Suppose $f : X \to Y$, and $g : Y \to X$ are both injective. Set $B = \cup_{n\in\mathbb{N}}(f \circ g)^n(Y \setminus f(X))$, and set $A = X \setminus g(B)$. Then we have $g(B) = X \setminus A$, and

$$f(A) = f(X) \setminus (f \circ g)(B) = Y \setminus ((Y \setminus f(X)) \cup (f \circ g)(B)) = Y \setminus B.$$

Hence if we define $\theta : X \to Y$ by $\theta(x) = \begin{cases} f(x) & \text{if } x \in A, \\ g^{-1}(x) & \text{if } x \in Y \setminus A = g(B), \end{cases}$ then θ gives a bijection. ∎

A set X is **countable** if $|X| \leq |\mathbb{N}|$. We say that X is **uncountable** if it is not countable.

Proposition 1.1.2.

1. *If* X *and* Y *are countable, then so is* $X \times Y$.
2. *If* I *is countable and* X_i *is countable for each* $i \in I$ *then* $\cup_{i\in I} X_i$ *is also countable.*

Proof. We let $p : \mathbb{N} \to \mathbb{N}$ be the map which takes n, to the nth prime number. Suppose $f : X \to \mathbb{N}$, and $g : Y \to \mathbb{N}$ are injective, and consider $h : X \times Y \to \mathbb{N}$ by $h(x, y) = p(f(x))^{g(y)}$. If $p(f(x_1))^{g(y_1)} = h(x_1, y_1) = h(x_2, y_2) = p(f(x_2))^{g(y_2)}$ then by uniqueness or prime factorization we have $p(f(x_1)) = p(f(x_2))$ and $g(y_1) = g(y_2)$. As p, f, and g are injective we then have $x_1 = x_2$ and $y_1 = y_2$. Thus, h is injective.

Similarly, if I is countable, and X_i is countable for each $i \in I$, then consider $f : I \to \mathbb{N}$ injective, and for each $i \in I$ consider $f_i : X_i \to \mathbb{N}$ injective. We define $g : \cup_{i\in I} X_i \to \mathbb{N}$, by setting $g(x) = p(f(i))^{f_i(x)}$ where $f(i)$ is the smallest number so that $x \in X_i$. Then similar to above it is easy to check that g is injective and hence $\cup_{i\in I} X_i$ is countable. ∎

Corollary 1.1.3. *$\mathbb{Z}$ and $\mathbb{Q}$ are countable.*

Proof. We have $\mathbb{Z} = \mathbb{N} \cup \{0\} \cup -\mathbb{N}$ showing that $\mathbb{Z}$ is countable. Also, writing any rational number in reduced fraction form a/b with $a \in \mathbb{Z}$ and $b \in \mathbb{N} \setminus \{0\}$, defines an injective function $f(a/b) = (a, b) \in Z \times \mathbb{N}$. Since $\mathbb{Z} \times \mathbb{N}$ is countable, so is $\mathbb{Q}$. ■

Proposition 1.1.4 (Cantor's diagonalization method)**.** *Let X be a set, then $|X| < |2^X|$.*

Proof. The injective map $f : X \to 2^X$ given by $f(x) = \{x\}$ shows that we have $|X| \leq |2^X|$. Now, suppose we have an injective function $g : X \to 2^X$. We let $A = \{x \in X \mid x \notin g(x)\}$. Then, if $x \in X$ and $x \in g(x)$ we have $x \notin A$ and hence $g(x) \neq A$. Similarly, if $x \in X$ and $x \notin g(x)$ then $x \in A$ and hence $g(x) \neq A$. We therefore have produced a set which is not in the range of g showing that g is not surjective. As g was arbitrary we then have $|X| \neq |2^X|$. ■

Proposition 1.1.5. *$|\mathbb{R}| = |2^{\mathbb{N}}|$, and hence $\mathbb{R}$ is uncountable.*

Proof. Note that we have $|2^{\mathbb{N}}| = |2^{\mathbb{Z}}|$. Writing each real number in its binary expansion (If there is ambiguity we choose the representation which ends in zeros) gives an injective map from $\mathbb{R}$ to $2^{\mathbb{Z}}$. On the other hand, each sequence in $2^{\mathbb{N}}$ we may view as a decimal expansion, and this gives an injective map from $2^{\mathbb{N}}$ into $\mathbb{R}$. ■

1.1.2 Transfinite induction

A **relation** on X is a subset $R \subset X \times X$. We write xRy to mean $(x, y) \in R$. A relation R is an **equivalence relation** if the following properties hold:

- xRx for each $x \in X$.
- If xRy then yRx.
- If xRy and yRz then xRz.

A relation $\prec$ is a **partial ordering** if the following properties hold:

- $x \prec x$ for each $x \in X$.
- If $x \prec y$ and $y \prec z$ then $x \prec z$.
- If $x \prec y$ and $y \prec x$ then $x = y$.

We write $x \precneqq y$ if $x \prec y$ and $x \neq y$. An **order isomorphism** between two partially ordered sets is a bijection which preserves the partial orderes. A partial ordering $\prec$ is **linear** (or **total**) if for each $x, y \in X$ we have either $x \prec y$ or $y \prec x$.

If X is partially ordered by $\prec$, a **maximal element** of X is an element $x \in X$ such that if $x \prec y$ then we have $x = y$. If $E \subset X$, then an **upper bounded** for E is an element $x \in X$ such that $y \prec x$ for each $y \in E$. We may similarly define minimal elements and lower bounds. A linear ordering is said to be **well ordered** if every nonempty subset of X has a minimal element. For example, $\mathbb{N}$ is well ordered by its usual ordering. If $(X, \leq)$ is a well ordered set and $x \in X$ we define the **initial segment of** x to be $I_x = \{y \in X \mid y < x\}$. The elements of I_x are called **predecessors** of x. Note that either $I_x \cup \{x\} = X$, or else $I_x \cup \{x\} = I_y$ where y is the minimal element in $X \setminus (I_x \cup \{x\})$.

Proposition 1.1.6 (The principle of transfinite induction)**.** *Let X be a well ordered set. If $A \subset X$ is such that $x \in A$ whenever $I_x \subset A$, then $A = X$.*

Proof. By contraposition, if $A \neq X$ we let $x \in X \setminus A$ be the minimal element. Then we have $x \notin A$, and $I_x \subset A$ from the definition of x. ∎

Lemma 1.1.7. *Let X be a well ordered set and $A \subset X$, then $\cup_{x \in A} I_x$ is either an initial segment or $A = X$.*

Proof. If A^c is nonepmty then let x be the minimal element in A^c. It's then easy to see that $A = I_x$. ∎

Proposition 1.1.8 (The principle of transfinite recursion)**.** *Let X be a well ordered set, Y a set, and let $\mathcal{F} = \cup_{x \in X} Y^{I_x}$ denote the space of all functions from initial segments of X to Y. If $G : \mathcal{F} \to Y$, then there exists a unique function $g : X \to Y$ so that $g(x) = G(g_{|I_x})$ for each $x \in X$.*

Proof. We let $\mathcal{E}$ denote the family of functions $f : I \to Y$ such that $I = X$ or is an initial segment in X, and f satisfies the formula $f(x) = G(f_{|I_x})$ for all $x \in I$. If $f' : I' \to Y$ is another such function in $\mathcal{F}$ and $I \subset I'$, then set $A = \{x \in I \mid f'(x) = f(x)\}$. If $x \in I$ and $I_x \subset A$ then we have

$$f'(x) = G(f'_{|I_x}) = G(f_{|I_x}) = f(x),$$

hence $x \in A$. It then follows by transfinite induction that $A = I$ and hence $f'_{|I} = f$.

We may then consider J the union of all I such that there is $f : I \to Y$ with $f \in \mathcal{E}$. By Lemma 1.1.7 either $J = X$, or J is an initial segment. We define a function $g : J \to Y$ by letting $g(x) = f(x)$, where $f : I \to Y$ is in $\mathcal{E}$ and $x \in I$. By our remarks above it follows easily that g is well defined and $g \in \mathcal{E}$.

If $J \neq X$ then $J = I_x$ for some $x \in X$. We could then extend the function g to $\tilde{g} : J \cup \{x\} \to Y$ such that $\tilde{g}_{|J} = g$ and $\tilde{g}(x) = G(g)$. We would then have $\tilde{g} \in \mathcal{E}$ which would contradict the maximality of g. Thus, we conclude that $J = X$ and $g : X \to Y$ is our desired function. Uniqueness follows easily from the remarks above. ∎

Intuitively, the previous result states that if we have an initial value $(G(\emptyset))$, and a procedure for choosing a new value based on the ones previously chosen (this is the function G). Then we can defines resursively a unique function on all of X.

Lemma 1.1.9. *Let $(X, \leq)$ and $(Y, \prec)$ be well ordered sets, and suppose $f : X \to Y$ is an order isomorphism, then $f(I_x) = I_{f(x)}$ for each $x \in X$. Conversely, if $f : X \to Y$ is such that $f(I_x) = I_{f(x)}$ for each $x \in X$, then $f(X)$ is either Y or an initial segment in Y, and f is an order isomorphism onto its image.*

Proof. Suppose first that $f : X \to Y$ is an order isomorphism. If $x \in X$ and $a < x$, then $f(a) < f(x)$ and hence $f(I_x) \subset I_{f(x)}$. Considering the inverse of f gives the reverse inclusion $I_{f(x)} = f(f^{-1}(I_{f(x)})) \subset f(I_x)$.

Now suppose $f : X \to Y$ such that $f(I_x) = I_{f(x)}$ for each $x \in X$. Then if $y \in f(X)$ we have $I_y \subset f(X)$ and hence $f(X)$ is either Y, or an inital segment in Y. If $x_1 < x_2$ then we have $f(x_1) \in f(I_{x_2}) = I_{f(x_2)}$ and so $f(x_1) < f(x_2)$, and this also shows that f is injective. If $f(x_1) \leq f(x_2)$ then $f(x_2) \notin I_{f(x_1)} = f(I_{x_1})$ hence $x_1 \leq x_2$. Thus, f is an order isomorphism onto its image. ■

Proposition 1.1.10. *Suppose $(X, \leq)$ is a well ordered set and $I \subset X$ is either an initial segment, or is equal to X. If $f : X \to I$ is an order isomorphism, then $I = X$ and f is the identity map.*

Proof. Let $A = \{x \in X \mid f(x) = x\}$. If $x \in X$ is such that $I_x \subset A$, then Lemma 1.1.9 shows that $I_x = f(I_x) = I_{f(x)}$, hence $x = f(x)$ showing that $x \in A$. The result then follows by transfininte induction. ■

Theorem 1.1.11. *If X and Y are well ordered, then exactly one of the following holds:*

1. *X is order isomorphic to Y;*
2. *X is order isomorphic to an initial segment in Y;*
3. *Y is order isomorphic to an initial segment in X.*

Moreover, in each of the cases the order isomorphism is unique.

Proof. We may assume Y is non-empty. Suppose that Y is not order isomorphic to an initial segment in X. If $x \in X$ and $g : I_x \to Y$ with $g(I_x) \neq Y$ then let $G(g)$ denote the smalles element in $Y \setminus G(g)$, otherwise let $G(g)$ be the initial element in Y. By the principal of transfinite recursion there is a function $f : X \to Y$ so that $g(x) = G(g_{I_x})$ for each $x \in X$.

We set $A = \{x \in X \mid g(I_x) = I_{g(x)}\}$. If $I_a \subset A$ then from Lemma 1.1.9 we have that $g(I_a)$ is either Y, or an inital segment in Y, and we have that g defines an order isomorphism from I_a onto $g(I_a)$. As Y is not order isomorphic to an initial segment in X it follows that $g(I_a)$ is an initial segment in Y, say $g(I_a) = I_y$. Then from the definition of G we have $g(a) = y$ and hence $g(I_a) = I_{g(a)}$, showing that $a \in A$. By the principle of transfinite induction we then have $A = X$ and from Lemma 1.1.9 it follows that X is either order isomorphic to Y or to an initial segment in Y.

Thus, one of the three cases above must hold, and Proposition 1.1.10 shows that no more than one can hold, and that the order isomorphism is unique. ■

1.1.3 The axiom of choice

Consider the following four principles:

AC (The Axiom of Choice): If $\{A_i\}_{i\in I}$ is a nonempty collection of nonempty sets then $\prod_{i\in I} A_i$ is nonempty.

WO (The Well Ordering Principle) Every set X can be well ordered.

ZL (Zorn's Lemma) If X is a nonempty partially ordered set and every linearly ordered subset of X has an upper bound, then X has a maximal element.

HM (The Hausdorff Maximal Principle) Every partially ordered set has a maximal linearly ordered subset.

These principles are logically equivalent, and after this section we will use them in these notes without explicit reference. In this section we show that we have the implications AC $\implies$ WO $\implies$ ZL $\implies$ HM. The reverse implications are easier and we leave them as exercises.

Proposition 1.1.12. *The axiom of choice implies the well ordering principle.*

Proof. Let X be a nonempty set. By the axiom of choice there exists $f \in \prod_{Y\in 2^X\setminus X}(X\setminus Y)$. That is, $f : 2^X \setminus X \to X$ is a function such that $f(Y) \notin Y$ for each $Y \subsetneq X$.

We define an f-string to be a well ordered set $(A, \leq)$ such that $A \subset X$ and $a = f(I_a)$ for all $a \in A$. We let $\mathcal{F}$ denote the set of f-strings. If $(A, \leq)$ and $(B, \leq')$ are f-strings, and $h : A \to B$ is an order isomorphism, then let $E = \{a \in A \mid h(a) = a\}$. If $x \in A$ is such that $I_x \subset E$, then by Lemma 1.1.9 we have

$$x = f(I_x) = f(h(I_x)) = f(I_{h(x)}) = h(x).$$

It then follows from transfinite induction that $B = A$, and h is the identity map.

It then follows from Theorem 1.1.11 that given any two distinct f-strings we must have that one is the initial segment of the other, so that $\mathcal{F}$ is linearly ordered by inclusion. We let A denote the union over all sets in $\mathcal{F}$ and we let $\prec$ denote the induced relation on A. Since $\mathcal{F}$ is linearly ordered by inclusion it follows easily that each $(A, \leq)$ is well ordered and and each initial segment of A is an f-string. From this it then follows that $(A, \leq)$ itself is an f-string, and is then the unique maximal f-string.

If $A \neq X$ then we could create the larger f-string $(A \cup \{f(A)\}, \leq')$ where $\leq'$ agrees with $\leq$ when restricted to A, and $a \leq' f(A)$ for all $a \in A$. This would then contradict the maximality of $(A, \leq)$ and so we conclude that $A = X$, and hence X is well ordered by $\leq$. ■

Proposition 1.1.13. *The well ordering principle implies Zorn's lemma.*

Proof. Let $(X, \prec)$ be a nonempty partially ordered set such that every linearly ordered subset has an upper bound. We let $\mathcal{C}$ denote the set of all linearly ordered subsets of X. By the well ordering principle there exist well orders $\leq_1$ and $\leq_2$ on X and 2^X respectively. To simplify notation we set $Y = 2^X$.

We define $f : \mathcal{C} \to X$ as follows: If $C \in \mathcal{C}$ and C does not contain a maximal element in X, then we let $f(C)$ be the $\leq_1$-least element which is not in C, but is a $\prec$-upper bound for C, otherwise if C contains a maximal element $x_0 \in X$ then we set $f(C) = x_0$.

We now recursively define a function $g : Y \to X$ such that $g(y) \prec g(z)$ whenever $y \leq_2 z$. Indeed, suppose $y \in Y$ and that g has been defined on I_y, then $g(I_y)$ is a well ordered (and hence linearly ordered) subset of X and we may set $g(y) = f(g(I_y))$.

By Cantor's diagonalization argument g cannot be injective. Thus, there exists some $y \in Y$ such that $g(y) = g(z)$ for some $z \in I_y$. It then follows that $f(g(I_y)) = g(y) \in g(I_y)$. By construction of f we must have that $f(g(I_y))$ is a maximal element. ■

Proposition 1.1.14. *Zorn's lemma implies the Hausdorff maximal principle.*

Proof. Let X be a nonempty partially ordered set and let $\mathcal{C}$ denote the space of all linearly ordered subsets. Then $\mathcal{C}$ is ordered by inclusion. If $\mathcal{C}_0 \subset \mathcal{C}$ is a subset which is linearly ordered by inclusion then we may consider $C = \cup_{C_0 \in \mathcal{C}_0} C_0$. Then $C \subset X$ is also linearly ordered and is an upper bound for $\mathcal{C}_0$ with resped to the inclusion order. We may then apply Zorn's Lemma to $\mathcal{C}$ to produce a maximal linearly ordered subset. ■

1.1.4 Ordinals and Cardinals

Proposition 1.1.15. *Suppose $X \neq \emptyset$, then $|X| \leq |Y|$ if and only if there exists a surjection from Y to X.*

Proof. Suppose that $f : X \to Y$ is injective, and take $x \in X$. We may then define a surjective function $g : Y \to X$, by letting $g(y) = \begin{cases} f^{-1}(x) & \text{if } x \in F(X) \\ x & \text{otherwise.} \end{cases}$
Conversely, if $g : Y \to X$ is surjective, then for each $x \in X$ we may choose $f(x) \in Y$ so that $g(f(x)) = x$. If $f(x) = f(y)$ then $x = g(f(x)) = g(f(y)) = y$. Thus, $f : X \to Y$ is injective. ■

Proposition 1.1.16. *For any sets X, Y, either $|X| \leq |Y|$, or $|Y| \leq |X|$.*

Proof. If we well order X and Y, then this follows immediately from Theorem 1.1.11. ■

We would like to define an ordinal as an order isomorphic equivalence class of well ordered sets, the previous proposition would then show that we have a linear ordering on the collection of ordinals. We should be somewhat careful here though as there is no reason that the collection of all well ordered sets should be a set itself, and we have only discussed equivalence relations and orderings on sets. One option to make this precise is to work outside of the universe of sets, that is, we can call the collection of all well ordered sets a "class", and then we can introduce equivalence relations and well orderings for classes. Another option to make this notion precise is the following definition proposed by von Neumann:

A set α is an **ordinal** if every element of α is also a proper subset of α, and if α is well ordered with respect to set inclusion.

The first few ordinals are then $0 = \emptyset$, $1 = \{0\}$, $2 = \{0, 1\}$, $3 = \{0, 1, 2\}$. The first infinite ordinal is $\omega = \{0, 1, 2, 3, 4, \ldots\}$. For each ordinal α we may consider the set $\alpha \cup \{\alpha\}$, which is again an ordinal, this is the successor ordinal which we denote by $\alpha + 1$. Any ordinal which is not a successor ordinal is called a limit ordinal. If two ordinals are order isomorphic, then it follows easily by induction that they must be the same. More generally, if α and β are ordinals, then by Theorem 1.1.11 either α and β are order isomorphic, in which case $\alpha = \beta$, or else one (say α) is isomorphic to (and hence equal to) an initial segment of the other, in which case we have $\alpha \subset \beta$.

The following proposition shows that von Neumann's definition captures all order isomorphism classes.

Proposition 1.1.17. *Let X be a well ordered set, then there exists a unique ordinal α such that X and α are order isomorphic.*

Proof. We let A denote the subset of X consisting of all points x such that there exists an ordinal α_x, and an order isomorphism $f_x : I_x \to \alpha_x$. Note that if $x, y \in A$ with $x < y$, then we obtain an order isomorphism from α_x to an initial segment in α_y, hence we have $\alpha_x \subset \alpha_y$. We then have $\alpha_y = \cup_{x \leq y} \alpha_x$, and $f_{y|I_x} = f_x$.

If $I_z \subset A$, and z is not a successor, then we set $\alpha = \cup_{x \in I_z} \alpha_x$. Note that α is again an ordinal. We define the function $f : I_z \to \alpha$ by setting $f(x) = f_y(x)$ for some $x < y < z$. (note that such a y exists since z is not a successor). Then f is well defined and implements an order isomorphism between I_z and α, showing that $z \in A$.

Similarly, if $I_z \subset A$, and z is a successor to $y \in A$, then we consider the ordinal $\alpha = \alpha_y \cup \{\alpha_y\}$, and we define the function $f : I_z \to \alpha$ by $f(x) = f_y(x)$ for $x < y$, and $f(y) = \alpha_y$. We then again see that $z \in A$.

By transfinite induction we then have $A = X$, and the result then follows easily. ■

Just as there is a first infinite ordinal, there is also a first uncountable ordinal:

Proposition 1.1.18. *There is a unique uncountable ordinal ω_1 such that I_x is countable for each $x \in \omega_1$.*

Proof. We let ω_1 be the set of all countable ordinals α. Then ω_1 is ordered by inclusion. If $\alpha \in \omega_1$ then α is a countable ordinal and hence so are all the ordinals contained in α, thus $\alpha \subset \omega_1$. We therefore have that ω_1 is an ordinal, and I_x is countable for each $x \in \omega_1$. Note that ω_1 cannot be countable since $\omega_1 \notin \omega_1$.

If Ω' were another such ordinal, then we could not have $\Omega' \subset \omega_1$ since ω_1 is not countable. We similarly could not have $\omega_1 \subset \Omega'$. Hence we must have $\omega_1 = \Omega'$. ■

If we fix a set X, then by the well ordering principle there exists a well ordering on X and hence a bijection between X and some ordinal α. The **cardinality** of X is the smallest ordinal such that there exists such a bijection. We denote the cardinality of X by $|X|$, and note that this is consistent with our notation above.

1.1.5 Exercises

Exercise 1.1.19. If X is countably infinite then there is a bijection from X onto $\mathbb{N}$.

Exercise 1.1.20. We have $|\mathbb{R}^{\mathbb{N}}| = |\mathbb{R}|$.

Exercise 1.1.21. Let $2^{<\mathbb{N}}$ denote the set of finite sequences in $\{0,1\}$, then $2^{<\mathbb{N}}$ is countable.

Exercise 1.1.22. Let X be a countably infinite set. There exists an uncountable family $\mathcal{F} \subset 2^X$ so that for any distinct pair $A, B \in \mathcal{F}$ we have that $A \cap B$ is finite. (Hint: It may help to consider the case $X = 2^{<\mathbb{N}}$.)

A complex number α is **algebraic** if it is the solution of a polynomial having rational coefficients.

Exercise 1.1.23. The set of algebraic numbers is countable.

Exercise 1.1.24. The Hausdorff maximal principle implies Zorn's lemma.

Exercise 1.1.25. Zorn's lemma implies the well ordering principle.

Exercise 1.1.26. The well ordering principle implies the axiom of choice.

Let X be a set, and $\leq$ be a linear ordering on X. We say that the linear order is **dense** if for all $x < y$ there exists $z \in X$ such that $x < z < y$.

Exercise 1.1.27 (Cantor's back-and-forth method)**.** Let $(X, \leq)$ and $(Y, \leq)$ be countable dense linear orderings which do not have upper or lower bounds. Enumerate $X = \{x_1, x_2, \ldots\}$, and $Y = \{y_1, y_2, \ldots\}$.

1. There exist increasing sequences of finite sets $A_n \subset X$, $B_n \subset Y$, and order preserving bijections $f_n : A_n \to B_n$ such that $x_n \in A_n$, $y_n \in B_n$, and $f_{n+1|A_n} = f_n$, for all $n \geq 1$.

2. There exists an order preserving bijection $f : X \to Y$.

A (undirected) **graph** consists of a pair (V, E) where V is a set (the **vertex set**) and $E \subset V \times V$ (the **edge set**) such that $(v, w) \in E$ if and only if $(w, v) \in E$. A **subgraph** is a graph (V_0, E_0) with $V_0 \subset V$, and $E_0 \subset E$.

If (V, E) is a graph, two vertices $v, w \in V$ are **adjacent** if $(v, w) \in E$. We let $N(v)$ denote the set of vertices which are adjacent to v. A graph (V, E) is **locally finite** if $|N(v)| < \infty$ for each $v \in V$. A **finite simple path** is an injective function $p : \{1, 2, \ldots, n\} \to V$, such that $(p(k), p(k + 1)) \in E$ for all $1 \leq k < n$; we say that n is the **length** of the path. A ray is an injective function $p : \mathbb{N} \to V$ such that $(p(k), p(k + 1)) \in E$ for all $1 \leq k$. A graph is **connected** if for any distinct vertices $v, w \in V$, there exists a finite simple path $p : \{1, 2, \ldots, n\} \to V$ such that $p(1) = v$ and $p(n) = w$.

Exercise 1.1.28 (König's lemma)**.** If a locally finite connected graph (V, E) has infinitely many vertices, then (V, E) admits a ray.

1.2 Metric spaces

Let X be a set. A **semimetric** on X is a function $d : X \times X \to [0, \infty)$ such that for all $x, y, z \in X$ the following properties hold:

1. $d(x, y) = d(y, x)$.
2. $d(x, z) \leq d(x, y) + d(y, z)$.

If, in addition, we have that $x = y$ if and only if $d(x, y) = 0$, then d is a **metric**. A **metric space** is a pair (X, d) consisting of a set X and a metric d on X. When d is understood we will sometimes refer to the metric space X.

Examples of metric spaces include:

1. Euclidean space $\mathbb{R}^n$ with metric $d(x, y) = \|x - y\|$.
2. If X is any set and for all $x, y \in X$ we have $d(x, y) = \begin{cases} 0 & \text{if } x = y; \\ 1 & \text{if } x \neq y, \end{cases}$ then (X, d) is a metric space.
3. If (X, d) is a metric space and $A \subset X$, then $(A, d_{|A \times A})$ is a metric space.
4. If (X_1, d_1) and (X_2, d_2) are metric spaces, then $X_1 \times X_2$ is a metric space with metric $d((x_1, x_2), (y_2, y_2)) = \max(d_1(x_1, y_1), d_2(x_2, y_2))$.
5. If (X, d) is a metric space and $f : [0, \infty) \to [0, \infty)$ is a strictly increasing function satisfying $f(0) = 0$, and $f(s + t) \leq f(s) + f(t)$ for all $s, t \in \mathbb{R}$ (e.g., $f(t) = \sqrt{t}$, or $f(t) = \frac{t}{t+1}$), then $(X, f \circ d)$ is again a metric space.

If (X, d) is a metric space, $x \in X$ and $r > 0$, then the **ball** of radius r about x is $B(r, x) = \{y \in X \mid d(x, y) < r\}$; we call any such set $B(r, x)$ an r-ball. A set $A \subset X$ is **open** if for each $x \in A$ there exists $r > 0$ such that $B(r, x) \subset A$. A set $A \subset X$ is **closed** if A^c is open. Both $\emptyset$ and X are open. If a set is both closed and open then we say it is **clopen**. Note that the collection of open sets is closed under finite intersections and arbitrary unions. Taking complements shows that the collection of closed sets is closed under finite unions and arbitrary intersections. If $E \subset X$ then the **closure** $\overline{E}$ of E is the intersection of all closed sets which contian E. We say the E is **dense** if $\overline{E} = X$.

If (X_1, d_1) and (X_2, d_2) are metric spaces, and $f : X_1 \to X_2$, then f is

1. **isometric** if $d_2(f(x), f(y)) = d_1(x, y)$, for all $x, y \in X_1$;
2. **Lipschitz continuous** if there exists $K \geq 0$ (a **Lipschitz constant**), such that for all $x, y \in X_1$ we have $d_2(f(x), f(y)) \leq K d_1(x, y)$;
3. **contractive** if f is Lipschitz continuous with Lipschitz constant 1;
4. **uniformly continuous** if for all $\varepsilon > 0$, there exists $\delta > 0$ so that $f(B(\delta, x)) \subset B(\varepsilon, f(x))$ for all $x \in X_1$;
5. **continuous at** $x \in X_1$ if for each $\varepsilon > 0$, there exists $\delta > 0$ so that $f(B(\delta, x)) \subset B(\varepsilon, f(x))$.
6. **continuous** if f is continuous at each point $x \in X_1$.
7. a **homeomorphism** if f is bijective and both f and f^{-1} are continuous.

If X is a metric space, a sequence $\{x_n\}_{n \in \mathbb{N}} \subset X$ has a limit point $x \in X$ if for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ so that $x_n \in B(\varepsilon, x)$ for $n \geq N$. We say that $\{x_n\}_n$ converges to x and write $\lim_{n\to\infty} x_n = x$ if this is the case. $\{x_n\}_{n \in \mathbb{N}}$ is convergent if it converges to some point $x \in X$.

Proposition 1.2.1. *If (X_1, d_1) and (X_2, d_2) are metric spaces, the following are equivalent:*

1. *$f : X_1 \to X_2$ is continuous;*
2. *for any convergent sequence $\{x_n\}_{n \in \mathbb{N}}$ we have that $\{f(x_n)\}_{n \in \mathbb{N}}$ is also convergent and $\lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n)$;*
3. *$f^{-1}(O)$ is open for each open set $O \subset X_2$.*

Proof. First, suppose $f : X_1 \to X_2$ is continuous, and $\{x_n\}_{n \in \mathbb{N}}$ is a sequence such that $x = \lim_{n\to\infty} x_n$. Fix $\varepsilon > 0$. Since f is continuous there exists $\delta > 0$ so that $f(B(\delta, x)) \subset B(\varepsilon, f(x))$. Since $x = \lim_{n\to\infty} x_n$, there exists $N \in \mathbb{N}$ so that $x_n \in B(\delta, x)$ for all $n \geq N$. Hence, $f(x_n) \in f(B(\delta, x)) \subset B(\varepsilon, f(x))$ for all $n \geq N$. Since $\varepsilon > 0$ was arbitrary we have $f(x) = \lim_{n\to\infty} f(x_n)$.

Next, suppose that $O \subset X_2$ is open, but $f^{-1}(O)$ is not open. Then there exists $x \in f^{-1}(O)$ such that for all $n \in \mathbb{N}$ we have $B(1/n, x) \not\subset f^{-1}(O)$. For each $n \in \mathbb{N}$ choose $x_n \in B(1/n, x)$. Since O is open there exists $\varepsilon > 0$ so that $B(\varepsilon, f(x)) \subset O$. We then have $x = \lim_{n\to\infty} x_n$, and $f(x_n) \notin O \supset B(\varepsilon, f(x))$ for all $n \in \mathbb{N}$, hence $\{f(x_n)\}_{n \in \mathbb{N}}$ does not converge to $f(x)$.

Finally, suppose that $f^{-1}(O)$ is open whenever $O \subset X_2$ is open. Fix $x \in X$ and $\varepsilon > 0$. Then $B(\varepsilon, f(x))$ is open and hence so is $f^{-1}(B(\varepsilon, f(x)))$. Therefore, there exists $\delta > 0$ so that $B(\delta, x) \subset f^{-1}(B(\varepsilon, f(x)))$. Hence, we have $f(B(\delta, x)) \subset B(\varepsilon, f(x))$ showing that f is continuous. ∎

A sequence of functions $f_n : X \to Y$ is said to **converge pointwise** to a function $f : X \to Y$, if $f(x) = \lim_{n\to\infty} f_n(x)$ for each $x \in X$. The sequence $\{f_n\}_{n\in\mathbb{N}}$ **converges uniformly** to f if $\lim_{n\to\infty} \sup_{x\in X} |f(x) - f_n(x)| = 0$.

Proposition 1.2.2. *Suppose $f_n : X \to Y$ are continuous, and $\{f_n\}_{n\in\mathbb{N}}$ converges to $f : X \to Y$ uniformly, then f is continuous.*

Proof. Fix $\varepsilon > 0$, and $x \in X$. Since $f_n \to f$ uniformly, there exists $n \in \mathbb{N}$ so that $\sup_{y\in X} |f(y) - f_n(y)| < \varepsilon/3$. Since f_n is continuous there exists an open set O containing x so that for $y \in O$ we have $|f_n(y) - f_n(x)| < \varepsilon/3$. For $y \in O$ we then have

$$|f(y) - f(x)| \leq |f(y) - f_n(y)| + |f_n(y) - f_n(x)| + |f_n(x) - f(x)| < \varepsilon.$$

Thus, f is continuous. ∎

A sequence $\{x_n\}_{n\in\mathbb{N}}$ is **Cauchy** if for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ so that $d(x_n, x_m) < \varepsilon$ for all $n, m \geq N$. A metric space is **complete** if every Cauchy sequence is convergent.

Proposition 1.2.3. *$\mathbb{R}^n$ with its Euclidean metric is complete.*

Proof. A sequence in $\mathbb{R}^n$ is Cauchy if and only if its coordinates are Cauchy, and, similarly, a sequence converges if and only if its coordinates converge. Thus, the general result follows from $\mathbb{R}$. Suppose $\{x_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ is Cauchy. Then there exists $N \in \mathbb{N}$ so that $|x_N - x_m| < 1$ for all $m \geq N$. Hence $\{x_n\}_{n\in\mathbb{N}}$ is bounded. We let $x = \limsup_{n\to\infty} x_n$.

Fix $\varepsilon > 0$, then there exists $N \in \mathbb{N}$ so that $|x_n - x_m| < \varepsilon/2$ for all $n, m \geq N$. Also, there exists $n \geq N$ so that $|x - x_n| < \varepsilon/2$. Then for all $m \geq N$ we have $|x - x_m| \leq |x - x_n| + |x_n - x_m| < \varepsilon$. It then follows that $x = \lim_{n\to\infty} x_n$. ∎

Proposition 1.2.4. *A closed subset of a complete metric space is complete, and a complete subspace of an arbitrary metric space is closed.*

Proof. Suppose (X, d) is complete and $F \subset X$ is closed. Let $\{x_n\}_{n\in\mathbb{N}} \subset F$ be Cauchy. Then it is also Cauchy in X and by completeness there exists $x \in X$ so that $x_n \to x$. If $\varepsilon > 0$ then there exists $N \in \mathbb{N}$ so that $x_n \in B(\varepsilon, x)$ for all $n \geq N$. In particular we have $B(\varepsilon, x) \cap F \neq \emptyset$ and hence $x \notin F^c$ as this is open and $\varepsilon > 0$ was arbitrary.

Conversely, Suppose $F \subset X$ is a subspace which is not closed. Then F^c is not open and hence there exists $x \in F^c$, so that $B(1/n, x) \cap F \neq \emptyset$ for all $n \in \mathbb{N}$. Take $x_n \in B(1/n, x) \cap F$ for each $n \in \mathbb{N}$. Then $x = \lim_{n\to\infty} x_n$ and hence $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy. However, $x \notin F$ and hence $\{x_n\}_{n\in\mathbb{N}}$ does not converge to a point in F. Thus, F is not complete. ∎

If (X, d) is a metric space, then we let Cauchy(X) denote the set of all Cauchy sequences in X. On Cauchy$(X)^2$ we define the function $\overline{d}$ by $\overline{d}(\{x_n\}_{n\in\mathbb{N}}, \{y_n\}_{n\in\mathbb{N}}) = \lim_{n\to\infty} d(x_n, y_n)$. We leave it as an exercise to verify that this limit actually exists. Using the properties of the metric d it is then easy to see that we have $\overline{d}(s,t) = \overline{d}(t,s)$ and $\overline{d}(s,r) \leq \overline{d}(s,t) + \overline{d}(t,r)$, for all Cauchy sequences s, t, r. We define an equivalence relation on Cauchy(X) by $s \sim t$ if and only if $\overline{d}(s,t) = 0$. It's easy to see that this is indeed an equivalence relation, and that $\overline{d}(s,t)$ only depends on the equivalence classes that s and t lie in. Thus, setting $\overline{X} = \text{Cauchy}(X)/\sim$, we may view $\overline{d}$ as a function on $\overline{X} \times \overline{X}$ where it then gives a metric. We also have a natural isometric embedding $\pi : X \to \overline{X}$ which takes a point $x \in X$ to the constant sequence $\pi(x) = \{x\}_{n\in\mathbb{N}}$. We call the metric space $(\overline{X}, \overline{d})$ the **completion of** (X, d), and we usually view X as a subspace by identifying X with $\pi(X)$.

Proposition 1.2.5. *$(\overline{X}, \overline{d})$ is complete, and X is a dense subspace.*

Proof. We first show that $\pi(X)$ is a dense subspace. Suppose that $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy and $\varepsilon > 0$ is given. Then there exists $N \in \mathbb{N}$ so that $|x_n - x_m| < \varepsilon$ for all $n, m \geq N$. In particular, we have that $|x_N - x_m| < \varepsilon$ for all $m \geq N$. Hence, we have $\overline{d}(\pi(x_N), \{x_n\}_{n\in\mathbb{N}}) \leq \varepsilon$.

Next we show that $(\overline{X}, \overline{d})$ is complete. Note that if $\{s_n\}_{n\in\mathbb{N}}, \{t_n\}_{n\in\mathbb{N}} \subset \overline{X}$ such that $0 = \lim_{n\to\infty} d(s_n, t_n)$, then $\{s_n\}_{n\in\mathbb{N}}$ is Cauchy (resp. convergent) if and only if $\{t_n\}_{n\in\mathbb{N}}$ is Cauchy (resp. convergent). Thus, it is enough to consider Cauchy sequences which are valued in the dense subspace $\pi(X)$. Suppose therefore that $\{\pi(x_n)\}_{n\in\mathbb{N}}$ is Cauchy, with $x_n \in X$. If we set $s = \{x_n\}_{n\in\mathbb{N}}$ then it follows easily that we have $0 = \lim_{n\to\infty} \overline{d}(s, \pi(x_n))$. Hence $(\overline{X}, \overline{d})$ is complete. ■

If $E \subset X$, then E is **bounded** if there exists $K > 0$, such that $d(x,y) \leq K$, for all $x, y \in E$. If $\{V_i\}_{i\in I}$ is a family of subsets of X such that $E \subset \cup_{i\in I} V_i$, then $\{V_i\}_{i\in I}$ is a **cover** of E. E is **totally bounded** if for any $\varepsilon > 0$, there is a finite collection of ε-balls which cover E. Note that totally bounded sets are also bounded.

Lemma 1.2.6. *A metric space (X, d) is totally bounded if and only if every sequence has a Cauchy subsequence.*

Proof. Suppose (X, d) is totally bounded and $\{x_n\}_{n\in\mathbb{N}}$ is a sequence. Fix $\varepsilon > 0$. We will inductively define a decreasing sequence of infinite subsets $A_j \subset \mathbb{N}$, such that $d(x_n, x_m) < 2/j$ for all $n, m \in A_j$, and $j \geq 2$. We first set $A_1 = \mathbb{N}$. Suppose now that A_{j-1} has been chosen for $j \geq 2$. Since E is totally bounded, there exist a finite collection of $1/j$-balls $O_1, \ldots, O_k$ which cover E. Since A_{j-1} is infinite for some O_i we must have that $A_j = \{k \in A_{j-1} \mid x_k \in O_i\}$ is infinite.

We new choose a subsequence by taking $n_j \in A_j$ so that n_j is strictly increasing. If $\varepsilon > 0$, and $j \in \mathbb{N}$ so that $2/j < \varepsilon$ then we have $d(x_{n_k}, x_{n_l}) < 2/j < \varepsilon$ for all $k, l \geq j$. Therefore we have that the subsequence $\{x_{n_j}\}_{j\in\mathbb{N}}$ is Cauchy.

Conversely, if E is not totally bounded then there exists $\varepsilon_0 > 0$ so that there is no cover of E by finitely many ε_0-balls. We may therefore inductively construct a sequence $x_n \in E$ so that $d(x_n, x_m) \geq \varepsilon_0$ for all $n, m \in \mathbb{N}$. We then have that no subsequence of $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy. ∎

Lemma 1.2.7. *Let (X, d) be a totally bounded metric space, then every open cover has a countable subcover.*

Proof. Suppose that $\{V_i\}_{i\in I}$ is an open cover. For each $n \in \mathbb{N}$ take $\{x_1^n, \ldots, x_{k_n}^n\} \subset X$, so that $\cup_{j=1}^{k_n} B(1/n, x_j) = X$. We then let $\mathcal{O}_{n,m}$ be the collection of all open balls $B(1/m, x_j)$ which are contained in some V_i, for $i \in I$. We set $\mathcal{O} = \cup_{n,m\in\mathbb{N}} \mathcal{O}_{n,m}$.

If $x \in E$ then we have $x \in V_i$ for some $i \in I$. We then have $B(1/n, x) \subset V_i$ for some $n \in \mathbb{N}$. For some $1 \leq j \leq k_{2n}$ we then have $x \in B(1/2n, x_j^{2n}) \subset V_i$. Thus, we see that $\mathcal{O}$ covers X and is countable. Moreover, each set in $\mathcal{O}$ is contained in V_i for some $i \in I$, thus a countable subcollection of $\{V_i\}_{i\in I}$ must cover X. ∎

Theorem 1.2.8. *If $E \subset X$, the following are equivalent:*

1. *E is complete and totally bounded.*
2. *(The Bolzano-Weierstrass Property) Every sequence in E has a subsequence which converges to a point in E.*
3. *(The Heine-Borel Property) If $\{V_i\}_{i\in I}$ is a cover of E by open sets, then there exists a finite set $F \subset I$ such that $\{V_i\}_{i\in F}$ is also a cover of E.*

Proof. ($1 \implies 2$) Suppose that $E \subset X$ is complete and totally bounded. Let $\{x_n\}_{n\in\mathbb{N}} \subset E$ be a sequence. By Lemma 1.2.6 there exists a Cauchy subsequence $\{x_{n_j}\}_{j\in\mathbb{N}}$, and since E is complete we must have that this subsequence converges. Therefore E satisfies the Bolzano-Weierstrass property.

($2 \implies 1$) If E is not totally bounded then Lemma 1.2.6 shows that X has a sequence which has no Cauchy (and hence no convergent) subsequence. Similarly, if E is not complete then there exists a Cauchy sequence $\{x_n\}_{n\in\mathbb{N}}$ which does not converge, and it then follows easily that no subsequence can converge either. We have therefore shown the equivalence between the Bolzano-Weierstrass property and being complete and totally bounded.

($3 \implies 1$) If E is not totally bounded then there exists $\varepsilon_0 > 0$ so that there is no cover of E by finitely many ε_0-balls. However, all ε_0-balls cover E and hence E does not have the Heine-Borel Property. Also, if E is not complete, then we may consider the completion $\overline{E}$ and take a point $x \in \overline{E} \setminus E$. Then consider $O_n = \{y \in E \mid \overline{d}(y, x) > 1/n\}$. We then have that $\{O_n\}_{n\in\mathbb{N}}$ is an increasing sequence of open sets, such that $\cup_{n\in\mathbb{N}} O_n = E$. However, $O_n \neq E$ for any $n \in \mathbb{N}$ since E is dense in $\overline{E}$. Hence, we again have shown that E does not have the Heine-Borel property.

($1 \implies 3$) Suppose that E is totally bounded and $\{V_i\}_{i\in I}$ is an open cover which does not have a finite subcover. By Lemma 1.2.7 we may pass to a countable cover so that we may assume $\{V_i\}_{i\in I}$ is countable and then sequence this as $\{V_n\}_{n\in\mathbb{N}}$. We inductively define a sequence $\{x_n\}_{n\in\mathbb{N}}$ by taking $x_n \notin \cup_{k=1}^n V_n$. Note that by construction, each open set V_n can contain at most finitely many elements in the sequence $\{x_n\}_{n\in\mathbb{N}}$. By Lemma 1.2.6 there exists a Cauchy subsequence $\{x_{n_j}\}_{j\in\mathbb{N}}$. If this subsequence converged to some point $x \in E$, then x would be contained in some open set V_n and it would follow that infinitely many x_{n_j}'s would belong to V_n contradicting our remark above. Thus, we must have that $\{x_{n_j}\}_{j\in\mathbb{N}}$ is a Cauchy sequence which does not converge and hence X is not complete. ■

Any set E which satisfies the conditions of the previous theorem is called a **compact** set. Note that homeomorphisms preserve open sets and hence from the Heine-Borel Property we see that homeomorphisms preserve compact sets. This is not the case however for complete sets.

Proposition 1.2.9. *Let (X, d) and (Y, ρ) be metric spaces with X compact. Suppose that $f : X \to Y$ is continuous. Then $f(X)$ is compact and f is uniformly continuous.*

Proof. If $\{O_i\}_{i\in I}$ is an open cover of $f(X)$, then since f is continuous we have that $\{f^{-1}(O_i)\}_{i\in I}$ is an open cover of X. By the Heine-Borel property there exists a finite subcover $f^{-1}(O_1), \ldots, f^{-1}(O_n)$. We then have that $O_1, \ldots, O_n$ covers $f(X)$ and so by the Heine-Borel property we have that $f(X)$ is compact.

To see that f is uniformly continuous we fix $\varepsilon > 0$. Since f is continuous, for each $x \in X$ there exists $\delta_x > 0$ so that $f(B(\delta_x, x)) \subset B(\varepsilon/2, f(x))$. Then $\{B(\delta_x/2, x)\}_{x\in X}$ covers X and by the Heine-Borel property there is a finite subcover $B(\delta_{x_1}/2, x_1), \ldots, B(\delta_{x_n}/2, x_n)$.

Set $\delta = \min_{1\leq i\leq n}\{\delta_{x_i}/2\}$. Then if $1 \leq i \leq n$ and $x \in B(\delta_{x_i}/2, x_i)$ we have $B(\delta, x) \subset B(\delta_{x_i}, x_i)$ and hence

$$f(B(\delta, x)) \subset f(B(\delta_{x_i}, x_i)) \subset B(\varepsilon/2, f(x_i)).$$

Therefore, $f(B(\delta, x)) \subset B(\varepsilon, f(x))$. Since $B(\delta_{x_1}/2, x_1), \ldots, B(\delta_{x_n}/2, x_n)$ covers X it follows that f is uniformly continuous. ■

1.2.1 Exercises

Exercise 1.2.10. Suppose that X is a set and $d : X \times X \to [0, \infty)$ is a semimetric on X. We define a relation $\sim$ on X by $x \sim y$ if $d(x, y) = 0$. Then $\sim$ is an equivalence relation on X and we have a well defined metric on $X/\sim$ given by $\tilde{d}([x], [y]) = d(x, y)$.

Exercise 1.2.11. There are two homeomorphic metric spaces (X_1, d_1) and (X_2, d_2) such that (X_1, d_1) is complete, while (X_2, d_2) is not.

A metric space (X, d) is **separable** if it contains a countable dense set.

Exercise 1.2.12. A compact metric space is separable.

We let $\ell^\infty(\mathbb{N})$ denote the set of uniformly bounded sequences from $\mathbb{N}$ to $\mathbb{C}$. We consider this as a complete metric space whose metric is given by $d(f,g) = \|f-g\|_\infty = \sup_{n\in\mathbb{N}} |f(n)-g(n)|$.

Exercise 1.2.13 (Kuratowski)**.** Every bounded seprable metric space is isometric to a subspace of $\ell^\infty(\mathbb{N})$.

1.3 Normed spaces

We assume the reader is familiar with the basic properties of vector spaces. Let $\mathbb{K} = \mathbb{R}$, or $\mathbb{K} = \mathbb{C}$, and suppose that V is a $\mathbb{K}$-vector space. A **seminorm** on V is a map $V \ni v \mapsto \|v\| \in [0,\infty)$ which satisfies

1. $\|v+w\| \leq \|v\| + \|w\|$;
2. $\|kv\| = |k|\|v\|$, for $k \in \mathbb{K}$, and $v, w \in V$.

If, in addition, we have that $\|v\| = 0$ if and only if $v = 0$, then we say that $\|\cdot\|$ is a **norm**. Associated with a (pre)norm is a (pre)metric d which is given by $d(v,w) = \|v+w\|$. A **normed space** is a pair $(V, \|\cdot\|)$ where V is a vector space and $\|\cdot\|$ is a norm on V. If the associated metric is complete then the normed space is a **Banach space**. Examples of Banach spaces include:

1. $\ell^1_n = \mathbb{K}^n$, with norm $\|(\alpha_1,\ldots,\alpha_n)\|_1 = \sum_{k=1}^n |\alpha_k|$.
2. More generally, if $1 \leq p < \infty$, $\ell^p_n = \mathbb{K}^n$, with norm $\|(\alpha_1,\ldots,\alpha_n)\|_p = \left(\sum_{k=1}^n |\alpha_k|^p\right)^{1/p}$.
3. $\ell^\infty_n = \mathbb{K}^n$, with norm $\|(\alpha_1,\ldots,\alpha_n)\|_\infty = \max\{|\alpha_k| \mid 1 \leq k \leq n\}$.

If $(V, \|\cdot\|_V)$, and $(W, \|\cdot\|_W)$, are normed spaces, and $T : V \to W$ is a linear operator, then we say that T is **bounded** if there exists $K > 0$ so that $\|Tv\|_W \leq K\|v\|_V$ for all $v \in V$. We let $\mathcal{B}(V,W)$ (or $\mathcal{B}(V)$ if $V = W$) denote the set of bounded linear operators. Then $\mathcal{B}(V,W)$ is a $\mathbb{K}$-vector space, where the vector space structure is taken pointwise, i.e., $(T+S)(v) = T(v) + S(v)$, and $(kT)(v) = k(T(v))$, for $k \in \mathbb{K}$, $v \in V$, and $T, S \in \mathcal{B}(V,W)$. If $T \in \mathcal{B}(V,W)$ then the **operator norm** of T is given by $\|T\|_{\mathcal{B}(V,W)} = \sup_{v\in V, \|v\|_V \leq 1} \|Tv\|_V$. The space $\mathcal{B}(V,W)$, together with its operator norm, is a normed space.

1.3.1 Algebras

We again let $\mathbb{K} = \mathbb{R}$, or $\mathbb{K} = \mathbb{C}$. A $\mathbb{K}$-**algebra** is a $\mathbb{K}$-vector space A, together with a binary operation $A \times A \ni (a,b) \mapsto ab \in A$ (called multiplication, or composition), such that

1. $(ab)c = a(bc)$;
2. $\alpha(ab) = (\alpha a)b = a(\alpha b)$;

3. $a(b+c) = (ab) + (ac)$;

4. $(a+b)c = (ac) + (bc)$, for $\alpha \in \mathbb{K}$, $a, b, c \in A$.

Examples of algebras include:

1. The vector space of $n \times n$ matrices $\mathbb{M}_n(\mathbb{K})$, together with matrix multiplication.

2. The space of $\mathbb{K}$-polynomials with its usual vector space structure and multiplication.

3. ℓ_n^∞ where multiplication is taken coordinate-wise $(\alpha_1, \ldots, \alpha_n) \cdot (\beta_1, \ldots, \beta_n) = (\alpha_1\beta_1, \ldots, \alpha_n\beta_n)$.

A **normed algebra** is an algebra A, which also has a norm $\| \cdot \|$ which satisfies $\|ab\| \leq \|a\|\|b\|$, for $a, b \in A$. A **Banach algebra** is a normed algebra where the norm is complete.

The space $\ell_n^\infty(\mathbb{K})$ is a normed algebra. Also, if V and W are normed spaces then $\mathcal{B}(V, W)$, with its operator norm, is a normed algebra.

If (X, d) is a metric space, we let $C_b(X)$ denote the space of all complex-valued continuous functions which are uniformly bounded (If X is compact then the boundedness is automatic and we use the notation $C(X)$ instead). For $f \in C_b(X)$, the **uniform norm** of f is given by $\|f\|_\infty = \sup_{x\in X} |f(x)|$.

Proposition 1.3.1. *Let (X, d) be a metric space, then $C_b(X)$, endowed with the uniform norm, is a Banach algebra.*

Proof. First, note that $C_b(X)$ is clearly an algebra, and $\|\cdot\|_\infty$ is clearly a norm on $C_b(X)$. If $f, g \in C_b(X)$, then $\|fg\|_\infty = \sup_{x\in X} |f(x)g(x)| \leq \sup_{x,y\in X} |f(x)g(y)| = \|f\|_\infty \|g\|_\infty$ so that $C_b(X)$ is a normed algebra.

Suppose $\{f_n\}_{n\in\mathbb{N}} \subset C_b(X)$ is Cauchy. Therefore, for each $x \in X$ the sequence $\{f_n(x)\}_{n\in\mathbb{N}}$ is Cauchy and hence converges to some $f(x) \in \mathbb{C}$. We then have that $0 = \lim_{n\to\infty} \|f_n - f\|_\infty$, and $f \in C_b(X)$ by Proposition 1.2.2. Therefore $C_b(X)$ is complete and hence is a Banach algebra. ∎

Note that if X is a set and d is the metric $d(x, y) = \begin{cases} 0 & \text{if } x = y; \\ 1 & \text{if } x \neq y, \end{cases}$ then every function is continuous and hence $C_b(X)$ is the space of all uniformly bounded functions.

1.3.2 Exercises

In the following we consider vector spaces over a field $\mathbb{K}$, where $\mathbb{K} = \mathbb{R}$, or $\mathbb{K} = \mathbb{C}$.

Recall that if V is a vector space and $V_0 \subset V$ is a subspace, then the quotient space V/V_0 is defined to be the set of cosets $\{v + V_0 \mid v \in V\}$. This is naturally a vector space whose vector space operations satisfy $\alpha(v_1 + V_0) + (v_2 + V_0) = (\alpha v_1 + v_2) + V_0$, for all $v_1, v_2 \in V$, and scalar α.

Exercise 1.3.2. Suppose that V is a $\mathbb{K}$-vector space and $\|\cdot\|_0$ is a seminorm on V. Set $V_0 = \{v \in V \mid \|v\|_0 = 0\}$. Then V_0 is a linear subspace and we have a well defined norm on V/V_0 given by $\|v + V_0\| = \|v\|_0$, for each $v \in V$.

Exercise 1.3.3. Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed spaces, and $T : V \to W$ a linear operator. Then T is bounded if and only if T is continuous.

Exercise 1.3.4. Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed spaces, then the operator norm on $\mathcal{B}(V, W)$ is indeed a norm, and that with this norm $\mathcal{B}(V)$ is a normed algebra. Moreover, $\mathcal{B}(V, W)$ is a Banach space if W is a Banach space.

Exercise 1.3.5. Let V be a finite dimensional $\mathbb{K}$-vector space, and suppose $\|\cdot\|_1$, and $\|\cdot\|_2$ are norms on V. Then the identity map from $(V, \|\cdot\|_1)$ to $(V, \|\cdot\|_2)$ is a homeomorphism.

Exercise 1.3.6. Let V be a finite dimensional normed space. Then the closed unit ball $\overline{B}(1, 0)$ is compact, and V is a Banach space.

Exercise 1.3.7 (Riesz' lemma)**.** Let $(V, \|\cdot\|)$ be a normed space, $W \subset V$ a proper closed subspace, and fix $0 < \alpha < 1$. Then there exits $x \notin W$ with $\|x\| = 1$ so that $\inf_{y \in W} \|x - y\| \geq \alpha$. (Hint: Start with $x_0 \notin W$, set $d = \inf_{y \in W} \|x_0 - y\|$, take $x_1 \in W$ so that $\|x_0 - x_1\| \geq d - \varepsilon$ for some suitably chosen $\varepsilon > 0$, and show that $x = \|x_0 - x_1\|^{-1}(x_0 - x_1)$ works.)

Exercise 1.3.8. Let V be a normed space such that the closed unit ball $\overline{B}(1, 0)$ is compact. Then V is finite dimensional.

If $(V, \|\cdot\|)$ is a normed space, a series $\sum_{n=1}^{\infty} x_n$ is said to converge if the partial sums $\sum_{n=1}^{k} x_n$ converge as k tends to infinite. A series $\sum_{n=1}^{\infty} x_n$ is said to converge absolutely if $\sum_{n=1}^{\infty} \|x_n\| < \infty$.

Exercise 1.3.9. Let $(V, \|\cdot\|)$ be a normed space over $\mathbb{K}$. Then V is a Banach space if and only if every absolutely convergent series converges.

Exercise 1.3.10. Let $(V, \|\cdot\|)$ be a normed space over $\mathbb{K}$. Then the metric space completion $\overline{V}$ of of V has a vector space structure which extends the vector space structure of V. Thus, every normed space is a dense linear subspace of a Banach space.

Chapter 2

Measure and integration

Suppose we wanted to assign the notion of size (or measure) to a collection $\mathcal{M}$ of certain subsets of a $\mathbb{R}^n$. That is, for a subset $E \in \mathcal{M}$ we want to assign a number $0 \leq \mu(E) \leq \infty$ which tells us in some sense how large E is. Then we might want the following properties to hold:

(a) $\mathcal{M} = 2^{\mathbb{R}^n}$.

(b) μ is **countably additive**: If $\{E_j\}_{j=1}^\infty$ is a sequence of disjoint sets in $\mathcal{M}$, then $\mu(\cup_{j=1}^\infty E_j) = \sum_{j=1}^\infty \mu(E_j)$.

(c) If E can be transformed to F using translations, rotations, and reflections, then $\mu(E) = \mu(F)$.

(d) μ assigns a finite, nonzero value to the unit cube.

Unfortunately, these conditions are mutually inconsistent. This was first noticed by Vitali in 1905. Suppose we had such a function $\mu : 2^{\mathbb{R}} \to [0,\infty]$. Consider the equivalence relation on $[0,1)$ which is given by $s \sim t$ if $t - s \in \mathbb{Q}$. Take $E \subset [0,1)$ so that E contains exactly one element from each equivalence class. For each $t \in \mathbb{Q}$ consider the set $E_t = E + t \mod 1$, i.e.,

$$E_t = ((E+t) \cap [t,1)) \cup ((E+t-1) \cap [0,t)).$$

Then $\{E_t\}_{t\in\mathbb{Q}}$ is a countable family of pairwise disjoint sets which cover $[0,1)$. Note that for each $t \in \mathbb{Q}$ we have

$$\begin{aligned}\mu(E_t) &= \mu((E+t) \cap [t,1)) + \mu((E+t-1) \cap [0,t)) \\ &= \mu(E \cap [0,1-t)) + \mu(E \cap [1-t,1)) = \mu(E).\end{aligned}$$

Hence $\mu([0,1)) = \mu(\cup_{t\in\mathbb{Q}} E_t) = \sum_{t\in\mathbb{Q}} \mu(E_t) = \sum_{t\in\mathbb{Q}} \mu(E)$, so that $\mu([0,1)) \in \{0,\infty\}$. A contradiction then follows easily.

We must therefore compromise of some of the conditions above. Conditions (c) and (d) seem essential to having a good notion of size, thus we look to weaken conditions (a) or (b). One thing we might try is to weaken countable additivity to **finite additivity**:

(b') If $\{E_j\}_{j=1}^k$ is a finite sequence of disjoint sets in $\mathcal{M}$, then $\mu(\cup_{j=1}^\infty E_j) = \sum_{j=1}^k \mu(E_j)$.

The question of whether there exists a function μ satisfying $(a), (b'), (c)$, and (d) is quite interesting, and we'll come back to this later. (It turns out that such a μ exists when $n \leq 2$, and does not exist otherwise!)

Another possibility is to not try to measure every subset of $\mathbb{R}^n$, but rather only a certain nice class $\mathcal{M}$ which excludes Vitali's set above. We would want $\mathcal{M}$ to contains all intervals, and to be closed under taking countable unions and complements. In this case, one can indeed obtain such a μ, as was first shown by Lebesgue in 1901 (his dissertation!). Before we present Lebesgue's proof we first take a detour to the abstract setting.

2.1 Measurable sets and functions

Let X be a nonempty set. An **algebra** of subsets of X is a nonempty collection $\mathcal{A}$ of subsets of X which is closed under finite unions and complements. A **σ-algebra** is a nonempty collection $\mathcal{E}$ of subsets of X which is closed under countable unions and complements. Observe that σ-algebras are also closed under countable intersection. Also, observe that we have $\emptyset, X \in \mathcal{E}$.

Note that the intersection of any family of σ-algebras is again a σ-algebra. It follows that if $\mathcal{A}$ is any collection of subsets of X, then there is a unique smallest σ-algebra $\mathcal{M}(\mathcal{A})$ which contains $\mathcal{A}$. $\mathcal{M}(\mathcal{A})$ is the σ-algebra **generated by $\mathcal{A}$**. If X is a metric space, then the **Borel σ-algebra** is the σ-algebra $\mathcal{B}(X)$ generated by the open subsets of X.

A **measurable space** is a pair, consisting of a set X, together with a σ-algebra of subsets of X. Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be two measurable spaces. A function $f : X \to Y$ is **measurable** if $f^{-1}(E) \in \mathcal{M}$ for all $E \in \mathcal{N}$. We denote by $\mathcal{M}(X; Y)$ the set of all measurable functions from X to Y (with the underlying σ-algebras implicit). We denote by $\mathcal{M}(X) = \mathcal{M}(X; \mathbb{C})$ where $\mathbb{C}$ is endowed with the Borel σ-algebra. Thus, $f \in \mathcal{M}(X)$ if and only if $f^{-1}(E) \in \mathcal{M}$ for any Borel set $E \subset \mathbb{C}$.

Lemma 2.1.1. *Suppose $(X, \mathcal{M}), (Y, \mathcal{N})$, and $(Z, \mathcal{P})$ are measurable spaces and $f : X \to Y$, $g : Y \to Z$ are measurable, then $g \circ f : X \to Z$ is measurable.*

Proof. If $E \in \mathcal{P}$ then $(g \circ f)^{-1}(E) = f^{-1}(g^{-1}(E))$ and the result is immediate. ■

Proposition 2.1.2. *Suppose $\mathcal{N}$ is generated as a σ-algebra by $\mathcal{E} \subset \mathcal{N}$. A function $f : X \to Y$ is measurable if and only if $f^{-1}(E) \in \mathcal{M}$ for all $E \in \mathcal{E}$.*

Proof. We let $\mathcal{A} = \{E \subset Y \mid f^{-1}(E) \in \mathcal{M}\}$. Then $\mathcal{E} \subset \mathcal{A}$ and hence it is enough to show that $\mathcal{A}$ is a σ-algebra. Note that $\emptyset \in \mathcal{A}$. If $E \in \mathcal{A}$, then $f^{-1}(E^c) = f^{-1}(E)^c$ and hence $E^c \in \mathcal{A}$. Also, if $\{E_n\}_{n \in \mathbb{N}} \subset \mathcal{A}$, then $f^{-1}(\cup_{n \in \mathbb{N}} E_n) = \cup_{n \in \mathbb{N}} f^{-1}(E_n)$ and hence $\cup_{n \in \mathbb{N}} E_n \in \mathcal{A}$. Therefore, $\mathcal{A}$ is a σ-algebra. ■

Corollary 2.1.3. *Suppose X and Y are metric spaces and $f : X \to Y$ is continuous, then f is measurable with respect to the Borel σ-algebras.*

Proof. Since a function is continuous if and only if the inverse images of open sets are open, and since the Borel σ-algebra is generated by open sets this follows from the previous proposition. ∎

Proposition 2.1.4. *Let (X, d) be a separable metric space, then $\mathcal{B}(X)$ is generated by the open balls $B(r, x)$, for $x \in X$ and $r > 0$.*

Proof. Let $O \subset X$ be open and let $\{x_n\}_{n\in\mathbb{N}} \subset O$ be a countable dense subset. For each $n \in \mathbb{N}$ we let r_n denote the supremum over all $r > 0$ so that $B(r, x_n) \subset O$. Then $B(r_n, x_n) \subset O$, and by density it follows that $\cup_{n\in\mathbb{N}} B(r_n, x_n) = O$.

Thus, any open set is contained in the σ-algebra generated by open balls and hence this is also true for any Borel set. ∎

Corollary 2.1.5. *Suppose $f : X \to \mathbb{R}$. Then the following conditions are equivalent:*

1. $f \in \mathcal{M}(X; \mathbb{R})$.
2. $f^{-1}(O) \in \mathcal{M}$ *for any open set* $O \subset \mathbb{R}$.
3. $f^{-1}((a, b)) \in \mathcal{M}$ *for any* $a, b \in \mathbb{R}$.
4. $f^{-1}((-\infty, b)) \in \mathcal{M}$ *for any* $b \in \mathbb{R}$.

Proposition 2.1.6. *Let $(X, \mathcal{M})$ be a measurable space.*

1. *If $f \in \mathcal{M}(X)$, and $\phi : \mathbb{C} \to \mathbb{C}$ is continuous, then $\phi \circ f \in \mathcal{M}(X)$.*
2. *If $f, g \in \mathcal{M}(X)$, and $\alpha \in \mathbb{C}$, then $\alpha f, f+g, fg, |f|, \mathrm{Re}\,(f), \mathrm{Im}\,(f) \in \mathcal{M}(X)$.*
3. *If $f, g \in \mathcal{M}(X; \mathbb{R})$ then $\max\{f, g\}, \min\{f, g\} \in \mathcal{M}(X; \mathbb{R})$.*

Proof. The first assertation follows from Lemma 2.1.1 and Corollary 2.1.3. It then follows that if f is measurable then so is αf, $|f|$, $\mathrm{Re}\,(f)$, and $\mathrm{Im}\,(f)$, since multiplication by α, absolute value, and taking real and imaginary parts are continuous functions.

More generally, consder $\mathbb{C}^2$ with the metric $d((a, b), (x, y)) = \max\{|a-x|, |b-y|\}$. Then we have $B(r, (a, b)) = B(r, a) \times B(r, b)$, and if $f, g \in \mathcal{M}(X)$ then the function given by $(f, g)(x) = (f(x), g(x))$ satisfies

$$(f, g)^{-1}(B(r, (a, b))) = f^{-1}(B(r, a)) \cap g^{-1}(B(r, b))$$

and hence is measurable. Then if $\phi : \mathbb{C}^2 \to \mathbb{C}$ is continuous we must have that $\phi \circ (f, g)$ is again measurable.

Since addition and multiplication are continuous on $\mathbb{C}$, and since maximum and minimum are continuous on $\mathbb{R}$, it then follows that if $f, g \in \mathcal{M}(X)$ then $f+g, fg \in \mathcal{M}(X)$, and if $f, g \in \mathcal{M}(X; \mathbb{R})$ then $\max\{f, g\}, \min\{f, g\} \in \mathcal{M}(X; \mathbb{R})$. ∎

Proposition 2.1.7. *Suppose* $\{f_n\}_{n\in\mathbb{N}} \subset \mathcal{M}(X)$*, and* $f : X \to \mathbb{C}$ *so that* $f_n(x) \to f(x)$*, for each* $x \in X$*. Then* $f \in \mathcal{M}(X)$*.*

Proof. Since a function f is measurable if and only if its real and imaginary parts are measurable we may assume that $f_n \in \mathcal{M}(X;\mathbb{R})$ for each $n \in \mathbb{N}$.

If $r \in \mathbb{R}$, then $f(t) < r$ if and only if there exists $k, N \in \mathbb{N}$ so that $f_n(t) < r - 1/k$ for all $n \geq N$. Hence,

$$f^{-1}((-\infty, r)) = \cup_{k\in\mathbb{N}} \cup_{N\in\mathbb{N}} \cap_{n\geq N} f_n^{-1}((-\infty, r - 1/k)),$$

and thus $f^{-1}((-\infty, r))$ is measurable. It then follows that f is measurable from Corollary 2.1.5. ∎

If $E \in \mathcal{M}$, then the **characteristic** (or **indicator**) function on E is the function $1_E : X \to \mathbb{C}$ given by $1_E(x) = \begin{cases} 1 & \text{if } x \in E, \\ 0 & \text{if } x \notin E. \end{cases}$ Clearly, characteristic functions are measurable. A **simple function** is a finite complex linear combination of characteristic functions. Simple functions are also measurable by Proposition 2.1.6, and from the previous proposition we have that any pointwise limit of simple functions is then measurable.

Proposition 2.1.8. *If* $f \in \mathcal{M}(X)$*, then* f *is a pointwise limit of simple functions. If* f *is bounded then* f *is a uniform limit of simple functions.*

Proof. By considering the real and imaginary parts it is enough to consider the case $f \in \mathcal{M}(X;\mathbb{R})$. For $N \in \mathbb{N}$, and $-N^2 \leq k \leq N^2$ we let $E_{N,k} = f^{-1}\left(\left[\frac{k}{N}, \frac{k+1}{N}\right)\right)$, and set

$$f_N = \sum_{-N^2 \leq k \leq N^2} \frac{k}{N} 1_{E_{N,k}}.$$

Then if $f(x) \in [-N, N)$ we have $|f(x) - f_N(x)| \leq 1/N$, and the result follows. ∎

2.1.1 Exercises

Exercise 2.1.9. Suppose we have an algebra $\mathcal{C} \subset 2^X$ with the proeprty that if $E_n \in \mathcal{C}$ and $E_n \subset E_{n+1}$, for $n \geq 1$, then $\cup_{n=1}^{\infty} E_n \in \mathcal{C}$. Then $\mathcal{C}$ is a σ-algebra.

Exercise 2.1.10. Suppose $\mathcal{M} \subset 2^X$ is an algebra, such that if $\{E_n\}_{n=1}^{\infty} \subset \mathcal{M}$ are pairwise disjoint then $\cup_{n=1}^{\infty} E_n \in \mathcal{M}$. Then $\mathcal{M}$ is a σ-algebra.

Exercise 2.1.11. Suppose $(X, \mathcal{M})$ is a measurable space and $f, g \in \mathcal{M}(X;\mathbb{R})$. The sets

$$\{x \in X \mid f(x) < g(x)\} \qquad \text{and} \qquad \{x \in X \mid f(x) = g(x)\}$$

are measurable.

Exercise 2.1.12. Suppose $(X, \mathcal{M})$ is a measurable space and $\{f_n\}_{n\in\mathbb{N}} \subset \mathcal{M}(X;\mathbb{R})$. Set

$$\mathcal{C} = \{x \in X \mid \{f_n(x)\}_{n\in\mathbb{N}} \text{ converges}\}.$$

Then $\mathcal{C}$ is measurable.

Exercise 2.1.13. Suppose $(X, \mathcal{M}, \mu)$ is a measure space, $(Y, \mathcal{N})$ is a measurable space and $\theta : X \to Y$ is measurable. For each set $E \subset Y$ we set $\theta_*\mu(E) = \mu(\theta^{-1}(E))$. Then $\theta_*\mu$ is a measure on $(Y, \mathcal{N})$ called the **push forward measure of μ with respect to θ**.

Exercise 2.1.14. The set Borel σ-algebra $\mathcal{B} \subset 2^{\mathbb{R}}$ has cardinality $|\mathbb{R}|$. Hint: To show $|\mathcal{B}| \leq |\mathcal{R}|$ let $\mathcal{B}_0$ denote the set of open intervals, and inductively define for each ordinal $\alpha < \omega_1$ the set $\mathcal{B}_{\alpha+1}$ to consist of all sets of the form $(\cup_{i=1}^\infty E_i) \cup (\cup_{j=1}^\infty F_j^c)$, where $E_i, f_i \in B_\alpha$, and for each limit ordinal $\alpha < \omega_1$ set $\mathcal{B}_\alpha = \cup_{\beta<\alpha}\mathcal{B}_\beta$. Then show that $|\mathcal{B}_\alpha| \leq |\mathbb{R}|$ for each $\alpha < \omega_1$ and $\mathcal{B} = \cup_{\alpha<\omega_1}\mathcal{B}_\alpha$.

2.2 Measures

If $(X, \mathcal{M})$ is a measurable space, then a **measure** on $(X, \mathcal{M})$ is a set function $\mu : \mathcal{M} \to [0, \infty]$ that satisfies

1. $\mu(\emptyset) = 0$.
2. μ is **countably additive**: if $\{E_n\}_{n=1}^\infty$ is a sequence of disjoint sets in $\mathcal{M}$, then $\mu(\cup_{n=1}^\infty E_n) = \sum_{n=1}^\infty \mu(E_n)$.

A **measure space** is a triple $(X, \mathcal{M}, \mu)$ where $(X, \mathcal{M})$ is a measurable space and μ is a measure on $(X, \mathcal{M})$.

Here are some basic properties of measures:

Proposition 2.2.1. *Let $(X, \mathcal{M}, \mu)$ be a measure space.*

1. *(**Monotonicity**) If $E, F \in \mathcal{M}$ and $E \subset F$, then $\mu(E) \leq \mu(F)$.*
2. *(**Subadditivity**) If $\{E_n\}_{n=1}^\infty \subset \mathcal{M}$, then $\mu(\cup_{n=1}^\infty E_n) \leq \sum_{n=1}^\infty \mu(E_n)$.*
3. *(**Continuity from below**) If $\{E_n\}_{n=1}^\infty \subset \mathcal{M}$ and $E_1 \subset E_2 \subset \cdots$ then $\mu(\cup_{n=1}^\infty E_n) = \lim_{n\to\infty} \mu(E_n)$.*
4. *(**Continuity from above**) If $\{E_n\}_{n=1}^\infty \subset \mathcal{M}$ and $E_1 \supset E_2 \supset \cdots$, with $\mu(E_1) < \infty$, then $\mu(\cap_{n=1}^\infty E_n) = \lim_{n\to\infty} \mu(E_n)$.*

Proof. If $E, F \in \mathcal{M}$ with $E \subset F$, then we have $\mu(F) = \mu(E) + \mu(F \setminus E) \geq \mu(E)$ showing monotonicity.

If $\{E_n\}_{n=1}^\infty \subset \mathcal{M}$, then setting $F_n = E_n \setminus (\cup_{k<n} E_k)$ we have $F_n \subset E_n$, $\{F_n\}_{n=1}^\infty \subset \mathcal{M}$ are pairwise disjoint, and $\cup_{n=1}^\infty F_n = \cup_{n=1}^\infty E_n$. From monotonicity we then obtain subadditivity:

$$\mu(\cup_{n=1}^\infty E_n) = \mu(\cup_{n=1}^\infty F_n) = \sum_{n=1}^\infty \mu(F_n) \leq \sum_{n=1}^\infty \mu(E_n).$$

If $\{E_n\}_{n=1}^\infty \subset \mathcal{M}$ and $E_1 \subset E_2 \subset \cdots$, then setting $F_1 = E_1$, and $F_n = E_n \setminus E_{n-1}$ for $n > 1$ we have that $\{F_n\}_{n=1}^\infty$ are pairwise disjoint and hence

$$\begin{aligned}\mu(\cup_{n=1}^\infty E_n) = \mu(\cup_{n=1}^\infty F_n) &= \sum_{n=1}^\infty \mu(F_n) \\ &= \lim_{N\to\infty} \sum_{n=1}^N \mu(F_n) = \lim_{N\to\infty} \mu(\cup_{n=1}^N F_n) = \lim_{N\to\infty} \mu(E_N).\end{aligned}$$

If $\{E_n\}_{n=1}^\infty \subset \mathcal{M}$ and $E_1 \supset E_2 \supset \cdots$, then taking $F_n = E_1 \setminus E_n$, we have $F_1 \subset F_2 \subset \cdots$. By continuity from below we have $\mu(\cup_{n=1}^\infty F_n) = \lim_{n\to\infty} \mu(F_n)$. Since $\mu(E_1) = \mu(E_n) + \mu(F_n) = \mu(\cup_{n=1}^\infty F_n) + \mu(\cap_{n=1}^\infty E_n)$, and since $\mu(E_1) < \infty$ we have

$$\mu(\cup_{n=1}^\infty E_n) = \mu(E_1) - \mu(\cup_{n=1}^\infty F_n) = \lim_{n\to\infty} \mu(E_1) - \mu(F_n) = \lim_{n\to\infty} \mu(E_n).$$

■

A measure μ is **finite** if $\mu(X) < \infty$. μ or $(X, \mathcal{M}, \mu)$ is **σ-finite**, if $X = \cup_{n=1}^\infty E_n$ where $E_n \in \mathcal{M}$ with $\mu(E_n) < \infty$. μ or $(X, \mathcal{M}, \mu)$ is **semifinite** if for all $E \in \mathcal{M}$ with $\mu(E) > 0$ there exists $A \in \mathcal{M}$ with $A \subset E$ such that $0 < \mu(A) < \infty$.

Note that if $(X, \mathcal{M}, \mu)$ is σ-finite then it must also be semifinite. Indeed, if $X = \cup_{n=1}^\infty E_n$ with $E_n \in \mathcal{M}$, $\mu(E_n) < \infty$, and if $E \in \mathcal{M}$ with $\mu(E) > 0$, then $\mu(E \cap E_n) \leq \mu(E_n) < \infty$, for all n and we have $0 < \mu(E \cap E_n)$ for at least one n since $0 < \mu(E) \leq \sum_{n=1}^\infty \mu(E \cap E_n)$.

A measure μ or measure space $(X, \mathcal{M}, \mu)$ has the **essential suprema property** if for any $\mathcal{E} \subset \mathcal{M}$ there exists $E \in \mathcal{M}$ such that $\mu(A \setminus E) = 0$ for all $A \in \mathcal{E}$, and if $E_0 \in \mathcal{M}$ is any other measurable set which satisfies $\mu(A \setminus E_0) = 0$ for all $A \in \mathcal{E}$ then we also have $\mu(E \setminus E_0) = 0$. A measure μ or measure space $(X, \mathcal{M}, \mu)$ is **localizable** if it is semifinite and has the essential suprema property.

Here are some examples of measure spaces:

1. If X is a set and $\mathcal{M} = 2^X$, then the **counting measure on** X is given by $\mu(E) = |E|$ if E is finite, and $\mu(E) = \infty$ if E is infinite. It's not hard to check that this space is always localizable and it is σ-finite if and only if X is countable.

2. If X is a nonempty set and $\mathcal{M} = 2^X$, then the **Dirac measure** (or **point mass**) at $x_0 \in X$ is given by $\mu(E) = 1$ if $x_0 \in E$, and $\mu(E) = 0$ if $x_0 \notin E$.

3. If $(X, \mathcal{M}, \mu)$ is a measure space and $E \in \mathcal{M}$ then we may consider a new measure μ_E on $(X, \mathcal{M})$ given by $\mu_E(F) = \mu(F \cap E)$. This is the **restriction measure** on E.

4. Suppose X is a set and $\mathcal{M}$ consists of all sets $E \subset X$ such that either E or E^c is countable. Then counting measure restricted to $\mathcal{M}$ gives a measure. This measure is always semifinite and satisfies the essential suprema property if and only if X is countable.

5. Suppose $(X, \mathcal{M})$ is a measurable space and $\mathcal{N} \subset \mathcal{M}$ is a non-empty collection of subsets such that $\mathcal{N}$ is closed under countable union and whenever we have $E \in \mathcal{N}$ and $F \in \mathcal{M}$ with $F \subset E$ then $F \in \mathcal{N}$. We define the measure μ^∞ on $\mathcal{M}$ by setting $\mu^\infty(E) = \begin{cases} 0 & \text{if } E \in \mathcal{N}, \\ \infty & \text{if } E \notin \mathcal{N}, \end{cases}$ Then μ gives a measure on $\mathcal{M}$. This will be semifinite if and only if $\mathcal{N} = \mathcal{M}$.

Given a measure space $(X, \mathcal{M}, \mu)$, we say a set $E \subset X$ is **σ-finite** if $E = \cup_{n=1}^\infty E_n$ where $E_n \in \mathcal{M}$ with $\mu(E_n) < \infty$. We say that E is a **null set** if $\mu(E) = 0$. We say that E is **conull** if E^c is null. A property is said to hold **almost everywhere** (or **μ-almost everywhere**) if it holds on a conull set. The collection of null sets is non-empty, and closed under countable unions and taking measurable subsets, therefore given any measure space $(X, \mathcal{M}, \mu)$ we may consider the corresponding measure μ^∞ as described above. Then μ^∞ will satisfy the essential suprema property if and only if μ does.

In practice most interesting measure spaces one encounters are localizable. In part because these are the spaces in which a nice integration theory can be developed. The latter two examples above show that there do exist more general measure spaces, however we shall view these spaces as pathalogical.

Lemma 2.2.2. *Suppose $(X, \mathcal{M}, \mu)$ is a measure space and $\{F_n\}_{n=1}^\infty \subset \mathcal{M}$ is a countable partition of X so that μ_{F_n} has the essential suprema property for each $n \geq 1$, then μ has the essential suprema property.*

Proof. Suppose $\mathcal{E} \subset \mathcal{M}$, and for each n take $E_n \in \mathcal{M}$ so that $\mu(F_n \cap (A \setminus E_n)) = 0$ for all $A \in \mathcal{E}$, and if $E_0 \in \mathcal{M}$ is such that $\mu(F_n \cap (A \setminus E_0)) = 0$ for all $A \in \mathcal{E}$ then we have $\mu(F_n \cap (E_n \setminus E_0)) = 0$.

Set $E = \cup_{n=1}^\infty (E_n \cap F_n)$. If $A \in \mathcal{E}$ then we have

$$\mu(A \setminus E) = \sum_{n=1}^\infty \mu(F_n \cap (A \setminus E)) = \sum_{n=1}^\infty \mu(F_n \cap (A \setminus E_n)) = 0.$$

Also, if $E_0 \in \mathcal{M}$ such that $\mu(A \setminus E_0) = 0$ for all $A \in \mathcal{E}$ then we have

$$\mu(E \setminus E_0) = \sum_{n=1}^\infty \mu(F_n \cap (E \setminus E_0)) = \sum_{n=1}^\infty \mu(F_n \cap (E_n \setminus E_0)) = 0.$$

∎

Proposition 2.2.3. *Suppose $(X, \mathcal{M}, \mu)$ is a σ-finite measure space, then $(X, \mathcal{M}, \mu)$ is localizable.*

Proof. We already noted above that μ is semifinite, thus we only need to show that it satisfies the essential suprema property. By the previous lemma it is enough to consider the case when μ is finite. Suppose $\mathcal{E} \subset \mathcal{M}$, and let

$$\mathcal{E}^+ = \{E \in \mathcal{M} \mid \mu(A \setminus E) = 0 \text{ for all } A \in \mathcal{E}\}.$$

Note that $X \in \mathcal{E}^+$ and $\mathcal{E}^+$ is closed under countable intersection. Let $a = \inf\{\mu(E) \mid E \in \mathcal{E}^+\}$, then there exists a sequence $\{E_k\}_{k=1}^\infty \subset \mathcal{E}^+$ so that $\mu(E_k) \to a$. We set $E = \cap_{k=1}^\infty E_k$ so that $E \in \mathcal{E}^+$ and $\mu(E) \leq \inf_{k\to\infty} \mu(E_k) = a \leq \mu(E)$.

If $E_0 \in \mathcal{E}^+$, then $E_0 \cap E \in \mathcal{E}^+$ and hence $\infty > \mu(E_0 \cap E) \geq a = \mu(E)$. We then have $0 = \mu(E) - \mu(E_0 \cap E) \geq \mu(E \setminus E_0)$. ■

Proposition 2.2.4. *If $(X, \mathcal{M}, \mu)$ is a semifinite measure space, then for all $E \in \mathcal{M}$ we have*

$$\mu(E) = \sup\{\mu(A) \mid A \subset E, A \in \mathcal{M}, \text{ and } \mu(A) < \infty\}.$$

Proof. If $\mu(E) < \infty$ then this is obvious, therefore we may assume that $\mu(E) = \infty$. We let $a = \sup\{\mu(A) \mid A \subset E, A \in \mathcal{M}, \text{ and } \mu(A) < \infty\}$, and take $A_n \subset E, A_n \in \mathcal{M}$ such that $\mu(A_n) \to a$. We set $B_k = \cup_{n=1}^k A_n$ and $B = \cup_{n=1}^\infty A_n$. Then $\mu(B_k) \to \mu(B)$, and $B_k \subset E$, hence $a \geq \mu(B_k) \geq \mu(A_k) \to a$, so that $\mu(B) = a$. If we had $a < \infty$ then $\mu(E \setminus B) = \infty$ and so by semifiniteness there exists $A_0 \in \mathcal{M}$, $A_0 \subset E \setminus B$ so that $0 < \mu(A_0) < \infty$. We then have $A_n \cup A_0 \subset E$ and $\mu(A_n \cup A_0) = \mu(A_n) + \mu(A_0)$, therefore

$$a \geq \mu(A_n \cup A_0) = \mu(A_n) + \mu(A_0) \to a + \mu(A_0) > a,$$

which cannot happen. Thus, we must have $a = \infty = \mu(E)$. ■

Proposition 2.2.5. *Suppose $(X, \mathcal{M}, \mu)$ has the essential suprema property, and $\mathcal{F} \subset \mathcal{M}(X, [0, \infty])$. Then there exists $h \in \mathcal{M}(X, [0, \infty])$ so that*

$$\mu(\{x \in X \mid h(x) < f(x)\}) = 0$$

for each $f \in \mathcal{F}$, and if $\tilde{h} \in \mathcal{M}(X, [0, \infty])$ is any other function with this property then we have

$$\mu(\{x \in X \mid \tilde{h}(x) < h(x)\}) = 0.$$

Proof. For each $k \geq 0$, $n \geq 1$ consider the collection $\mathcal{E}_{k,n} = \{f^{-1}([k/n, \infty] \mid f \in \mathcal{F}\}$, and let $E_{k,n}$ be such that $\mu(A \setminus E_{k,n}) = 0$ for all $A \in \mathcal{E}_{k,n}$ and if E_0 is another measurable set with this property then we have $\mu(E \setminus E_0) = 0$. It then follows that $\mu(E_{k,n} \setminus E_{k',n'}) = 0$ whenever $k/n \geq k'/n'$.

Let $h(x) = \sup_{k\geq 0, n\geq 1}\{k/n \mid x \in E_{k,n}\}$. Then for each $f \in \mathcal{F}$ we have

$$\mu(\{x \in X \mid h(x) < f(x)\}) = \mu(\cup_{k\geq 0, n\geq 1}\{x \in X \mid h(x) \leq k/n < f(x)\}) = 0.$$

Moreover, if $\tilde{h}$ is another measurable function with this property and if for each $k \geq 0, n \geq 1$ we set $\tilde{E}_{n,k} = h^{-1}([k/n, \infty])$, then we have $\mu(f^{-1}(k/n, \infty]) \setminus \tilde{E}_{n,k}) = 0$ for every $f \in \mathcal{F}$ and hence $\mu(E_{k,n} \setminus \tilde{E}_{k,n}) = 0$. It then follows that

$$\mu(\{x \in X \mid \tilde{h}(x) < h(x)\}) = 0.$$

■

We call a function h in the previous proposition an **essential supremum** of $\mathcal{F}$.

2.2.1 Outer measures

An **outer measure** is a set function $\mu^* : 2^X \to [0, \infty]$ that satisfies

1. $\mu^*(\emptyset) = 0$.
2. (**Monotonicity**) $\mu^*(A) \leq \mu^*(B)$ if $A \subset B$.
3. (**Subadditivity**) $\mu^*(\cup_{n\in\mathbb{N}} A_n) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$.

Proposition 2.2.6. *Suppose $\mathcal{S} \subset 2^X$ and $\mu_0 : \mathcal{S} \to [0, \infty]$ is such that $\emptyset \in \mathcal{S}$, and $\mu_0(\emptyset) = 0$. For $E \subset X$ define*

$$\mu^*(A) = \inf \left\{ \sum_{n=1}^{\infty} \mu_0(E_n) \mid E_n \in \mathcal{S} \text{ and } A \subset \cup_{n=1}^{\infty} E_n \right\}.$$

Then μ^* is an outer measure.

Proof. Since $\emptyset$ covers itself we clearly have $\mu^*(\emptyset) = 0$.

If $A \subset B \subset X$, then as any cover of B also covers A it follows that the set for which we are taking the infimum for B is contained in the corresponding set for A. Therefore $\mu^*(A) \leq \mu^*(B)$.

Fix $\varepsilon > 0$. If $\{A_n\}_{n\in\mathbb{N}} \subset 2^X$, then for each $n \in \mathbb{N}$ there exists $\{E_j^n\}_{j\in\mathbb{N}} \subset \mathcal{S}$ so that $\sum_{j\in\mathbb{N}} \mu_0(E_j^n) < \mu^*(A_n) + \varepsilon 2^{-n}$. We then have $\cup_{n\in\mathbb{N}} A_n \subset \cup_{n,j\in\mathbb{N}} E_j^n$, and so

$$\mu^*(\cup_{n\in\mathbb{N}} A_n) \leq \sum_{n,j\in\mathbb{N}} \mu_0(E_j^n) < \varepsilon + \sum_{n\in\mathbb{N}} \mu^*(A_n).$$

As $\varepsilon > 0$ was arbitrary it then follows that

$$\mu^*(\cup_{n\in\mathbb{N}} A_n) \leq \sum_{n\in\mathbb{N}} \mu^*(A_n).$$

■

The outer measure μ^* in the previous proposition is called the **outer measure associated to** μ_0.

If μ^* is an outer measure on X, then a set $A \subset X$ is μ^***-measurable** if

$$\mu^*(S) = \mu^*(S \cap A) + \mu^*(S \cap A^c) \text{ for all } S \subset X.$$

Theorem 2.2.7 (Carathéodory)**.** *Suppose μ^* is an outer measure on X, then the collection $\mathcal{M}$ of all μ^*-measurable sets is a σ-algebra, and the restriction of μ^* to $\mathcal{M}$ is a measure.*

Proof. Since $\mu^*(\emptyset) = 0$ we have $\mu^*(S) = \mu^*(\emptyset) + \mu^*(S)$ for each $S \subset X$, hence $\emptyset \in \mathcal{M}$. Also, note that $\mathcal{M}$ is clearly closed under taking complements.

If $A, B \in \mathcal{M}$, then for each $S \subset X$ we have

$$\begin{aligned}\mu^*(S) &= \mu^*(S \cap A) + \mu^*(S \cap A^c)\\ &= \mu^*(S \cap A \cap B) + \mu^*(S \cap A \cap B^c) + \mu^*(S \cap A^c \cap B) + \mu^*(S \cap A^c \cap B^c)\\ &\geq \mu^*(S \cap (A \cup B)) + \mu^*(S \cap (A \cup B)^c) \geq \mu^*(S).\end{aligned}$$

We therefore have that $A \cup B \in \mathcal{M}$ and if A and B are disjoint then taking $S = A \cup B$ we have

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

It then follows easily that $\mathcal{M}$ is closed under unions of finite families, and hence $\mathcal{M}$ is an algebra. To show that $\mathcal{M}$ is a σ-algebra it is then enough to show that $\mathcal{M}$ is closed under taking countable unions of pairwise disjoint families.

If $\{A_n\}_{n\in\mathcal{N}} \subset \mathcal{M}$ is a sequence of pairwise disjoint sets, then set $B_n = \cup_{k=1}^n A_k$ and $B = \cup_{k=1}^\infty A_k$. Since $A_n \in \mathcal{M}$, if $S \subset X$, and $n > 1$ we have

$$\begin{aligned} \mu^*(S \cap B_n) &= \mu^*(S \cap B_n \cap A_n) + \mu^*(S \cap B_n \cap A_n^c) \\ &= \mu^*(S \cap A_n) + \mu^*(S \cap B_{n-1}). \end{aligned}$$

By induction it then follows easily that

$$\mu^*(S \cap B_n) = \sum_{k=1}^n \mu^*(S \cap A_n).$$

Hence,

$$\begin{aligned} \mu^*(S) &= \mu^*(S \cap B_n) + \mu^*(S \cap B_n^c) \\ &\geq \sum_{k=1}^n \mu^*(S \cap A_k) + \mu^*(S \cap B^c). \end{aligned}$$

Taking $n \to \infty$ then gives

$$\begin{aligned} \mu^*(S) &\geq \sum_{k=1}^\infty \mu^*(S \cap A_k) + \mu^*(S \cap B^c) \\ &\geq \mu^*(S \cap B) + \mu^*(S \cap B^c) \geq \mu^*(S). \end{aligned} \tag{2.1}$$

Thus, $B \in \mathcal{M}$, showing that $\mathcal{M}$ is a σ-algebra. Taking $S = B$ in (2.1) shows

$$\mu^*(B) = \sum_{k=1}^\infty \mu^*(A_k).$$

Hence μ^* defines a measure on $\mathcal{M}$. ■

2.2.2 Carathéodory's extension theorem

If $\mathcal{A} \subset 2^X$ is an algebra, a function $\mu_0 : \mathcal{A} \to [0, \infty]$ is a **premeasure** if

1. $\mu_0(\emptyset) = 0$.
2. Whenever $\{E_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ are disjoint such that $\cup_{n=1}^\infty E_n \in \mathcal{A}$, then we have $\mu_0(\cup_{n=1}^\infty E_n) = \sum_{n=1}^\infty \mu_0(E_n)$.

Theorem 2.2.8 (Carathéodory's extension theorem)**.** *Suppose $\mathcal{A} \subset 2^X$ is an algebra, $\mu_0 : \mathcal{A} \to [0, \infty]$ is a premeasure, and μ^* is the associated outer measure, then every set $E \in \mathcal{A}$ is μ^*-measurable and we have $\mu^*(E) = \mu_0(E)$.*

Moreover, if $\mathcal{M}$ denotes the σ-algebra generated by $\mathcal{A}$, and if μ^* defines a semifinite measure on $\mathcal{M}$, then μ^* is the unique measure on $\mathcal{M}$ which extends μ_0.

Proof. If $E, A_n \in \mathcal{A}$, for $n \geq 1$ with $E \subset \cup_{n=1}^\infty A_n$, then setting $B_n = E \cap (A_n \setminus (\cup_{k=1}^{n-1} A_k))$ we have that $B_n \subset A_n$, and $\{B_n\}_{n=1}^\infty$ is a family of pairwise disjoint sets in $\mathcal{A}$ such that $E = \cup_{n=1}^\infty B_n$. We therefore have $\mu_0(E) = \sum_{n=1}^\infty \mu_0(B_n) \leq \sum_{n=1}^\infty \mu_0(A_n)$. Thus, it follows that $\mu_0(E) \leq \mu^*(E) \leq \mu_0(E)$. Hence, μ^* is an extension of μ_0.

If $A \in \mathcal{A}$, $S \subset X$, and $\varepsilon > 0$, then we may take $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ so that $S \subset \cup_{n=1}^\infty A_n$ and $\sum_{n=1}^\infty \mu_0(A_n) \leq \mu^*(S) + \varepsilon$. Since μ_0 is finitely additive on $\mathcal{A}$ it then follows that

$$\begin{aligned}\mu^*(S) + \varepsilon &\geq \sum_{n=1}^\infty (\mu_0(A_n \cap A) + \mu_0(A_n \cap A^c)) \\ &\geq \mu^*(S \cap A) + \mu^*(S \cap A^*) \geq \mu^*(S).\end{aligned}$$

As this was for $\varepsilon > 0$ arbitrary we then have that A is μ^*-measurable.

Suppose now that $\mathcal{M}$ is the σ-algebra generated by $\mathcal{A}$, and let ν be another measure on $(X, \mathcal{M})$ so that $\nu(A) = \mu_0(A)$ for all $A \in \mathcal{A}$. Then for $E \in \mathcal{M}$, if $E \subset \cup_{n=1}^\infty A_n$ with $A_n \in \mathcal{A}$ we have

$$\nu(E) \leq \sum_{n=1}^\infty \nu(A_n) = \sum_{n=1}^\infty \mu_0(A_n),$$

and it follows that $\nu(E) \leq \mu^*(E)$.

If we have $E \in \mathcal{M}$ such that $\mu^*(E) < \infty$, and if $\varepsilon > 0$, then there exist $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ so that $\mu^*(E) + \varepsilon > \mu^*(\cup_{n=1}^\infty A_n)$, and hence setting $A = \cup_{n=1}^\infty A_n$ we have $\mu^*(A \setminus E) < \varepsilon$. Therefore,

$$\begin{aligned}\mu^*(E) \leq \mu^*(A) = \nu(A) &= \nu(E) + \nu(A \setminus E) \\ &\leq \nu(E) + \mu^*(A \setminus E) < \nu(E) + \varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary we then have $\mu^*(E) \leq \nu(E)$ whenever $\mu^*(E) < \infty$. If μ^* gives a semifinite measure on $\mathcal{M}$ then by Proposition 2.2.4 it follows that for all $E \in \mathcal{M}$ we have

$$\begin{aligned}\nu(E) &\geq \sup\{\nu(A) \mid A \subset E, A \in \mathcal{M}, \text{ and } \nu(A) < \infty\} \\ &\geq \sup\{\mu^*(A) \mid A \subset E, A \in \mathcal{M}, \text{ and } \mu^*(A) < \infty\} = \mu^*(E),\end{aligned}$$

and hence in this case we have $\nu(E) = \mu^*(E)$ for all $E \in \mathcal{M}$. ∎

2.2.3 Exercises

A measure space $(X, \mathcal{M}, \mu)$ is **complete** if every subset of a null set is measurable (and hence also null).

Exercise 2.2.9. Suppose $(X, \mathcal{M}, \mu)$ is a measure space and let $\mathcal{N} = \{E \in \mathcal{M} \mid \mu(E)\}$ be the space of null sets. We let $\overline{\mathcal{M}} = \{E \cup F \mid E \in \mathcal{M}$ and $F \subset N$ for some $N \in \mathcal{N}\}$. Then $\overline{\mathcal{M}}$ is a σ-algebra and there is a unique extension $\overline{\mu}$ of μ to a complete measure on $\overline{\mathcal{M}}$.

The measure space $(X, \overline{\mathcal{M}}, \overline{\mu})$ from the previous theorem is called the **completion** of $(X, \mathcal{M}, \mu)$.

Exercise 2.2.10. If μ^* is an outer measure on X, $\mathcal{M}$ is the collection of all μ^*-measurable sets, and μ is the restriction of μ^* to $\mathcal{M}$, then $(X, \mathcal{M}, \mu)$ is a complete measure space.

Let $(X, \mathcal{M}, \mu)$ be a measure space. A function $f \in \mathcal{M}(X)$ is **essentially bounded** if there exists $M \in [0, \infty)$ such that $\mu(\{x \in X \mid |f(x)| > M\}) = 0$. We let $\mathcal{L}^\infty(X, \mu)$ denote the space of all (complex valued) essentially bounded functions, and for $f \in \mathcal{L}^\infty(X, \mu)$ we set

$$\|f\| = \inf\{M \in [0, \infty) \mid \mu(\{x \in X \mid |f(x)| > M\}) = 0\}.$$

For clarity, we sometimes may write $\|f\|_\infty$ instead of $\|f\|$.

Exercise 2.2.11. $\mathcal{L}^\infty(X, \mu)$ is an algebra and $\| \cdot \|_\infty$ gives a seminorm on $\mathcal{L}^\infty(X, \mu)$.

We let $L^\infty(X, \mu)$ be the normed algebra obtained from $\mathcal{L}^\infty(X, \mu)$ by identifying two functions f and g when $\|f - g\|_\infty = 0$, i.e., when $f = g$ almost everywhere.

Exercise 2.2.12. If $\{f_n\}_{n \in \mathbb{N}} \subset \mathcal{L}^\infty(X, \mu)$ is Cauchy with respect to $\| \cdot \|_\infty$, then there exists $f \in \mathcal{L}^\infty(X, \mu)$ such that $\|f - f_n\|_\infty \to 0$, hence $L^\infty(X, \mu)$ is a Banach algebra.

Exercise 2.2.13. Let $(X, \mathcal{M}, \mu)$ be a finite measure space and for $E, F \in \mathcal{M}$ set $\rho(E, F) = \mu(E \Delta F)$. Then ρ gives a semimetric on $\mathcal{M}$.

Exercise 2.2.14. Let $(X, \mathcal{M}, \mu)$ be a finite measure space and ρ defined as above. Then ρ is a complete semimetric. Hint: If $\{E_n\}_{n \in \mathbb{N}}$ is Cauchy, by passing to a subsequence we may suppose $\mu(E_n \Delta E_m) \leq \max\{2^{-n}, 2^{-m}\}$, and in this case setting $F_m = \cup_{k \geq m} E_k$, we have that $\{F_m\}_{m \in \mathbb{N}}$ is again Cauchy, and $\mu(F_m \Delta E_n) < 2^{-n+4}$ for $m > n$.

2.3 Borel measures on $\mathbb{R}$

By a **Borel measure** on a metric space, we mean a measure on the Borel σ-algebra.

Lemma 2.3.1. *Let $F : \mathbb{R} \to \mathbb{R}$ be increasing and right continuous. Let $\mathcal{I}$ be the collection of intervals of the form $(a, b]$, for $-\infty \le a < b < \infty$, or of the form (a, ∞) for $-\infty \le a < \infty$, and set $\mu_0((a, b]) = F(b) - F(a)$ if $b < \infty$, and $\mu_0((a, \infty)) = F(\infty) - F(a)$, where $F(\pm\infty) = \lim_{t\to\pm\infty} F(t)$. We let $\mathcal{A}$ denote the algebra consisting of finite unions of intervals in $\mathcal{I}$. Then μ_0 extends to a premeasure on $\mathcal{A}$.*

Proof. Note that if $(a, b] = \cup_{k=1}^n (a_k, b_k]$, then after rearranging we may assume that $a = a_1 < b_1 = a_2 < b_2 = \cdots < b_{k-1} = a_k < b_k = b$, and we have that

$$\mu_0((a, b]) = F(b) - F(a) = \sum_{k=1}^n F(b_k) - F(a_k) = \sum_{k=1}^n \mu_0((a_k, b_k]).$$

We similarly have that if $I_1, \ldots, I_n \in \mathcal{I}$ are disjoint and $\cup_{k=1}^n I_k = (a, \infty)$, then $\mu_0((a, \infty)) = \sum_{k=1}^n \mu_0(I_k)$. From this it then follows easily that we obtain a well defined finitely additive set function $\mu_0 : \mathcal{A} \to [0, \infty]$ by setting $\mu_0(\cup_{k=1}^n I_k) = \sum_{k=1}^n \mu_0(I_k)$ for pairwise disjoint sets $I_1, \ldots, I_n \in \mathcal{I}$.

We will now show that μ_0 is a premeasure on $\mathcal{A}$. Suppose that $\{I_j\}_{j=1}^\infty$ is a pairwise disjoint sequence of intervals in $\mathcal{A}$, such that $\cup_{j=1}^\infty I_j = I \in \mathcal{I}$. Then we have

$$\mu_0(I) = \mu_0(\cup_{j=1}^n I_j) + \mu_0(I \setminus \cup_{j=1}^n I_j) \ge \mu_0(\cup_{j=1}^n I_j) = \sum_{j=1}^n \mu_0(I_j).$$

Taking a limit as $n \to \infty$ we see that $\mu_0(I) \ge \sum_{n=1}^\infty \mu_0(I_j)$.

For the reverse inequality we first assume that $I = (a, b]$, where a and b are finite. Fix $\varepsilon > 0$. As F is right continuous there exists $\delta > 0$ so that $F(a + \delta) - F(a) < \varepsilon$. Similarly, if $I_j = (a_j, b_j]$ then there exist $\delta_j > 0$ so that $F(b_j + \delta_j) - F(b_j) < \varepsilon 2^{-j}$. Since the open intervals $(a_j, b_j + \delta_j)$ cover the compact set $[a + \delta, b]$ there exists $n \in \mathbb{N}$ and $j_1, \ldots, j_n$, so that $[a + \delta, b] \subset \cup_{i=1}^n (a_{j_1}, b_{j_i} + \delta_{j_i})$. We may further assume that no subcollection also covers $[a + \delta, b]$ and by reordering and reindexing $j_1, \ldots, j_n$ as $1, \ldots, n$ we may then assume that

$$a_1 < a + \delta \le a_2 < b_1 + \delta_1 \le a_3 < \cdots \le a_n < b_{n-1} + \delta_{n-1} \le b < b_n + \delta_n.$$

We then have

$$\begin{aligned}
\mu_0(I) &< F(b) - F(a + \delta) + \varepsilon \\
&\le F(b_n + \delta_n) - F(a_1) + \varepsilon \\
&= F(b_n + \delta_n) - F(a_n) + \sum_{j=1}^{n-1} (F(a_{j+1}) - F(a_j)) + \varepsilon \\
&\le F(b_n + \delta_n) - F(a_n) + \sum_{j=1}^{n-1} (F(b_j + \delta_j) - F(a_j)) + \varepsilon \\
&\le \sum_{j=1}^n F(b_n) - F(a_n) + 2\varepsilon \le \sum_{j=1}^\infty \mu_0(I_j) + 2\varepsilon.
\end{aligned}$$

As $\varepsilon > 0$ was arbitrary it then follows that $\mu_0(I) \leq \sum_{j=1}^{\infty} \mu_0(I_j)$.

For the case of a general interval $I \in \mathcal{I}$, we can easily check that have $\mu_0(I) = \lim_{a\to-\infty,b\to\infty} \mu_0(I\cap[a,b))$ and $\sum_{j=1}^{\infty} \mu_0(I_j) = \lim a \to -\infty, b \to \infty \sum_{j=1}^{\infty} \mu_0(I_j\cap [a,b))$. Using the case $-\infty < a < b < \infty$ above and taking limits then shows that $\mu_0(I) = \sum_{j=1}^{\infty} \mu_0(I_j)$.

If we now consider general sets $E, E_j \in \mathcal{A}$, such that $\{E_j\}_{j=1}^{\infty}$ is pairwise disjoint and $E = \cup_{j=1}^{\infty} E_j$, then writing each set as a finite union of disjoint intervals, and using finite additivity of μ_0 it then follows that $\mu_0(E) = \sum_{j=1}^{\infty} \mu_0(E_j)$. Hence, μ_0 is a premeasure on $\mathcal{A}$. ■

Theorem 2.3.2. *Let $F : \mathbb{R} \to \mathbb{R}$ be increasing and right continuous. Then there is a unique Borel measure μ_F on $\mathbb{R}$ such that $\mu_F((a,b]) = F(b) - F(a)$ for all $a, b \in \mathbb{R}$.*

Conversely, if μ is a Borel measure on $\mathbb{R}$ which is finite on all bounded Borel sets and we define

$$F(x) = \begin{cases} \mu((0,x]) & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -\mu((-x,0]) & \text{if } x < 0, \end{cases}$$

then F is increasing, right continuous, and $\mu = \mu_F$.

Proof. We let $\mathcal{A}$ denote the algebra in Lemma 2.3.1, and note that the σ-algebra generated by $\mathcal{A}$ is the Borel σ-algebra. By Lemma 2.3.1 there exists a premeasure μ_0 on $\mathcal{A}$ so that $\mu_0((a,b]) = F(b) - F(a)$ for $a, b \in \mathbb{R}$. By Carathéodory's extension theorem there then exists a Borel measure μ_f on $\mathbb{R}$ such that $\mu_F((a,b]) = F(b) - F(a)$. Moreover, since μ_F is σ-finite it then also follows from Carathéodory's extension theorem that μ_F is the unique Borel measure with this property.

If μ is a Borel measure on $\mathbb{R}$ which is finite on all bounded Borel sets and if we define F as above, then it follows from monotonicity that F is increasing and from continuity from above/below that F is right continuous. Moreover, we see easily that for $a, b \in \mathbb{R}$ we have $\mu((a,b]) = F(b) - F(a)$. By uniqueness of the measure μ_F it then follows that $\mu = \mu_F$. ■

Given $F : \mathbb{R} \to \mathbb{R}$ increasing and right continuous, the completion of the corresponding measure μ_F is called the **Lebesgue-Stieltjes measure** associated to F, and F is called a **distribution function** associated to μ_F. It's easy to check that two distribution functions associated to the same measure must differ by a constant.

2.3.1 Lebesgue measure on $\mathbb{R}$

The Lebesgue-Stieltjes measure corresponding to the function $F(x) = x$ is called **Lebesgue measure** on $\mathbb{R}$ and usually denoted by λ. A set $E \subset \mathbb{R}$ is **Lebesgue measurable** if it is λ^*-measurable where λ^* is the outer measure corresopnding to λ. Note that by Exercise 2.2.10 if $E \subset \mathbb{R}$ satisfies $\lambda^*(E) = 0$, then E is Lebesgue measurable.

Theorem 2.3.3. *If $E \subset \mathbb{R}$ is Borel, then so is $E + s$ and rE for all $s, r \in \mathbb{R}$. Moreover, we have $\lambda(E + s) = \lambda(E)$ and $\lambda(rE) = |r|\lambda(E)$.*

Proof. Since addition and multiplication are continuous it follows from Corollary 2.1.3 that $E+s$ and rE are Borel. By uniqueness of the Lebesgue-Stieltjes measure to show that $\lambda(E+s) = \lambda(E)$ and $\lambda(rE) = |r|\lambda(E)$, it suffices to show these equalities when E is a half open interval, in which case this is obvious. ■

Note that every point $x \in \mathbb{R}$ has Lebesgue measure zero. It follows that every countable set has Lebesgue measure zero. There are also uncountable sets with Lebesgue measure zero. The **Cantor set** C is the set of all $x \in [0,1]$ that have a base-3 expansion $x = \sum_{n=1}^{\infty} a_n 3^{-n}$ with $a_n \neq 1$ for all n (note that such an expansion, if it exists, must be unique). We may may obtain C by starting with the unit interval $[0,1]$ and removing the open middle third $(\frac{1}{3}, \frac{2}{3})$, then removing the open middle thirds $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$ of the two remaining interval, etc.

Proposition 2.3.4. *The Cantor set C is compact, contains no non-trivial open interval, and has no isolated points. Moreover, C has cardinality $|\mathbb{R}|$ and satisfies $\lambda(C) = 0$.*

Proof. C is obtained by removing open intervals, thus C is a decreasing union of closed subsets of $[0,1]$, and so C itself is a closed subset of $[0,1]$ which must then be compact. If $x = \sum_{n=1}^{\infty} a_n 3^{-n} \in C$ with $a_n \in \{0,2\}$ for all n, then for each $n \in \mathbb{N}$ consider $x_n \in C$ which has the same expansion as x except for the nth coefficient, which is either 0 if $a_n = 2$, or 2 if $a_n = 0$. Then $\{x_n\}_{n\in\mathbb{N}} \subset C$ is an infinite sequence such that $x_n \to x$. Hence, C has no isolated points.

By considering the lengths of the intervals removed from C we have

$$\lambda(C) = 1 - \sum_{n=1}^{\infty} \frac{2^{n-1}}{3^n} = 0.$$

If $x = \sum_{n=1}^{\infty} a_n 3^{-n} \in C$ with $a_n \neq 1$ for all n, then set $f(x) = \sum_{n=1}^{\infty} b_n 2^{-n}$ where $b_n = a_n/2$. Then the series describing $f(x)$ is a base-2 expansion and every number in $[0,1]$ can be expressed in this way, thus $f : C \to [0,1]$ is a surjection which shows that $|C| = |\mathbb{R}|$. ■

Corollary 2.3.5. *Let $\mathcal{L} \subset 2^{\mathbb{R}}$ denote the σ-algebra of Lebesgue measurable subsets, then $|\mathcal{L}| = |2^{\mathbb{R}}|$. Hence, $\mathcal{B}(\mathbb{R}) \subsetneq \mathcal{L} \subsetneq 2^{\mathbb{R}}$.*

Proof. We clearly have $|\mathcal{L}| \leq |2^{\mathbb{R}}|$, and if C is the Cantor set then $\lambda(C) = 0$, hence any subset of C is Lebesgue measurable and so we have $|2^{\mathbb{R}}| = |2^C| \leq |\mathcal{L}|$.

By Exercise 2.1.14 we have $|\mathcal{B}(\mathcal{R})| = |\mathbb{R}|$, hence $\mathcal{B}(\mathbb{R}) \subsetneq \mathcal{L}$. Also, Vitali's set E constucted at the beginning of the chapter cannot be Lebesgue measurable hence $\mathcal{L} \subsetneq 2^{\mathbb{R}}$. ■

Note that the function $f : C \to [0,1]$ in the proof of Proposition 2.3.4 is monotone increasing. Moreover, if $x, y \in C$, with $x < y$, then $f(x) = f(y)$

only if x and y are the endpoints of one of the intervals removed from $[0,1]$. In this case we have $f(x) = f(y) = m2^{-n}$ where m, n are integers. Thus we may extend f on the interval (x, y) by letting it be constant $m2^{-n}$. In this way we extend f to a monotone increasing function $\tilde{f} : [0,1] \to [0,1]$. The function $\tilde{f}$ is called the **Cantor function**.

Theorem 2.3.6. *Let $f : [0,1] \to [0,1]$ be the Cantor function. Then the following hold:*

1. *f is continuous.*
2. *The derivative of f exists, and equals zero, almost everywhere.*
3. *There exists a Lebesgue measurable set $E \subset [0,1]$ such that $f(E)$ is not Lebesgue measurable.*

Proof. Note that f is surjective and hence cannot have any jump discontinuities. Since f is monotone increasing it must then be continuous.

f is constant on each middle third, and hence the derivative of f exists and equals zero, on each middle third, and as we say above, the union of these open intervals is conull.

If we consider Vitali's example of a non-measurable set $E \subset [0,1]$, then setting $E_0 = f^{-1}(E) \cap C$ we have that E_0 is contained in a measure zero set and hence must be measurable. Since $f(C) = [0,1]$ we have $f(E_0) = E$. ∎

2.3.2 Regularity of Borel measures

Theorem 2.3.7. *Suppose μ is a finite Borel measure on a metric space (X, d). Then μ is* **regular***: For $E \subset X$ Borel we have*

$$\begin{aligned}\mu(E) &= \inf\{\mu(G) \mid E \subset G \text{ and } G \text{ is open}\}\\ &= \sup\{\mu(F) \mid F \subset E \text{ and } F \text{ is closed}\}.\end{aligned}$$

Proof. We let Σ denote the family of Borel sets E which satisfy the conclusion of the theorem. If $E \subset X$ is closed then $G_n = \{x \in X \mid d(x, E) < 1/n\}$ is open for each $n \in \mathbb{N}$ and we have $\cap_{n=1}^{\infty} G_n = E$. By continuity from above of measures we have that $\mu(E) = \lim_{n\to\infty} \mu(G_n)$. Hence it follows that $E \in \Sigma$. It is also clear that Σ is closed under taking complements.

If $\{E_n\}_{n=1}^{\infty} \subset \Sigma$ and $\varepsilon > 0$, then there exist $F_n, G_n \subset X$ with F_n closed and G_n open such that $F_n \subset E_n \subset G_n$ and $\mu(G_n \setminus F_n) < \varepsilon 2^{-n}$. If we set $G = \cup_{n=1}^{\infty}$ and $F = \cup_{n=1}^{\infty} F_n$ then we have $F \subset \cup_{n=1}^{\infty} E_n \subset G$, and $G \setminus F \subset \cup_{n=1}^{\infty}(G_n \setminus F_n)$ hence $\mu(G \setminus F) \leq \sum_{n=1}^{\infty} \mu(G_n \setminus F_n) < \varepsilon$. By continuity from above we have $\lim_{n\to\infty} \mu(G \setminus (\cup_{k=1}^{n} F_k)) = \mu(G \setminus F) < \varepsilon$. Hence, for n large enough we have $\mu(G \setminus (\cup_{k=1}^{n} F_k)) < \varepsilon$. Since G is open and $\cup_{k=1}^{n} F_k \subset F \subset E$ is closed it then follows that $\cup_{n=1}^{\infty} E_n \in \Sigma$. Thus, Σ is a σ-algebra which contains the closed sets and hence must contain all Borel sets. ∎

A set $E \subset X$ is a **G_δ-set** if it a countable intersection of open sets. A set $E \subset X$ is an **F_σ-set** if it is a countable union of closed sets.

Corollary 2.3.8. *Suppose μ is a σ-finite Borel measure on a metric space (X,d). Then for every Borel set $E \subset X$ there exists an F_σ-set F, and a G_δ-set G such that $F \subset E \subset G$ and $\mu(G \setminus F) = 0$.*

Proof. This follows easily from Theorem 2.3.7 when μ is finite. If μ is σ-finite we may write X as a disjoint union $X = \cup_{n=1}^{n} E_n$ where $E_n \subset X$ is Borel and $\mu(E_n) < \infty$. Suppose $E \subset X$ is Borel and $\varepsilon > 0$. For each $n \geq 1$ consider the Borel measure μ_n on X given by $\mu_n(A) = \mu(A \cap E_n)$, then μ_n is a finite measure and so there exist F_σ-sets $F_n^1 \subset E \cap E_n$, $F_n^2 \subset E^c \cap E_n$ so that $\mu((E \cap E_n) \setminus F_n^1) = \mu((E^c \cap E_n) \setminus F_n^2) = 0$.

If we set $F^1 = \cup_{n=1}^\infty F_n^1 \subset E$ and $F^2 = \cup_{n=1}^\infty F_n^2 \subset E^c$, then we have $\mu(E \setminus F^1) = \mu(E^c \setminus F^2) = 0$. Then $G = (F^2)^c$ is G_δ and satisfies $E \subset G$, and $\mu(G \setminus E) = \mu(E^c \setminus F^2) = 0$. Hence, $\mu(G \setminus F^1) = \mu(G \setminus E) + \mu(E \setminus F^1) = 0$. ∎

Corollary 2.3.9. *Suppose μ_F is a Lebesgue-Stieltjes measure on $\mathbb{R}$, and $E \subset \mathbb{R}$ is a Borel set such that $\mu_F(E) < \infty$. Then for every $\varepsilon > 0$, there exists $G \subset \mathbb{R}$ such that G is a finite union of intervals and $\mu_F(E \Delta G) < \varepsilon$.*

Proof. Suppose $E \subset \mathbb{R}$ is Borel and $\varepsilon > 0$. Take $t > 0$ so that $\mu(E \setminus (-t,t)) < \varepsilon/3$. By the previous Corollary there exists a sequence of open set G_n such that $\mu(E \Delta \cap_{n=1}^\infty G_n) = 0$. Thus, setting $G_n' = G_n \cap (-t,t)$ we have $\lim_{k\to\infty} \mu((\cap_{n=1}^k G_n') \setminus E) = 0$, and hence for some k we have $\mu((\cap_{n=1}^k G_n') \setminus E) < \varepsilon/3$. Since $\cap_{n=1}^k G_n'$ is open it is a countable union of intervals, hence there exists a set G which is a finite union of intervals such that $\mu((\cap_{n=1}^k G_n') \setminus G) < \varepsilon/3$. We then have

$$\mu(E \Delta G) \leq \mu(E \setminus (-t,t)) + \mu((\cap_{n=1}^k G_n') \setminus E) + \mu((\cap_{n=1}^k G_n') \setminus G) < \varepsilon.$$

∎

Theorem 2.3.10. *Suppose μ is a finite Borel measure on a complete separable metric space (X,d). Then μ is* **tight***: For $E \subset X$ Borel we have*

$$\mu(E) = \sup\{\mu(K) \mid K \subset A \text{ and } K \text{ is compact}\}.$$

Proof. Since μ is regular it is enough to consider the case when E is closed, and then restricting to E we might as well assume that $E = X$. So it suffices to show $\mu(X) = \sup\{\mu(K) \mid K \text{ is compact}\}$.

Fix $\varepsilon > 0$. Let $\{x_i\}_{i=1}^\infty$ be a countable dense set in X. If $n \geq 1$ then $\cup_{i=1}^\infty \overline{B}(1/n, x_i) = X$ and hence $\lim_{k\to\infty} \mu(\cup_{i=1}^k \overline{B}(1/n, x_i)) = \mu(X)$. We take k_n so that $\mu(X \setminus \cup_{i=1}^{k_n} \overline{B}(1/n, x_i)) < \varepsilon 2^{-n}$. Let $K = \cap_{n=1}^\infty \cup_{i=1}^{k_n} \overline{B}(1/n, x_i)$. Then K is closed and totally bounded, and hence compact. We also have $\mu(X \setminus K) \leq \sum_{n=1}^\infty \mu(X \setminus \cup_{i=1}^{k_n} \overline{B}(1/n, x_i)) < \varepsilon$. ∎

Theorem 2.3.11 (Lusin's Theorem)**.** *Suppose μ is a finite Borel measure on a metric space (X,d), and $f \in \mathcal{M}(X)$. For each $\varepsilon > 0$ there exists a closed set $F \subset X$ such that $\mu(F^c) < \varepsilon$ and $f_{|F}$ is continuous.*

Proof. Fix $\varepsilon > 0$, and take an enumeration of the rationals $\mathbb{Q} = \{q_n\}_{n=1}^\infty$. Then $E_{n,k} = f^{-1}((q_n, q_k))$ is measurable and hence there exist $F_{n,k}$ closed and $V_{n,k}$ opne so that $F_{n,k} \subset E_{n,k} \subset V_{n,k}$ with $\mu(V_{n,k} \setminus F_{n,k}) < \varepsilon 2^{-n-k}$. Set $U = \cup_{n,k}(V_{n,k} \setminus F_{n,k})$ and $F = U^c$. Then $\mu(U) < \varepsilon$, and F is closed. Moreover, $f^{-1}((q_n, q_k)) \cap F = V_{n,k} \cap F$. Since every open set is a union of sets of the form (q_n, q_k) it then follows easily that $f_{|F}$ is continuous. ■

Corollary 2.3.12. *Suppose μ_F is a Lebesgue-Stieltjes measure on $\mathbb{R}$, and $f \in \mathcal{M}(\mathbb{R})$ is such that f vanishes outside a finite measure set. Then for all $\varepsilon > 0$ there exists a continuous function $g \in C_0(\mathbb{R})$ so that*

$$\mu_F(\{x \in \mathbb{R} \mid f(x) \neq g(x)\}) < \varepsilon.$$

Proof. Fix $\varepsilon > 0$ and take $t_0 > 0$ so that $\mu_F(((-\infty, -t_0) \cup (t_0, \infty)) \cap \{x \in \mathbb{R} \mid f(x) \neq 0\}) < \varepsilon/4$. By considering the restriction of μ_F to $[-t_0, t_0]$ the previous theorem gives a closed set $E \subset [-t_0, t_0]$ so that $f_{|E}$ is continuous and $\mu_F([-t_0, t_0] \setminus E) < \varepsilon/4$.

We let $a = \inf E$ and $b = \sup E$, and take $a' < a$, and $b' > b$ so that $\mu_F([a', a)) + \mu_F((b, b']) < \varepsilon/2$. If $t \in E^c$, $a < t < b$ we let (t_1, t_2) denote the largest interval in E^c which contains t. We define g so that

$$g(t) = \begin{cases} f(t) & \text{if } t \in E, \\ 0 & \text{if } t < a', \text{ or } t > b', \\ \frac{t-a'}{a-a'} f(a) & \text{if } \in [a-1, a), \\ \frac{b'-t}{b'-b} f(b) & \text{if } t \in (b, b+1], \\ \frac{t-t_1}{t_2-t_1} f(t_2) + \frac{t_2-t}{t_2-t_1} f(t_1) & \text{if } t \in (t_1, t_2). \end{cases}$$

We then have that $g \in C_0(\mathbb{R})$ and g agrees with f on E, hence it follows easily that $\mu_F(\{x \in \mathbb{R} \mid f(x) = g(x)\}) < \varepsilon$. ■

2.3.3 Exercises

Exercise 2.3.13. For every $\varepsilon > 0$, there exists a compact set $K \subset [0, 1]$ which contains no isolated points and no non-trival open interval such that $\lambda(K) > 1 - \varepsilon$.

Exercise 2.3.14. Let $E \subset \mathbb{R}$ be a Borel set such that $\lambda(E) < \infty$. Then the maps $\mathbb{R} \ni t \mapsto \lambda(E \Delta tE)$, and $t \mapsto \lambda(E \Delta (E + t))$ are continuous.

Exercise 2.3.15. If $(X, \mathcal{M}, \mu)$ is a measure space and $\{A_j\}_{j=1}^\infty \subset \mathcal{M}$, we set $\liminf_{j\to\infty} A_j = \cup_{N=1}^\infty \cap_{k=N}^\infty A_k$. Then $\mu(\liminf A_j) \leq \liminf \mu(A_j)$.

Exercise 2.3.16. There exists a measurable function $f : [0, 1] \to \mathbb{R}$ so that for all $(a, b) \subset [0, 1]$ and $(c, d) \subset \mathbb{R}$ we have

$$\lambda(\{x \in (a, b) \mid f(x) \in (c, d)\}) > 0.$$

Exercise 2.3.17. Show that there exists a Borel set $A \subset [0, 1]$ such that $0 < m(A \cap I) < m(I)$ for every subinterval I of $[0, 1]$.

2.4 Integration

2.4.1 Integrable functions

Let $(X, \mathcal{M}, \mu)$ be a measure space. If $E \in \mathcal{M}$ has finite measure, and $f = \sum_{n=1}^k \alpha_n 1_{E_n}$ is a simple function with respect to some measurable partition $\{E_n\}_{n=1}^k$ of E, then we define the **integral** of f to be

$$\int f = \sum_{n=1}^k \alpha_n \mu(E_n).$$

Note that if we have another representation $f = \sum_{m=1}^l \beta_m 1_{F_m}$ then by writing $F_m = \cup_{n=1}^k F_m \cap E_n$ we see that $\sum_{n=1}^k \alpha_n \mu(F_m \cap E_n) = \beta_m \mu(F_m)$. Summing over m then shows that the integral is well defined. Depending on the situation we also use the following notation for the integral

$$\int f\, d\mu; \quad \int_X f\, d\mu; \quad \int_X f(x)\, d\mu(x); \quad \int_X f(x)\, dx$$

If $A \subset X$ is measurable we write $\int_A f$ for the integral $\int_X 1_A f$.

We let $\mathcal{L}_0^1(X)$ denote the set of all simple functions having a decomposition $f = \sum_{n=1}^k \alpha_n 1_{E_n}$ where the $\{E_n\}_{n=1}^k$ gives a partition of a finite measure set E. If $I \subset \mathbb{C}$ we denote by $\mathcal{L}_0^1(X; I)$ those functions in $\mathcal{L}_0^1(X)$ which take values in I, and we also set $\mathcal{L}_0^1(X)_+ = \mathcal{L}_0^1(X; [0\infty))$. We note that $\mathcal{L}_0^1(X)$ is a vector space over $\mathbb{C}$. We also note that the integral defines a linear functional on $\mathcal{L}_0^1(X)$.

If $f \in \mathcal{L}_0^1(X)_+$ then we clearly have $\int f \geq 0$. Linearity then shows that for $f, g \in \mathcal{L}_0^1(X; \mathbb{R})$ we have

$$\int f \leq \int g, \qquad \text{if } f \leq g. \tag{2.2}$$

Also, note that for $f \in \mathcal{L}_0^1(X)$ the triangle inequality in $\mathbb{C}$ shows that

$$\left| \int f \right| \leq \int |f|. \tag{2.3}$$

Similarly, if we have $f, g \in \mathcal{L}_0^1(X)$, then taking a partition of a set of finite measure such that both f and g are simple functions with respect to this partition, the triangle inequality in $\mathbb{C}$ shows that

$$\int |f + g| \leq \int |f| + \int |g|. \tag{2.4}$$

Therefore, we may define the L^1**-seminorm** on $\mathcal{L}_0^1(X)$ as

$$\|f\|_1 = \int |f|. \tag{2.5}$$

Finally, note that $\|f\|_1 = 0$ if and only if $f = 0$ almost everywhere.

We let $L^1(X)$ be the Banach space completion of $\mathcal{L}_0^1(X)$ after identifying functions which agree almost everywhere. Depending on the situation we may use the terminology

$$L^1(X); \quad L^1(\mu); \quad L^1(X,\mu); \quad L^1(X,\mathcal{M},\mu).$$

Note that since $|\int f| \leq \int |f|$ it follows that the integral is continuous with respect to the seminorm $\|\cdot\|_1$. Hence, the integral extends continuously to $L^1(X)$, and we use the same terminology here.

Every vector in $L^1(X)$ is the limit of a Cauchy sequence of functions in $\mathcal{L}_0^1(X)$. We now wish to find a more tractable realization of vectors in $L^1(X)$.

Lemma 2.4.1. *Let $\{f_n\}_{n\in\mathbb{N}} \subset \mathcal{L}_0^1(X)$ be a Cauchy sequence, then there exists a subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ which converges almost everywhere to a measurable function f, and such that for all $\varepsilon > 0$ there exists a measurable set $A \subset X$ with $\mu(A) < \varepsilon$, such that $\{f_{n_k}\}_{k\in\mathbb{N}}$ converges to f uniformly on A^c.*

Proof. Suppose $\{f_n\}_{n\in\mathbb{N}} \subset \mathcal{L}_0^1(X)$ is Cauchy. Passing to a subsequence we may assume that $\|f_n - f_{n+1}\|_1 \leq 4^{-n}$, for all $n \in \mathbb{N}$. Set

$$E_n = \{x \in X \mid |f_n(x) - f_{n+1}(x)| \geq 2^{-n}\}.$$

Then we have $|f_n - f_{n+1}| \geq 2^{-n} 1_{E_n}$, and hence it follows that

$$4^{-n} \geq \int |f_n - f_{n+1}| \geq \int 2^{-n} 1_{E_n} = 2^{-n}\mu(E_n).$$

For $N \in \mathbb{N}$ set $A_N = \cup_{n\geq N} E_n$, and set $A = \cap_{N\in\mathbb{N}} A_N$. Then

$$\mu(A_N) \leq \sum_{n\geq N} \mu(E_n) \leq \sum_{n\geq N} 2^{-n} < 2^{-N+1}.$$

Hence $\mu(A) = 0$.

If $x \in A_N^c$ and $n \geq N$ then $|f_n(x) - f_n(x+1)| < 2^{-n}$. It therefore follows that $\{f_n(x)\}_{n\in\mathbb{N}}$ is Cauchy and hence converges to some $f(x) \in \mathbb{C}$. Therefore there exists a function $f : A^c \to \mathbb{C}$ so that $f_n(x) \to f(x)$ for all $x \in A^c$. Note that we also have that f_n converges uniformly to f on each set A_N, and $\lim_{N\to\infty} \mu(A_N) = 0$. From Proposition 2.1.7 we see that $f : A^c \to \mathbb{C}$ is measurable, and we may extend f to a measurable function on X by setting $f(x) = 0$ for all $x \in A$. ■

Lemma 2.4.2. *Suppose $\{f_n\}_{n\in\mathbb{N}}, \{g_n\}_{n\in\mathbb{N}} \subset \mathcal{L}_0^1(X)$ are Cauchy sequences, and $f \in \mathcal{M}(X)$ such that both sequences converge almost everywhere to f. Then $\lim_{n\to\infty} \|f_n - g_n\|_1 = 0$.*

Proof. If we consider $h_n = f_n - g_n$, then $\{h_n\}_{n=1}^\infty$ is Cauchy in $\mathcal{L}_0^1(X)$ and satisfies $h_n \to 0$ almost everywhere. We must show that $\|h_n\|_1 \to 0$. Fix $\varepsilon > 0$, and take $N \in \mathbb{N}$ so that $\|h_n - h_m\|_1 < \varepsilon$ for all $n, m \geq N$. Using Lemma 2.4.1 and passing to a subsequence we may assume that there exists a measurable set

$A \subset X$ with $\mu(A) < \frac{\varepsilon}{1+\|h_N\|_\infty}$, such that $h_n \to 0$ uniformly on A^c. Let $E \subset X$ be a set of finite measure such that f_N vanishes outside of E. Then for n large we have

$$\begin{aligned}\int_{A\cup E^c} |h_n| &\leq \int_{A\cup E^c} |h_N| + \int_{A\cup E^c} |h_n - h_N| \\ &\leq \int_A |h_N| + \|h_n - h_N\|_1 \\ &\leq \mu(A)\|h_N\|_\infty + \|h_n - h_N\|_1 < 2\varepsilon.\end{aligned}$$

Hence,

$$\begin{aligned}\limsup_{n\to\infty} \|h_n\|_1 &= \limsup_{n\to\infty} \left(\int_{A\cup E^c} |h_n| + \int_{A^c\cap E} |h_n| \right) \\ &\leq 2\varepsilon + \mu(E) \limsup_{n\to\infty} \|h_{n|A^c}\|_\infty = 2\varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary we have that $\|h_n\|_1 \to 0$. ∎

We let $\mathcal{L}^1(X)$ denote the set of all measurable functions $f \in \mathcal{M}(X)$ such that f is the almost everywhere limit of a Cauchy sequence of functions in $\mathcal{L}_0^1(X)$. Functions in $\mathcal{L}^1(X)$ are said to be **integrable**. Clearly, $\mathcal{L}^1(X)$ is a vector space. Lemma 2.4.2 shows that we have a well defined map $\Xi : \mathcal{L}^1(X) \to L^1(X)$ which assigns to each function $f \in \mathcal{L}^1(X)$ a Cauchy sequence which converges almost everywhere to f. The map Ξ is clearly linear and the kernel consists of all measurable functions which are zero almost everywhere. We extend the integral to a linear map on $\mathcal{L}^1(X)$ by setting $\int f = \int \Xi(f)$. In other words, given a function $f \in \mathcal{L}^1(X)$ we have $\int f = \lim_{n\to\infty} \int f_n$ where $\{f_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{L}_0^1(X)$ which converges almost everywhere to f.

Lemma 2.4.1 shows that the map Ξ is surjective. Thus, we may think of the space $L^1(X)$ as the space of integrable functions where we identify functions which agree almost everywhere.

2.4.2 Properties of integration

Lemma 2.4.3. *If $\{f_n\}_{n\in\mathbb{N}} \subset \mathcal{L}_0^1(X)$ is Cauchy, and $f \in \mathcal{L}^1(X)$ such that $f_n \to f$ almost everywhere, then $\{|f_n|\}_{n\in\mathbb{N}}$ is also Cauchy, and $|f_n| \to |f|$ almost everywhere.*

Proof. This follows easily from the inequality $||a| - |b|| \leq |a - b|$ for all $a, b \in \mathbb{C}$. ∎

Theorem 2.4.4. *If $f \in \mathcal{L}^1(X)$ then $|f| \in \mathcal{L}^1(X)$. Moreover, the inequalities (2.2), (2.3), (2.4), and (2.5) hold for all functions in $\mathcal{L}^1(X)$.*

Proof. The fact that $|f| \in \mathcal{L}^1(X)$ if $f \in \mathcal{L}^1(X)$ follows from the previous lemma, which also shows that the function $f \mapsto |f|$ is uniformly continuous on bounded subsets of $\mathcal{L}_0^1(X)$. Since $f \mapsto \int f$ and $f \mapsto \|f\|_1$ are also uniformly continuous on bounded subsets of $\mathcal{L}_0^1(X)$, it follows that these maps are also continuous on the completion $\mathcal{L}^1(X)$. The inequalities (2.2), (2.3), (2.4), and (2.5) then follow by continuity, since they hold on the dense subspace $\mathcal{L}_0^1(X)$. ■

Recall that if we let $\mathcal{L}^\infty(X)$ denote the space of essentially bounded functions, then we have a complete semi-norm on $\mathcal{L}^\infty(X)$ given by $\|f\|_\infty = \inf\{M \in [0, \infty) \mid \mu(\{x \in X \mid |f(x)| > M\}) = 0\}$. An essentially bounded function f satisfies $\|f\|_\infty = 0$ if and only if f is zero almost everywhere. Thus, if we identify functions which agree almost everywhere then we obtain a Banach space $L^\infty(X)$. This is also a Banach algebra since $\|fg\|_\infty \leq \|f\|_\infty \|g\|_\infty$. We let $\mathcal{L}_0^\infty(X)$ denote the space of simple functions. Since any function $f \in \mathcal{L}^\infty(X)$ agrees almost everywhere with a bounded function it follows from Proposition 2.1.8 that $\mathcal{L}_0^\infty(X)$ is dense in $\mathcal{L}^\infty(X)$.

Theorem 2.4.5. *Suppose $f \in \mathcal{L}^1(X)$, and $g \in \mathcal{L}^\infty(X)$, then $gf \in \mathcal{L}^1(X)$ and $\|gf\|_1 \leq \|g\|_\infty \|f\|_1$.*

Proof. Suppose first that $f \in \mathcal{L}_0^1(X)$ and $g \in \mathcal{L}_0^\infty(X)$, say $f = \sum_{n=1}^N \alpha_n 1_{E_n}$ with $\mu(E_n) < \infty$, and $g = \sum_{m=1}^M \beta_m 1_{F_m}$. We may assume that $\{E_n\}_{n=1}^N$ and $\{F_m\}_{m=1}^M$ are each pariwise disjoint. We then have

$$gf = \sum_{n=1}^N \sum_{m=1}^M \beta_m \alpha_n 1_{F_m \cap E_n} \in \mathcal{L}_0^1(X),$$

and

$$\begin{aligned}
\|gf\|_1 &= \sum_{n=1}^N \sum_{m=1}^M |\beta_m \alpha_n| \mu(F_m \cap E_n) \\
&\leq \sum_{n=1}^N \sum_{m=1}^M \|g\|_\infty |\alpha_n| \mu(F_m \cap E_n) \\
&\leq \sum_{n=1}^N \|g\|_\infty |\alpha_n| \mu(E_n) = \|g\|_\infty \|f\|_1.
\end{aligned}$$

For the general case, suppose that $f \in \mathcal{L}^1(X)$ and $g \in \mathcal{L}^\infty(X)$. Take $f_n \in \mathcal{L}_0^1(X)$ so that $\{f_n\}_{n=1}^\infty$ is Cauchy in $\mathcal{L}_0^1(X)$ and $f_n \to f$ almost everywhere. Also, take $g_n \in \mathcal{L}_0^\infty(X)$ so that $\|g_n - g\|_\infty \to 0$. Then $g_n f_n \to gf$ almost everywhere, and from the triangle inequality and the argument above we have

$$\|g_n f_n - g_m f_m\|_1 \leq \|g_n\|_\infty \|f_n - f_m\|_1 + \|g_n - g_m\|_\infty \|f_m\|_1.$$

Therefore $\{g_n f_n\}_{n=1}^\infty$ is Cauchy in $\mathcal{L}_0^1(X)$. We then have that $gf \in \mathcal{L}^1(X)$ and

$$\|gf\|_1 = \lim_{n\to\infty} \|g_n f_n\|_1 \leq \lim_{n\to\infty} \|g_n\|_\infty \|f_n\|_1 = \|g\|_\infty \|f\|_1.$$

■

Corollary 2.4.6. *If $f \in \mathcal{L}^1(X)$, and $h \in \mathcal{M}(X)$ such that $|h| \leq |f|$, then $h \in \mathcal{L}^1(X)$. In particular, $f \in \mathcal{L}^1(X)$ if and only if $|f| \in \mathcal{L}^1(X)$.*

Proof. We let $g(x) = 0$ if $f(x) = 0$, and $g(x) = h(x)/f(x)$ otherwise. Then g is measurable and $\|g\|_\infty \leq 1$. Therefore $h = gf \in \mathcal{L}^1(X)$. ∎

Corollary 2.4.7. *If $\mu(X) < \infty$ then $\mathcal{L}^\infty(X) \subset \mathcal{L}^1(X)$, and $\|f\|_1 \leq \|f\|_\infty \mu(X)$.*

Proof. If $\mu(X) < \infty$ then $1_X \in \mathcal{L}^1(X)$ with $\|1_X\|_1 = \mu(X)$. Therefore for $f \in \mathcal{L}^\infty(X)$ we have $\|f\|_1 \leq \|f\|_\infty \|1_X\|_1 = \|f\|_\infty \mu(X)$. ∎

We set $\mathcal{L}^1(X)_+ = \mathcal{L}^1(X) \cap \mathcal{M}(X; [0, \infty))$, and we set $\mathcal{L}^\infty(X)_+ = \mathcal{L}^\infty(X) \cap \mathcal{M}(X)_+$.

2.4.3 Functions which agree almost everywhere

Let $(X, \mathcal{M}, \mu)$ be a measure space. So far we have introduced $\mathcal{M}(X)$, $\mathcal{L}^\infty(X)$, and $\mathcal{L}^1(X)$, as the spaces of all measurable, essentially bounded, and integrable functions respectively. It is often the case that we are interested in functions only up to measure zero, and so we consider the spaces $M(X)$, $L^\infty(X)$, and $L^1(X)$ which are respectivley the quotient of the above spaces where we have identified functions which agree almost everywhere. Note that $\mathcal{M}(X)$ does not depend on the measure μ, however $M(X)$ does.

The elements in $M(X)$ are equivalence classes of functions, however it is cumbersome to state this explicitly each time. Thus, in the sequel when we write $f \in M(X)$ (or $f \in L^\infty(X)$, $f \in L^1(X)$) we mean that we can take f to be any function in $\mathcal{M}(X)$ which represents this equivalence class. Similarly, if we write an expression, e.g., $f \leq g$ with $f, g \in M(X)$, then this is expression is meant to be understood as occuring almost everywhere.

As an example, we might say $\{f_n\}_{n=1}^\infty \subset M(X)$, and $f \in M(X)$ such that $f_n \to f$ almost everywhere. This is unambiguous as the countable union of measure zero sets has measure zero, and thus replacing f_n and f by functions which agree almost everywhere does not change the fact that $f_n \to f$ almost everywhere. As long as we restrict to countably many functions/operations at a time this will not cause any difficulty.

2.4.4 Convergence properties

We begin this subsection by improving Lemma 2.4.1 to the case when $f_n \in L^1(X)$.

Theorem 2.4.8. *Let $\{f_n\}_{n \in \mathbb{N}} \subset L^1(X)$ be a Cauchy sequence, then there exists a subsequence $\{f_{n_k}\}_{k \in \mathbb{N}}$ which converges almost everywhere to a measurable function f, and such that for all $\varepsilon > 0$ there exists a measurable set $A \subset X$ with $\mu(A) < \varepsilon$, such that $\{f_{n_k}\}_{k \in \mathbb{N}}$ converges to f uniformly on A^c.*

Proof. In the proof of Lemma 2.4.1 the only reason we needed $f_n \in \mathcal{L}^1_0(X)$ was so that $\|f_n\|_1$ was defined and satisfied $\|f_n\|_1 = \int f_n$, and that if $g \leq h$ then $\int g \leq \int h$. By Theorem 2.4.4 we also have these facts now for general functions in $L^1(X)$. Thus, the proof follows verbatim as in Lemma 2.4.1. ∎

Theorem 2.4.9 (The Monotone Convergence Theorem)**.** *Let $(X, \mathcal{M}, \mu)$ be a measure space. Suppose $\{f_n\}_{n=1}^\infty \subset L^1(X)_+$ is a sequence such that $f_n \leq f_{n+1}$ for all n, and such that $\int f_n$ is bounded. Then $\{f_n\}_{n=1}^\infty$ converges in L^1, and almost everywhere to a function $f \in L^1(X)_+$.*

Proof. Suppose $a = \sup \int f_n < \infty$. Since $\{f_n\}_{n=1}^\infty$ is increasing so is $\{\int f_n\}_{n=1}^\infty$, and for $n \leq m$ we have $\|f_m - f_n\|_1 = \int (f_m - f_n)$. Since $\int f_n \to a$ it then follows that $\{f_n\}_{n=1}^\infty$ is Cauchy in L^1. By the previous theorem there then exists a subsequence which converges in L^1 and almost everywhere to a function $f \in L^1$. Since we have an increasing sequence it then follows that $\{f_n\}_{n=1}^\infty$ converges almost everywhere and in L^1 to f. ∎

Corollary 2.4.10. *Let $(X, \mathcal{M}, \mu)$ be a measure space. Suppose $\{f_n\}_{n=1}^\infty \subset L^1(X)_+$ is a sequence such that $f_{n+1} \leq f_n$ for all j. Then $\{f_n\}_{n=1}^\infty$ converges in L^1, and almost everywhere to a function $f \in L^1(X)_+$.*

Proof. We apply the monotone convergence theorem to the sequence $\{f_1 - f_n\}_{n=1}^\infty$. ∎

Lemma 2.4.11 (Fatou's Lemma)**.** *If $\{f_n\}_{n=1}^\infty \subset L^1(X)_+$, is such that $\liminf_{n\to\infty} \int f_n < \infty$, then $\liminf_{n\to\infty} f_n(x)$ exists for almost every $x \in X$. Moreover, $\liminf_{n\to\infty} f_n$ is a measurable function which is in $L^1(X)$, and we have*

$$\int \liminf_{n\to\infty} f_n \leq \liminf_{n\to\infty} \int f_n.$$

Proof. Fix k and consider the decreasing sequence $\{g_m\}_{m=1}^\infty$ where

$$g_m = \inf\{f_k, f_{k+1}, \ldots, f_m\}.$$

Then $\{g_m\}_{m=1}^\infty$ decreases to $\inf_{m\geq k} f_m$, and applying the previous corollary we have that $\inf_{m\geq k} f_m$ is in $L^1(X)$, and

$$\int \inf_{m\geq k} f_m \leq \inf_{m\geq k} \int f_m.$$

By hypothesis we have that $\{\inf_{m\geq k} \int f_m\}_{m=1}^\infty$ is bounded, and since $\{\inf_{m\geq k} f_m\}_{k=1}^\infty$ is increasing we then have from the monotone convergence theorem that $\liminf_{n\to\infty} f_n$ exists almost everywhere, is in $L^1(X)$, and satisfies

$$\int \liminf_{n\to\infty} f_n = \lim_{n\to\infty} \int \inf_{m\geq n} f_m \leq \liminf_{n\to\infty} \int f_n.$$

∎

Theorem 2.4.12 (The Fatou-Lebesgue Theorem). *Let* $\{f_n\}_{n=1}^{\infty} \subset L^1(X;\mathbb{R})$. *If there exists a function* $g \in L^1(X)_+$ *such that* $|f_n| \leq g$ *for all* $n \geq 1$ *then* $\liminf_{n\to\infty} f_n$ *and* $\limsup_{n\to\infty} f_n$ *exist almost everywhere, are integrable, and we have*

$$\int \liminf_{n\to\infty} f_n \leq \liminf_{n\to\infty} \int f_n \leq \limsup_{n\to\infty} \int f_n \leq \int \limsup_{n\to\infty} f_n.$$

Proof. The first inequality follows from linearity of the integral and by applying Fatou's lemma to the non-negative functions $f_n + g$. The second inequality is obvious. The third inequality follows by applying Fatou's lemma to the non-negative functions $g - f_n$. ■

Theorem 2.4.13 (Lebesgue's Dominated Convergence Theorem). *Suppose* $\{f_n\}$ *is a sequence in* $L^1(X)$, *such that* $f_n \to f$ *almost everywhere. If there exists* $g \in L^1(X)_+$ *such that* $|f_n| \leq g$ *for all* $n \in \mathbb{N}$, *then* $f \in L^1(X)$ *and* $\int f_n \to \int f$.

Proof. Note that $|f| \leq g$ almost everywhere and so by Corollary 2.4.6 we have that $f \in L^1(X)$.

By considering separately the real and imaginary parts of f_n we see that it is enough to consider the case when f_n is real valued. In this case it follows from the Fatou-Lebesgue theorem that

$$\begin{aligned}\int f = \int \liminf_{n\to\infty} f_n &\leq \liminf_{n\to\infty} \int f_n \\ &\leq \limsup_{n\to\infty} \int f_n \\ &\leq \int \limsup_{n\to\infty} f_n = \int f,\end{aligned}$$

and the result then follows. ■

Theorem 2.4.14 (Egorov's Theorem). *Let* $(X, \mathcal{M}, \mu)$ *be a finite measure space, and suppose* $\{f_n\}_{n=1}^{\infty} \subset M(X)$, *and* $f \in M(X)$ *such that* $f_n \to f$ *almost everywhere. Then for each* $\varepsilon > 0$ *there exists* $A \subset X$ *measurable such that* $\mu(A) < \varepsilon$ *and* $f_n \to f$ *uniformly on* A^c.

Proof. We let $E_{n,k} = \{x \in X \mid |f_n(x) - f(x)| \geq 1/k\}$. Then $E_{n,k} \in \mathcal{M}$ and since $f_n(x) \to f(x)$ for almost every $x \in X$ we have that $\mu(\cap_{N=1}^{\infty} \cup_{n=N}^{\infty} E_{n,k}) = 0$, for every $k \in \mathbb{N}$. Thus, there exists N_k so that for each $k \in \mathbb{N}$ we have $\mu(\cup_{n=N_k}^{\infty} E_{n,k}) < \varepsilon 2^{-k}$. We set $A = \cup_{k=1}^{\infty} \cup_{n=N_k}^{\infty} E_{n,k}$, so that $\mu(A) < \varepsilon$.

If $k \in \mathbb{N}$, and $n \geq N_k$ we have $|f_n(x) - f(x)| < 1/k$ for all $x \in A^c$. Therefore $f_n \to f$ unifomrly on A^c. ■

We extend the integral to certain real valued functions as follows: If $g \in \mathcal{L}^1(X;\mathbb{R})$ and $f \in \mathcal{M}(X;[0,\infty))$ is not integrable, then we write $\int(f+g) = \infty$, and $\int -(f+g) = -\infty$. Many of the results above extend to this setting, although the inequalities become trivial in the case when f is not integrable.

2.4.5 Exercises

Exercise 2.4.15 (Chebyshev's inequality)**.** Let $(X, \mathcal{M}, \mu)$ be a measure space and suppose $f \in \mathcal{L}^1(X, \mu)$, then for each $\alpha > 0$ we have

$$\mu(\{x \in X \mid |f(x)| > \alpha\}) \leq \frac{1}{\alpha}\|f\|_1.$$

Exercise 2.4.16. There does not exist a metric d on $L^\infty([0,1], \lambda)$ so that a sequence of functions $\{f_n\}_{n=1}^\infty \subset L^\infty([0,1], \lambda)$ converge almost eveywhere to a function $f \in L^\infty([0,1], \lambda)$ if and only if $d(f_n, f) \to 0$. Hint: Find a sequence $\{f_n\}_{n=1}^\infty \subset L^\infty([0,1], \lambda)$ which does not converge almost everywhere to any function but such that every subsequence has a further subsequence which does converge almost everywhere to some function.

Exercise 2.4.17. Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space and suppose $f \in \mathcal{M}(X; [0, \infty))$. Define

$$I_1(f) = \sup \left\{ \int g \mid g \in \mathcal{L}_0^1(X; [0, \infty)), g \leq f \right\},$$

$$I_2(f) = \inf \left\{ \int g \mid g \in \mathcal{L}_0^1(X; [0, \infty)), f \leq g \right\}.$$

Show that f is integrable if and only if $I_1(f) < \infty$, and this this case we have $I_1(f) = I_2(f) = \int f$.

Exercise 2.4.18. Suppose $(X, \mathcal{M}, \mu)$ is a measure space and $f \in \mathcal{M}(X, [0, \infty))$. Set $F(\lambda) = \mu(f^{-1}([\lambda, \infty)))$. Then F is measurable and $F \in \mathcal{L}^1([0, \infty))$ if and only if $f \in \mathcal{L}^1(X)$. Moreover, in this case we have $\int f d\mu = \int F d\lambda$.

Exercise 2.4.19. Suppose $(X, \mathcal{M}, \mu)$ is a measure space, $(Y, \mathcal{N})$ is a measurable space, and $\theta : X \to Y$ is measurable. Then for all $f \in \mathcal{M}(Y)$ we have $f \circ \theta \in \mathcal{L}^1(X, \mu)$ if and only if $f \in \mathcal{L}^1(Y, \theta_* \mu)$, and in this case we have

$$\int f \circ \theta \, d\mu = \int f \, d(\theta_* \mu).$$

2.5 Product spaces

If $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ are measurable spaces, we let $\mathcal{M} \otimes \mathcal{N} \subset 2^{X \times Y}$ denote the σ-algebra generated by sets of the form $E \times F$ where $E \in \mathcal{M}$ and $F \in \mathcal{N}$. In other words, $\mathcal{M} \otimes \mathcal{N}$ is the smallest σ-algebra so that the projection maps $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ are measurable. More generally, if $\{(X_i, \mathcal{M}_i)\}_{i \in I}$ is a family of measurable spaces then we denote by $\otimes_{i \in I} \mathcal{M}_i$ the smallest σ-algebra so that the projection maps are measurable.

Proposition 2.5.1. *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be measure spaces, then there is measure ζ on $\mathcal{M} \otimes \mathcal{N}$ so that $\zeta(E \times F) = \mu(E)\nu(F)$, for $E \in \mathcal{M}$ and $F \in \mathcal{N}$. Here we use the convention $0 \cdot \infty = 0$. If μ and ν are σ-finite then this measure is unique.*

Proof. We let $\mathcal{A}$ denote the algebra generated by sets of the from $E \times F$ where $E \in \mathcal{M}$ and $F \in \mathcal{N}$. If $E \times F = \cup_{n=1}^{\infty} E_n \times F_n$, where $E, E_n \in \mathcal{M}$, and $F, F_n \in \mathcal{N}$, with $\mu(E), \nu(F) < \infty$. Then for $x \in X$ and $y \in Y$ we have

$$1_E(x)1_F(y) = 1_{E\times F}(x,y) = \sum_{n=1}^{\infty} 1_{E_n\times F_n}(x,y) = \sum_{n=1}^{\infty} 1_{E_n}(x)1_{F_n}(y).$$

Integrating with respect to x, and using the monotone convergence theorem gives

$$\mu(E)1_F(y) = \sum_{n=1}^{\infty} \mu(E_n)1_{F_n}(y).$$

If we then integrate with respect to y we obtain

$$\mu(E)\nu(F) = \sum_{n=1}^{\infty} \mu(E_n)\nu(F_n).$$

If $A \in \mathcal{A}$, then we may write A as a finite disjoint union $A = \cup_{i=1}^{n} E_i \times F_i$ where $E_i \in \mathcal{M}$ and $F_i \in \mathcal{N}$. From above we then see that setting $\zeta_0(A) = \sum_{i=1}^{n} \mu(E_i)\nu(F_i)$ gives a well defined premeasure on $\mathcal{A}$. Carathéodory's extension theorem then shows that this extends to a measure ζ, and if μ and ν are σ-finite then so is ζ and hence this measure is unique. ■

The measure constructed by Carathéodory's extension theorem in the previous proof is called the product measure and denoted by $\mu \times \nu$ (or μ^2 if $\mu = \nu$). We can of course generalize the above proposition easily to any finite number of measure spaces.

If $E \subset X \times Y$, and $x \in X$, $y \in Y$ then we define the **x-section** E_x and **y-section** E^y by $E_x = \{y \in Y \mid (x,y) \in E\}$, and $E^y = \{x \in X \mid (x,y) \in E\}$. Also, if $f : X \times Y \to \mathbb{C}$ we define the **x-section** f_x and **y-section** f^y by $f_x(y) = f^y(x) = f(x,y)$.

Proposition 2.5.2. *Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces. If $E \in \mathcal{M} \otimes \mathcal{N}$ then $E_x \in \mathcal{N}$ and $E^y \in \mathcal{M}$ for all $x \in X$ and $y \in Y$. Also, if $f : X \times Y \to \mathbb{C}$ is $\mathcal{M} \otimes \mathcal{N}$-measurable then f_x is $\mathcal{N}$-measurable and f^y is $\mathcal{M}$-measurable for all $x \in X$ and $y \in Y$.*

Proof. We let Σ denote the collection of subsets $E \subset X \times Y$ such that $E_x \in \mathcal{N}$ and $E^y \in \mathcal{M}$ for all $x \in X$ and $y \in Y$. Then Σ contains all sets of the form $A \times B$ with $A \in \mathcal{M}$ and $B \in \mathcal{N}$. Since $(\cup_{n=1}^{\infty} E_n)_x = \cup_{n=1}^{\infty} (E_n)_x$ and $(E^c)_x = (E_x)^c$, (and similarly for y) it follows that Σ is a σ-algebra and hence must contain $\mathcal{M} \otimes \mathcal{N}$.

If $f : X \times Y \to \mathbb{C}$ is $\mathcal{M} \otimes \mathcal{N}$-measurable, then as $(f_x)^{-1}(C) = (f^{-1}(C))_x$ (and similarly for y) it then follows that f_x and f^y are measurable. ■

If X is a set, then a family $\mathcal{E} \subset 2^X$ is called a **monotone class** if $X \in \mathcal{E}$ and $\mathcal{E}$ is closed under countable monotone unions and intersections, i.e., whenever

$\{E_n\}_{n=1}^\infty \subset \mathcal{E}$ with $E_1 \subset E_2 \subset \cdots$ then we have $\cup_{n=1}^\infty E_n \in \mathcal{E}$, and whenever $\{E_n\}_{n=1}^\infty \subset \mathcal{E}$ with $E_1 \supset E_2 \supset \cdots$ then we have $\cap_{n=1}^\infty E_n \in \mathcal{E}$. Given a family of monotone classes it is clear that the intersection is again a monotone class, thus for any collection of sets $\mathcal{E}_0$ there exists a smallest monotone class which contains $\mathcal{E}_0$, we call this the monotone class which is **generated by** $\mathcal{E}$.

Lemma 2.5.3 (The monotone class lemma)**.** *Suppose $\mathcal{A} \subset 2^X$ is an algebra, then the monotone class generated by $\mathcal{A}$ coincides with the σ-algebra generated by $\mathcal{A}$.*

Proof. We let $\mathcal{M}$ denote the monotone class generated by $\mathcal{A}$. Since a σ-algebra is a monotone class then it suffices to show that $\mathcal{M}$ is a σ-algebra. For this it suffices to show that $\mathcal{M}$ is closed under taking complements and finite unions since if $\{E_n\}_{n=1}^\infty \subset \mathcal{M}$ then $\cup_{n=1}^N E_n$ is a monotone increasing sequence and hence if these finite unions are in $\mathcal{M}$ then so is $\cup_{n=1}^\infty E_n$.

For $E \in \mathcal{M}$ we set $K(E) = \{F \in \mathcal{M} \mid E \setminus F, F \setminus E, E \cup F \in \mathcal{M}\}$. We will show that $K(E) = \mathcal{M}$ for each $E \in \mathcal{M}$. This is shown by the following six steps, each of which is easily verified:

1. If $F \in K(E)$ then $E \in K(F)$. (This follows from symmetry in the definition of $K(E)$ and $K(F)$.)

2. If $E \in \mathcal{A}$ then $A \subset K(E)$. (This follows since $\mathcal{A} \subset \mathcal{M}$).

3. $K(E)$ is a monotone class for all $E \in \mathcal{M}$. (This follows since $\mathcal{M}$ is a monotone class).

4. If $E \in \mathcal{A}$ then $K(E) = \mathcal{M}$. (This follows from (2) and (3) since $\mathcal{M}$ is the smallest monotone class which contains $\mathcal{A}$).

5. $\mathcal{A} \subset K(E)$ for all $E \in \mathcal{M}$. (This follows from (4) and (1)).

6. $K(E) = \mathcal{M}$ for all $E \in \mathcal{M}$. (This follows from (5) and (3)).

■

Lemma 2.5.4. *Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure speaces, and $A \in \mathcal{M} \otimes \mathcal{N}$. Then the functions $x \mapsto \nu(E_x)$ and $y \mapsto \mu(E^y)$ are measurable and*

$$\mu \times \nu(E) = \int \nu(E_x)\, d\mu(x) = \int \mu(E^y)\, d\nu(y).$$

Proof. We first consider the case when μ and ν are finite. We let Σ denote the family of sets in $\mathcal{M} \otimes \mathcal{N}$ such that the conclusion of the proposition holds. Then from the arguemnt in the proof of Proposition 2.5.1 we see that Σ contains the algebra $\mathcal{A}$ generated by sets of the form $E \times F$ with $E \in \mathcal{M}$ and $F \in \mathcal{M}$.

If $\{E_n\}_{n=1}^\infty \subset \Sigma$ such that $E_n \subset E_{n+1}$ for all $n \in \mathbb{N}$, then we have that $\cup_{n=1}^\infty (E_n)_x = (\cup_{n=1}^\infty E_n)_x$ and so by the monotone convergence theorem we have

$$\begin{aligned}\int \nu((\cup_{n=1}^{\infty} E_n)_x)\, d\mu(x) &= \lim_{N\to\infty} \int \nu((E_N)_x)\, d\mu(x)\\ &= \lim_{N\to\infty} \mu\times\nu(E_N)\\ &= \mu\times\nu(\cup_{n=1}^{\infty} E_n).\end{aligned}$$

And we similarly have $\int \mu((\cup_{n=1}^{\infty} E_n)^y)\, d\nu(y) = \mu\times\nu(\cup_{n=1}^{\infty} E_n)$. Thus, $\cup_{n=1}^{\infty} E_n \in \Sigma$. Since μ and ν are finite a similar argument shows that if $E_n \supset E_{n+1}$ for all $n \in \mathbb{N}$, then $\cap_{n=1}^{\infty} E_n \in \Sigma$. Therefore, Σ is a monotone class which contains $\mathcal{A}$ and by the monotone class lemma we have that $\mathcal{M} = \Sigma$.

If μ and ν are σ-finite, say $X = \cup_{n=1}^{\infty} X_n$ and $Y = \cup_{n=1}^{\infty} Y_n$ with $\mu(X_n), \nu(Y_n) < \infty$, then the result follows by first restricting to $X_n \times Y_n$ then using the monotone convergence theorem as we did above. ■

Theorem 2.5.5 (The Fubini-Tonelli Theorem). *Suppose* $(X, \mathcal{M}, \mu)$ *and* $(Y, \mathcal{N}, \nu)$ *are measure spaces, and* $f : X \times Y \to \mathbb{C}$ *is* $\mathcal{M} \otimes \mathcal{N}$*-measurable. Consider the following conditions:*

1. $f \in \mathcal{L}^1(X \times Y, \mu \times \nu)$.
2. *For almost every* $x \in X$, $f_x \in \mathcal{L}^1(Y, \nu)$, *and the function* $f_{L^1(Y)}(x) = \|f_x\|_{L^1(Y)}$ *is in* $L^1(X)$.
3. *For almost every* $y \in Y$, $f^y \in \mathcal{L}^1(X, \mu)$, *and the function* $f_{L^1(X)}(y) = \|f^y\|_{L^1(X)}$ *is in* $L^1(Y)$.

Then (1) *implies both* (2) *and* (3), *and if* μ *and* ν *are* σ*-finite then all three conditions are equivalent.*

Moreover, if (1) *(and hence also* (2) *and* (3)*) is satisfied then we have*

$$\begin{aligned}\int f\, d(\mu\times\nu) &= \int \left(\int f(x,y)\, d\nu(y) \right) d\mu(x) \qquad (2.6)\\ &= \int \left(\int f(x,y)\, d\mu(x) \right) d\nu(y).\end{aligned}$$

Proof. We first consider the case when μ and ν are σ-finite. If f is a characteristic function then the result follows from Lemma 2.5.4. By linearity we then have the result for simple functions. Suppose now that $f \in \mathcal{M}(X \times Y)_+$. Then there exists an increasing sequence of simple functions φ_n which are valued in the non-negative reals so that $\varphi_n(x,y) \to f(x,y)$ for all $(x,y) \in X \times Y$. By the monotone convergence theorem we then have

$$\begin{aligned}\int f\, d(\mu\times\nu) &= \lim_{n\to\infty} \int \varphi_n\, d(\mu\times\nu)\\ &= \lim_{n\to\infty} \int \left(\int \varphi_n(x,y)\, d\nu(y) \right) d\mu(x)\\ &= \int \left(\int f(x,y)\, d\nu(y) \right) d\mu(x).\end{aligned}$$

We similarly have $\int f\,d(\mu\times\nu) = \int\left(\int f(x,y)\,d\mu(x)\right)d\nu(y)$. Thus, for non-negative valued functions we see that the three conditions above are equivalent and that (2.6) holds. From linearity we then get the result for general measurable functions when μ and ν are σ-finite.

If μ or ν is not σ-finite but $f \in L^1(X\times Y, \mu\times\nu)$, then we see that $G_n = \{(x,y) \mid |f(x,y)| \geq 1/n\}$ must have finite measure for all $n \geq 1$. Therefore, there exist $E_k \in \mathcal{M}$ and $F_k \in \mathcal{N}$ so that $G_n \subset \cup_{k=1}^{\infty} E_k \times F_k \subset (\cup_{k=1}^{\infty} E_k) \times (\cup_{k=1}^{\infty} F_k)$, and $\sum_{k=1}^{\infty} \mu(E_k)\nu(F_k) < \infty$. In otherwords, we have $G_n \subset E \times F$ where E and F are σ-finite. It then follows that there exist σ-finite sets $\tilde{E}$, and $\tilde{F}$ so that $\{(x,y) \in X\times Y \mid f(x,y) \neq 0\} = \cup_{n=1}^{\infty} G_n \subset \tilde{E}\times\tilde{F}$. Restricting to the σ-finite measure spaces $(\tilde{E}, \mu_{\tilde{E}})$ and $(\tilde{F}, \nu_{\tilde{F}})$ we see that the result then follows from the σ-finite case. ■

We give the following 2 examples which show how the hypotheses of the Fubini-Tonelli theorem are necessary:

Example 2.5.6. Consider $[0,1]$ with Lebesgue measure, and let $f(x,y) = \frac{x^2-y^2}{(x^2+y^2)^2}$. Then fixing $x \neq 0$ we have

$$\begin{aligned}
\int_0^1 f(x,y)\,dy &= \int_0^1 \frac{x^2+y^2-2y^2}{(x^2+y^2)^2}\,dy \\
&= \int_0^1 \frac{1}{x^2+y^2}\,dy + \int_0^1 y\,d\left(\frac{1}{x^2+y^2}\right) \\
&= \int_0^1 \frac{1}{x^2+y^2}\,dy + \left.\frac{y}{x^2+y^2}\right|_{y=0}^{1} - \int_0^1 \frac{1}{x^2+y^2}\,dy \\
&= \frac{1}{x^2+1}.
\end{aligned}$$

We therefore have

$$\int_0^1 \left(\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}\,dy\right)dx = \tan^{-1}(x)|_{x=0}^{1} = \frac{\pi}{4},$$

and as $f(y,x) = -f(x,y)$ we have $\int_0^1\left(\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}\,dx\right)dy = \frac{-\pi}{4}$. We must therefore have that $\frac{x^2-y^2}{(x^2+y^2)^2} \notin L^1([0,1]^2, \lambda^2)$.

Example 2.5.7. Consider $[0,1]$ with its Borel σ-algebra $\mathcal{B}$, and consider Lebesgue measure λ on $[0,1]$, and also counting measure μ on $[0,1]$. In the product space $([0,1]^2, \mathcal{B}\otimes\mathcal{B}, \lambda\times\mu)$ we may consider the measurable subset $\Delta = \{(x,x) \mid x \in [0,1]\}$, and set $f = 1_\Delta$. Then for every $x \in [0,1]$ we have $f_x \in L^1([0,1],\mu)$ and $\int\left(\int f(x,y)\,d\lambda(y)\right)d\mu(x) = 0$. Similarly, for every $y \in [0,1]$ we have $f_y \in L^1([0,1],\lambda)$ and $\int\left(\int f(x,y)\,d\mu(x)\right)d\lambda(y) = 1$. So that also in this case the iterated integrals do not agree. We must therefore have that $\lambda\times\mu(\Delta) = \infty$, and we see that for non-σ-finite spaces the iterated integrals need not agree even for functions valued in the non-negative reals.

If λ is Lebesgue measure on $\mathbb{R}$ and $n \geq 1$, then **Lebesgue measure on** $\mathbb{R}^n$ is defined to be λ^n. When there is no danger of confusion we will just write λ for λ^n.

Theorem 2.5.8. *If $E \subset \mathbb{R}^n$ is Borel and $t \in \mathbb{R}^n$, then $\lambda(E+t) = \lambda(E)$.*

Proof. If E is a disjoint union of products of intervals than the formula $\lambda(E+t) = \lambda(E)$ clearly holds. As such sets form an algebra which generates the Borel σ-algebra, and since $E \mapsto \lambda(E+t)$ gives a Borel measure, the proposition then follows from uniqueness in Carathéodory's extension theorem. ∎

Theorem 2.5.9. *If $T \in GL_n(\mathbb{R})$ then $T_*\lambda = |\det T|^{-1}\lambda$, i.e., for all Borel sets $E \subset \mathbb{R}^n$ we have*

$$\lambda(T^{-1}(E)) = |\det T|^{-1}\lambda(E). \tag{2.7}$$

Proof. Since every invertible matrix can be row reduced to the identity matrix it follows that every linear transformation $T \in GL_n(\mathbb{R})$ is a composition of elementrary matrices of the following types:

1. $T_1(x_1, \ldots, x_j, \ldots, x_n) = (x_1, \ldots, cx_j, \ldots, x_n)$ with $c \neq 0$.
2. $T_2(x_1, \ldots, x_j, \ldots, x_n) = (x_1, \ldots, x_j + cx_i, \ldots, x_n)$ with $i \neq j$.
3. $T_3(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_n) = (x_1, \ldots, x_j, \ldots, x_i, \ldots, x_n)$.

Also, if we can show that (2.7) holds for matrices of the above type then as the determinant is multiplicative it also holds for their composition. Thus, it is enough to verity (2.7) for matrices of the above type. These all follow easily from the Fubini-Tonelli theorem. Writing

$$\lambda(E) = \int \cdots \int 1_E(x_1, \ldots, x_n)\, d\lambda(x_1) \ldots d\lambda(x_n)$$

we see that (2.7) holds for matrices of the first and second type by their corresponding formulas in one dimensions, while matrices of the third type just correspond to changing the orders of integration. ∎

2.5.1 Exercises

Exercise 2.5.10. Consider $\mathbb{N}$ with the counting measure, and consider the function $f : \mathbb{N}^2 \to \mathbb{C}$ given by $f(n,m) = 1$ if $n = m$, $f(n,m) = -1$ if $n = m+1$, and $f(n,m) = 0$ otherwise. Then $\sum_{n=1}^{\infty} \left(\sum_{m=1}^{\infty} f(n,m)\right)$ and $\sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} f(n,m)\right)$ both exist but are not equal.

2.6 Signed and complex measures

In this section we extend the notion of a measure to allow set functions which may give negative, or even complex, values. The Hahn and Jordan decomposition theorems below, together with the polar decomposition theorem for complex measures, give the main tools to relate this more general setting to the non-negative valued case we have already considered.

2.6.1 Signed measures

A **signed measure** on a measurable space $(X, \mathcal{M})$ is a function $\nu : \mathcal{M} \to [-\infty, \infty]$ such that

1. at most one of the values in $\{-\infty, \infty\}$ are obtained;
2. $\nu(\emptyset) = 0$;
3. if $\{E_n\}_{n=1}^\infty$ are pairwise disjoint measurable sets then
$$\nu(\cup_{n=1}^\infty E_n) = \sum_{n=1}^\infty \nu(E_n),$$
where this series converges absolutely if $\nu(\cup_{n=1}^\infty E_n)$ is finite.

If μ_1 and μ_2 are measures on $(X, \mathcal{M})$, at least one of which is finite, then we obtain a signed measure $\mu_1 - \mu_2$ on $(X, \mathcal{M})$ by setting $(\mu_1 - \mu_2)(E) = \mu_1(E) - \mu_2(E)$ for any $E \in \mathcal{M}$.

If ν is a signed measure on $(X, \mathcal{M})$, and $E \in \mathcal{M}$, then we say that E is **positive** (resp. **negative**, **null**) with respect to ν if for all $F \subset E$ measurable we have $\nu(F) \geq 0$ (resp. ≤ 0, $= 0$). Note that positive (resp. negative, null) sets are preserved under taking measurable subsets, and also under taking countable unions.

If μ is a measure on $(X, \mathcal{M})$ and $E \in \mathcal{M}$ with $\mu(E) < \infty$, the we may obtain a signed measure on $(X, \mathcal{M})$ by setting $\nu(F) = \mu(F \cap E^c) - \mu(F \cap E)$. In this case E is a negative set, E^c is a positive set, and a set is null for ν if and only if it is null for μ. The Hahn decomposition theorem shows that every signed measure (or its negative) arrises in this way.

Lemma 2.6.1. *Let ν be a signed measure on $(X, \mathcal{M})$, and suppose $E \in \mathcal{M}$ such that $-\infty < \nu(E) < 0$. Then there exists a negative set $N \subset E$ with $\nu(N) < 0$.*

Proof. Note first that there does not exist a measurable subset $F \subset E$ with $\nu(F) = \infty$, since otherwise we would have $\nu(E) = \nu(F) + \nu(E \setminus F) = \infty$. We inductively define a non-decreasing sequence of number $\{n_k\}_{k=0}^\infty \subset \mathbb{N} \cup \{\infty\}$, and pairwise disjoint subsets $\{E_k\}_{k=0}^\infty$ of E as follows: Set $E_0 = \emptyset$ and $n_0 = 1$. Having defined $n_0, \ldots, n_k$ and $E_0, \ldots, E_k$ we let n_{k+1} denote the smallest integer such that there exists a measurable subset $E_{k+1} \subset E \setminus (\cup_{j=0}^k E_j)$ with $\nu(E_{k+1}) \geq 1/n_{k+1}$. If no such number exists we set $n_{k+1} = \infty$ and $E_{k+1} = \emptyset$.

We then have that $\infty > \nu(\cup_{k=0}^\infty E_k) = \sum_{k=0}^\infty \nu(E_k) \geq \sum_{k=0}^\infty \frac{1}{n_k} \geq 0$ (here we use the convention $\frac{1}{\infty} = 0$). Since this series converges we must have $n_k \to \infty$. We set $N = E \setminus (\cup_{k=0}^\infty E_k)$ then we have $\nu(N) \leq \nu(E) < 0$. If $F \subset N$ is measurable then by our choice of n_k we must have $\nu(F) \leq 1/(n_k - 1)$ for each k. Since $n_k \to \infty$ this then shows that $\nu(F) \leq 0$, and hence N is negative. ∎

Theorem 2.6.2 (The Hahn decomposition theorem)**.** *Let $(X, \mathcal{M})$ be a measurable space and let ν be a signed measure on $(X, \mathcal{M})$. Then there exists a positive set $P \in \mathcal{M}$ so that $N = P^c$ is a negative set. Moreover if $\tilde{P}$ is another positive set such that $\tilde{P}^c$ is a negative set, then we have that $P \Delta \tilde{P}$ is null.*

Proof. We may assume that the value $-\infty$ is never obtained (otherwise consider $-\nu$). We let $a = \inf\{\nu(E) \mid E \in \mathcal{M}, E \text{ negative}\} \leq 0$. We take negative sets $E_n \in \mathcal{M}$ so that $\nu(E_n) \to a$, and we set $N = \cup_{n=1}^{\infty} E_n$. Then N is a negative set, and we have $\nu(N) = \nu(E_n) + \nu(N \setminus E_n) \leq \nu(E_n)$ for each $n \geq 1$, hence $\nu(N) = a$.

We claim that N^c is positive. Otherwise, there would exist a measurable set $E \subset N^c$ with $-\infty < \nu(E) < 0$, and by the previous lemma there would then exist a negative set $N_0 \subset E$ with $\nu(N_0) < 0$. However, we would then have that $N \cup N_0$ is negative and $\nu(N \cup N_0) = \nu(N) + \nu(N_0) < a$, contradicting our definition of a. This then finishes the existence part of the theorem.

Suppose now that $\tilde{N}$ is another negative set such that $\tilde{N}^c$ is positive, and let $F \subset \tilde{N} \setminus N$ be measurable. Then $F \subset N$ hence $\nu(F) \leq 0$, and $F \subset \tilde{N}^c$ hence $\nu(F) \geq 0$. Therefore $\nu(F) = 0$ and $\tilde{N} \setminus N$ is a null set. We similarly have that $N \setminus \tilde{N}$ is null and hence so is $N \Delta \tilde{N}$. ■

If $(X, \mathcal{M})$ is a measurable space, then two (signed) measures μ, η on $(X, \mathcal{M})$ are **singular**, which we write as $\mu \perp \eta$, if there exists $E \in \mathcal{M}$ so that E is a conull set for μ and E^c is a conull set for η. Note that this is a symmetric relation.

Theorem 2.6.3 (The Jordan decomposition theorem)**.** *Let $(X, \mathcal{M})$ be a measurable space and let ν be a signed measure on $(X, \mathcal{M})$. Then there exist unique singular measures ν_-, ν_+ on $(X, \mathcal{M})$, at least one of which is finite, so that $\nu = \nu_+ - \nu_-$.*

Proof. From the Hahn decomposition theorem there exists $P \in \mathcal{M}$ a positive set so that $N = P^c$ is a negative set. We define ν_+ by $\nu_+(E) = \nu(E \cap P)$ for all $E \in \mathcal{M}$ and we define ν_- by $\nu_-(E) = -\nu(E \cap N)$ for all $E \in \mathcal{M}$. That these define measures is easily seen from the definition of a signed measure. Moreover, we have $\nu(E) = \nu(E \cap P) + \nu(E \cap N) = \nu_+(E) - \nu_-(E)$ for all $E \in \mathcal{M}$, hence at least one of ν_- or ν_+ is finite and we have $\nu = \nu_+ - \nu_-$. Since, $\nu_+(N) = \nu(N \cap P) = -\nu_-(P) = 0$ we have $\nu_+ \perp \nu_-$.

Suppose now that η_1, η_2 are singular measure on $(X, \mathcal{M})$, at least one of which is finite, such that $\nu = \eta_1 - \eta_2$. Take $E \in \mathcal{M}$ so that E is a conull set for η_1 and E^c is a conull set for η_2. Then we clearly have that E is a positive set for ν, and E^c is a negative set for ν. Therefore, $P \Delta E$ is a null set for ν by the uniqueness part of the Hahn decomposition theorem. If we have $F \in \mathcal{M}$ then we have

$$\eta_1(F) = \eta_1(F \cap E) = \nu(F \cap E) = \nu(F \cap P) = \nu_+(F).$$

Hence, $\eta_1 = \nu_+$. We similarly have that $\eta_2 = \nu_-$ which then shows uniqueness. ■

If ν is a signed measure on $(X, \mathcal{M})$ then the measures ν_+ and ν_- are called respectively the **positive and negative variations** of ν. The measure $|\nu| = \nu_+ + \nu_-$ is called the **absolute variation** of ν. We also set $\|\nu\| = |\nu|(X)$ and call this the **total variation** of ν. It's easy to see that the absolutely variation satisfies

$$|\nu|(A) = \sup \sum_{k=1}^{\infty} |\nu(E_k)|, \tag{2.8}$$

where the supremum is taken over all measurable partitions of A.

2.6.2 Complex measures

A **complex measure** on a measurable space $(X, \mathcal{M})$ is a set function $\nu : \mathcal{M} \to \mathbb{C}$ such that

1. $\nu(\emptyset) = 0$;
2. if $\{E_n\}_{n=1}^{\infty}$ are pairwise disjoint measurable sets then

$$\nu(\cup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \nu(E_n),$$

   where this series converges absolutely.

Given a complex measure ν, we may consider the real and imaginary parts $\mathrm{Re}\,(\nu)$, $\mathrm{Im}\,(\nu)$, which give signed measures such that $\nu(E) = \mathrm{Re}\,(\nu)(E) + i\mathrm{Im}\,(\nu)(E)$ for each $E \in \mathcal{M}$. The **absolute variation** of ν is the set function $|\nu| : \mathcal{M} \to [0, \infty]$ given by equation (2.8). The **total variation** of ν is given by $\|\nu\| = |\nu|(X)$.

Proposition 2.6.4. *Let ν be a complex measure on $(X, \mathcal{M})$, then $|\nu|$ is a measure on $(X, \mathcal{M})$, $\|\nu\| < \infty$, and for all $A \in \mathcal{M}$ we have $|\nu(A)| \leq |\nu|(A)$.*

Proof. Suppose $\{A_n\}_{n=1}^{\infty} \subset \mathcal{M}$ is a sequence of pairwise disjoint sets. If $\{E_k\}_{k=1}^{\infty}$ gives a measurable partition of $\cup_{n=1}^{\infty} A_n$, then for each $n \geq 1$ we have a measurable partition of A_n given by $\{E_k \cap A_n\}_{k=1}^{\infty}$. We therefore have

$$\sum_{n=1}^{\infty} |\nu|(A_n) \geq \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} |\nu(E_k \cap A_n)| \geq \sum_{k=1}^{\infty} \left| \sum_{n=1}^{\infty} \nu(E_k \cap A_n) \right| = \sum_{k=1}^{\infty} |\nu(E_k)|.$$

Taking supremums over all such partitions then gives $\sum_{n=1}^{\infty} |\nu|(A_n) \geq |\nu|(\cup_{n=1}^{\infty} A_n)$.

Also, if $\varepsilon > 0$, and if $\{E_k^n\}_{k=1}^{\infty}$ is a measurable partition of A_n such that $\sum_{k=1}^{\infty} |\nu(E_k^n)| + \varepsilon 2^{-n} \geq |\nu|(A_n)$, then we have

$$|\nu|(\cup_{n=1}^{\infty} A_n) + \varepsilon \geq \sum_{n,k=1}^{\infty} |\nu(E_k^n)| + \varepsilon \geq \sum_{n=1}^{\infty} |\nu|(A_n).$$

Therefore, $|\nu|(\cup_{n=1}^{\infty} A_n) \geq \sum_{n=1}^{\infty} |\nu|(A_n)$, and hence $|\nu|$ is a measure.

We have $\|\nu\| \leq \|\mathrm{Re}\,(\nu)\| + \|\mathrm{Im}\,(\nu)\|$, and from the Hahn decomposition theorem we see that $\|\mathrm{Re}\,(\nu)\| + \|\mathrm{Im}\,(\nu)\| < \infty$. Hence, $\|\nu\| < \infty$. Also, if $E \in \mathcal{M}$ then the inequality $|\nu(E)| \leq |\nu|(E)$ follows easily from the definition of $|\nu|$. ■

2.6.3 Exercises

Exercise 2.6.5. Consider $[0,1]$ with the Borel σ-algebra. Let ν be counting measure and μ be Lebesgue measure on $[0,1]$, then there do not exist Borel measures ν_0, ν_1 on $[0,1]$ so that $\nu = \nu_0 + \nu_1$, $\nu_0 \perp \mu$, and $\nu_1 \ll \mu$.

If μ is a measure on $(X, \mathcal{M})$ and $f \in L^1(X, \mu)$ then we obtain a complex valued measure $f\mu$ by $(f\mu)(E) = \int_E f \, d\mu$.

Exercise 2.6.6. If $f \in L^1(X, \mu)$ then $|f\mu| = |f|\mu$, and $\|f\mu\| = \|f\|_1$.

We let $M_b(X)$ denote the space of all complex valued measures on $(X, \mathcal{M})$.

Exercise 2.6.7. The map $\nu \mapsto \|\nu\|$ gives a norm on $M_b(X)$, and with this norm $M_b(X)$ is a Banach space.

2.7 The Radon-Nikodym Theorem

If μ and ν are measures on $(X, \mathcal{M})$, then ν is **absolutely continuous** with respect to μ (and we write $\nu \ll \mu$) if every μ-null set is also a ν-null set. The terminology is justified by the following proposition:

Proposition 2.7.1. *Suppose μ and ν are measures on $(X, \mathcal{M})$ with ν finite, then $\nu \ll \mu$ if and only if for every $\varepsilon > 0$ there exists $\delta > 0$ so that if $E \in \mathcal{M}$ with $\mu(E) < \delta$ then we have $\nu(E) < \varepsilon$.*

Proof. Clearly the above condition implies that ν is absolutely continuous with respect to μ, thus we only need to show the converse. Suppose therefore that the condition above does not hold. Then there exists $\varepsilon > 0$ and $E_n \in \mathcal{M}$ so that $\mu(E_n) < 2^{-n}$, and $\nu(E_n) \geq \varepsilon$ for all $n \geq 1$. We let $F_k = \cup_{n=k}^\infty E_n$, and set $F = \cap_{k=1}^\infty F_k$. Then F_k is decreasing and $\mu(F_k) \to 0$, so that $\mu(F) = 0$. However, $\nu(F_k) \geq \varepsilon$ for all $k \geq 1$, hence $\nu(F) \geq \varepsilon$ showing that ν is not absolutely continuous with respect to μ. ■

Theorem 2.7.2 (The Lebesgue decomposition theorem)**.** *Suppose μ and ν are measures on $(X, \mathcal{M})$ such that ν is σ-finite, then there exist unique measures ν_0 and ν_1 on $(X, \mathcal{M})$ so that $\nu = \nu_0 + \nu_1$, $\nu_0 \perp \mu$, and $\nu_1 \ll \mu$.*

Proof. We first consider the case when $\nu(X) < \infty$. We let $\mathcal{N}$ denote the space of μ-null sets, and we set $a = \sup\{\nu(E) \mid E \in \mathcal{N}\}$. We take $E_n \in \mathcal{N}$ so that $\nu(E_n) \to a$ and we set $E = \cup_{n=1}^\infty E_n$. Then $E \in \mathcal{N}$ and $\nu(E) \geq \nu(E_n)$ for all $n \geq 1$, hence $\nu(E) = a$.

We let ν_0 be the measure given by $\nu_0(F) = \nu(F \cap E)$, and we let $\nu_1(F) = \nu(F \cap E^c)$. Then we clearly have $\nu = \nu_0 + \nu_1$, and we also have $\nu_0 \perp \mu$ since E is a μ-null set and E^c is a ν_0-null set. If $F \in \mathcal{N}$, then $F \cup E \in \mathcal{N}$ and hence $a \geq \nu(F \cup E) = \nu(F \cap E^c) + \nu(E) \geq a$, therefore we must have $\nu(F \cap E^c) = 0$. It therefore follows that $\nu_1(F) = \nu(F \cap E^c) = 0$ and hence $\nu_1 \ll \mu$.

If $\nu = \tilde{\nu}_0 + \tilde{\nu}_1$ is another decomposition with $\tilde{\nu}_0 \perp \mu$ and $\tilde{\nu}_1 \ll \mu$, then for all $F \in \mathcal{N}$ we have $\tilde{\nu}_0(F) = \nu(F) = \nu_0(F)$, and since $\tilde{\nu}_0 \perp \mu$ we then have that

$\tilde{\nu}_0 = \nu_0$, and it follows that $\tilde{\nu}_1 = \nu_1$. This then finishes the theorem in the case when $\nu(X) < \infty$.

For the general case, we write $X = \sup_{n=1}^\infty E_n$ where $\{E_n\}_{n=1}^\infty \subset \mathcal{M}$ is a pairwise disjoint sequence with $\nu(E_n) < \infty$. We consider the restriction of ν to E_n and from above there are unique measures ν_0^n, ν_1^n which have the conull set E_n and such that $\nu_0^n \perp \mu$ and $\nu_1^n \ll \mu$. If we set $\nu_0 = \sum_{n=1}^\infty \nu_0^n$ and $\nu_1 = \sum_{n=1}^\infty \nu_1^n$ then it is then easy to see that ν_0 and ν_1 are then the unique measures which satisfy the conclusion of the theorem. ■

Lemma 2.7.3. *Suppose η and μ are measures on $(X, \mathcal{M})$, with μ finite, $\eta \neq 0$ and such that $\eta \ll \mu$. Then there exists $\delta > 0$, and $E \in \mathcal{M}$ so that $\mu(E) > 0$ and $\eta \geq \delta\mu$ on E, i.e., $\eta(F) \geq \delta\mu(F)$ for all $F \in \mathcal{M}$, $F \subset E$.*

Proof. For each $n \in \mathbb{N}$ we consider a Hahn decomposition $X = P_n \cup N_n$ of $\eta - \frac{1}{n}\mu$. We set $P = \cup_{n=1}^\infty P_n$ and $N = \cap_{n=1}^\infty N_n$, so that $P = N^c$. Since N is a negative set for $\eta - \frac{1}{n}\mu$ for each n, we have $0 \leq \eta(N) \leq \frac{1}{n}\mu(N)$ for each n, and hence $\eta(N) = 0$, so that $\eta(P) > 0$. Since $\eta \ll \mu$ we than also have $\mu(P) > 0$.

Therefore, for some n we have $\mu(P_n) > 0$, and as P_n is a positive set for $\eta - \frac{1}{n}\mu$ we have $\eta(F) \geq \frac{1}{n}\mu(F)$ for all $F \in \mathcal{M}$, $F \subset P_n$. ■

Theorem 2.7.4 (The Radon-Nikodym theorem)**.** *Let μ and ν be σ-finite measures on $(X, \mathcal{M})$ such that $\nu \ll \mu$. Then there exists a unique $f \in M(X, \mu)$ so that for all $E \in \mathcal{M}$ we have*

$$\nu(E) = \int_E f \, d\mu. \tag{2.9}$$

Proof. We first consider the case when μ and ν are finite. We set

$$\mathcal{F} = \left\{ f \in \mathcal{M}(X, [0, \infty)) \mid \int_E f \, d\mu \leq \nu(E), \text{ for all } E \in \mathcal{M} \right\},$$

and

$$a = \sup_{f \in \mathcal{F}} \int f \, d\mu.$$

Note that if $f, g \in \mathcal{F}$ then $h = \max\{f, g\} \in \mathcal{F}$, since if we set $F_0 = \{x \in X \mid f(x) \geq g(x)\}$, then for $E \in \mathcal{M}$ we have $\int_E h \, d\mu \leq \int_{F_0 \cap E} f \, d\mu + \int_{F_0^c \cap E} g \, d\mu \leq \nu(F_0 \cap E) + \nu(F_0^c \cap E) = \nu(E)$.

We choose $f_n \in \mathcal{F}$ so that $\int f_n \, d\mu \to a$. Setting $h_n = \max\{f_1, \ldots, f_n\}$, we then have $h_n \in \mathcal{F}$, and $\{h_n\}_{n=1}^\infty$ is an increasing sequence. If we set $h = \lim_{n\to\infty} h_n$ then for each $E \in \mathcal{M}$ it follows from the monotone convergence theorem that $\int_E h \, d\mu = \lim_{n\to\infty} \int_E h_n \, d\mu \leq \nu(E)$. Therefore $h \in \mathcal{F}$, and we have $a \geq \int h \, d\mu = \lim_{n\to\infty} \int f_n \, d\mu = a$. So that $\int h \, d\mu = a$.

We claim that $\nu = \mu_h$. If not, then setting $\eta = \nu - \mu_h$, we have $\eta \ll \nu \ll \mu$, and $\eta \neq 0$, so that by Lemma 2.7.3 there exists $\delta > 0$ and $E \in \mathcal{M}$ with $\mu(E) > 0$ so that $\eta \geq \delta\mu$ on E, or equivalently, $\mu_h + \delta\mu \leq \nu$ on E. This would then show that $h + \delta 1_E \in \mathcal{F}$, and hence $a \geq \int (h + \delta 1_E) \, d\mu = a + \delta\mu(E) > a$, giving a contradiction. If $\tilde{h}$ were another such function then we would have $\int_E h\,d\mu = \int_E \tilde{h}\,d\mu$ for all $E \in \mathcal{M}$ and from this it follows that $h = \tilde{h}$, μ-almost everywhere.

In general, since μ and ν are σ-finite, we may decompose X as a countable disjoint union of measurable sets $X = \cup_{n=1}^{\infty} X_n$, such that $\mu(X_n), \nu(X_n) < \infty$. By the finite measure case above, there then exists a measurable function $f_n \in \mathcal{M}(X, \mu)$ so that for all $E \in \mathcal{M}$ we have $\nu(E \cap X_n) = \int_{E \cap X_n} f_n\,d\mu$. If we set $f = \sum_{n=1}^{\infty} f_n$ then it is easy to see that for all $E \in \mathcal{M}$ we have $\nu(E) = \int_E f\,d\mu$. Uniqueness follows similar to the finite case above. ■

The function $f \in M(X, \mu)$ in the previous theorem is called the **Radon-Nikodym derivative** of ν with respect to μ and denoted by $\frac{d\nu}{d\mu}$.

There is also a Radon-Nikodym theorem for complex measures:

Theorem 2.7.5 (The Radon-Nikodym theorem for complex measures). *Let μ be a σ-finite measure on $(X, \mathcal{M})$ and let ν be a complex measure on $(X, \mathcal{M})$ which is absolutely continuous with respect to μ, then there exists a unique $f \in L^1(X, \mu)$ so that $\nu(E) = \int_E f\,d\mu$ for all $E \in \mathcal{M}$.*

Proof. By considering the real and imaginary parts separately it is enough to consider the case when ν is a finite signed measure. We let $\nu = \nu_+ - \nu_-$ be the Jordan decomposition. Then ν_+ and ν_- are both absolutely continuous with respect to μ and so by the Radon-Nikodym theorem there exists $f_+, f_- \in M(X, \mu)$ so that $\nu_\pm = f_\pm \mu$. Note that since ν_+, and ν_- are finite measures we have that $f_+, f_- \in L^1(X, \mu)$. We then have $\nu = f\mu$ and $f \in L^1(X, \mu)$. Uniqueness follows just as in the previous theorem. ■

Corollary 2.7.6 (Polar decomposition for complex measures). *Let μ be a complex measure on $(X, \mathcal{M})$, then there exists $f : X \to \mathbb{T}$ measurable such that $\mu = f|\mu|$. Moreover, if $g : X \to \mathbb{T}$ is a measurable function such that $\mu = g|\mu|$ then $g = f$ $|\mu|$-almost everywhere.*

Proof. Since $\mu \ll |\mu|$ it follows from the Radon-Nikodym theorem that there exists a unique $f \in M(X, |\mu|)$ so that $\mu = f|\mu|$. We just need to show that $f(x) \in \mathbb{T}$ for $|\mu|$-almost every $x \in X$.

Suppose this were not the case. Then there exists $\alpha \notin \mathbb{T}$ so that α is in the essential range of f. We take $\delta > 0$ so that $2\delta < d(\alpha, \mathbb{T})$. We set $E = f^{-1}(B(\delta, \alpha))$ so that $|\mu|(E) > 0$.

Assume first that $|\alpha| > 1$. Since $\mu = f|\mu|$ it then follows that for any measurable set $F \subset E$ we have $|\mu(F)| = |f|\mu|(F)| \geq (\alpha - \delta)|\mu|(F)$. If $\cup_{n=1}^{\infty} E_n$ gives a measurable partition of E then we have $|\mu|(E) \geq (\alpha - \delta) \sum_{n=1}^{\infty} |\mu|(E_n)$. Taking a supremum over all partitions gives $|\mu|(E) \geq (\alpha - \delta)|\mu|(E)$, a contradiction.

If we had $|\alpha| < 1$, then a similar computation would show that $|\mu|(E) \leq (\alpha + \delta)|\mu|(E)$ which is again a contradiction. We must therefore have that $f(x) \in \mathbb{T}$ for $|\mu|$-almost every $x \in X$. ■

Lemma 2.7.7. *A measure space $(X, \mathcal{M}, \mu)$ is semifinite if and only if for all $f \in M(X, \mu)$ we have*

$$\|f\|_\infty = \sup \left\{ \left| \int fg\, d\mu \right| \mid g \in L^1(X, \mu), \|g\|_1 \leq 1 \right\}.$$

(Where we set $\|f\|_\infty = \infty$ if $f \notin L^\infty(X, \mu)$.)

Proof. Suppose first that μ is semifinite. If $f \in M(X, \mu)$, take $w \in L^\infty(X, \mu; \mathbb{T})$ so that $wf = |f|$. Set $\alpha = \|f\|_\infty = \||f|\|_\infty$ and fix $\varepsilon > 0$. If $\alpha < \infty$ then α is in the essential range of $|f|$ and hence $F = |f|^{-1}((\alpha - \varepsilon, \alpha])$ has positive measure. We let $E \subset F$ be a measurable set with finite positive measure (which exists by semifiniteness) and set $g = \frac{1}{\mu(E)} 1_E$ then $\|wg\|_1 = \|g\|_1 = 1$ and by Theorem 2.4.5 we have

$$\alpha \geq \left| \int fwg\, d\mu \right| = \frac{1}{\mu(E)} \int_E |f|\, d\mu \geq \alpha - \varepsilon.$$

So that $\|f\|_\infty = \sup\{| \int fg\, d\mu| \mid g \in L^1(X, \mu), \|g\|_1 \leq 1\}$.

Similarly, if $\alpha = \infty$ then $F = |f|^{-1}((N, \infty))$ has positive measure for all $N > 0$, and the same argument above then shows that $\|f\|_\infty = \infty = \sup\{| \int fg\, d\mu| \mid g \in L^1(X, \mu), \|g\|_1 \leq 1\}$.

Conversely, suppose μ is not semifinite. Then there exists $E \in \mathcal{M}$ so that $\mu(E) = \infty$, and for any measurable subset $F \subset E$ we have $\mu(F) \in \{0, \infty\}$, hence if $g \in L^1(X, \mu)$ we must have $g(x) = 0$ for almost every $x \in E$. Setting $f = 1_E$ we then have $\|f\|_\infty = 1$, while $\sup \left\{ \left| \int fg\, d\mu \right| \mid g \in L^1(X, \mu), \|g\|_1 \leq 1 \right\} = 0$. ■

Theorem 2.7.8. *Let $(X, \mathcal{M}, \mu)$ be a measure space and consider the map $\Psi : L^\infty(X, \mu) \to L^1(X, \mu)^*$ given by $\Psi(f)(g) = \int fg\, d\mu$. Then Ψ is isometric if μ is semifinite, and Ψ is surjective if μ has the essential suprema property.*

Proof. That Ψ maps into $L^1(X, \mu)^*$ follows from Theorem 2.4.5. From Lemma 2.7.7 we see that this map is injective if and only if μ is semifinite.

Suppose μ has the essential suprema property, and $\varphi \in L^1(X, \mu)^*$. Fix $E \subset X$ so that $\mu(E) < \infty$. Then $F \mapsto \varphi(1_{F \cap E})$ defines a complex measure on $(X, \mathcal{M})$ which is absolutely continuous with respect to μ and so by the Radon-Nikodym theorem there exists a function $f_E \in \mathcal{M}(X, \mu)$ so that $\varphi(1_{F \cap E}) = \int_{F \cap E} f_E\, d\mu$ for all $F \in \mathcal{M}$. Since $\mu(E) < \infty$ we have from Lemma 2.7.7 that $\|f_E\|_\infty \leq \|\varphi\|$.

Note that by uniqueness in the Radon-Nikodym we have that if $E_1, E_2 \in \mathcal{M}$ have finite measure then f_{E_1} and f_{E_2} agree almost everywhere on $E_1 \cap E_2$. If we let f denote an essential supremum of $\{f_E \mid \mu(E) < \infty\}$ as in Proposition 2.2.5, then $\|f\|_\infty \leq \|\varphi\|$ and for each $E \in \mathcal{M}$ with $\mu(E) < \infty$ we have $f(x) = f_E(x)$ for almost every $x \in E$.

It then follows that for every function $g \in L^1(X, \mu)$ such that $\mu(\{x \in X \mid g(x) \neq 0\}) < \infty$ we have $\varphi(g) = \int fg\, d\mu$. Since functions of this type are dense in $L^1(X, \mu)$ it follows that $\varphi(g) = \int fg\, d\mu$ for each $g \in L^1(X, \mu)$. ■

2.7.1 Exercises

Exercise 2.7.9. Suppose $\lambda \ll \nu \ll \mu$, then

$$\frac{d\mu}{d\lambda} = \frac{d\mu}{d\nu}\frac{d\nu}{d\lambda} \qquad \mu - \text{almost everywhere.}$$

Exercise 2.7.10. If $\nu \ll \mu$ and $\mu \ll \nu$ then

$$\frac{d\mu}{d\nu} = \left(\frac{d\nu}{d\mu}\right)^{-1} \qquad \mu - \text{almost everywhere.}$$

Exercise 2.7.11. If $\nu \ll \mu$ and $g \in \mathcal{M}(X)$, then g is μ-integrable if and only if $g\frac{d\nu}{d\mu}$ is ν-integrable, and in this case we have

$$\int g\, d\mu = \int g \frac{d\mu}{d\nu}\, d\nu.$$

Chapter 3

Point set topology

3.1 Topological spaces

Let X be a set. A **topology** on X is a family $\mathcal{T}$ of subsets of X, which contains $\emptyset$ and X, and is closed under finite intersections and arbitrary unions. A topological space is a pair $(X, \mathcal{T})$ consisting of a set X, together with a topology $\mathcal{T}$ on X. When $\mathcal{T}$ is understood we sometimes refer to the topological space X. The following are examples of topological spaces:

1. If X is any set then 2^X and $\{\emptyset, X\}$ are both topologies on X, called the **discrete** and **trivial** (or **indiscrete**) topologies respectively.

2. If (X, d) is a metric space and $\mathcal{T}$ consists of all open subsets of X then $(X, \mathcal{T})$ is a topological space. In this case we call $(X, \mathcal{T})$ **metrizable**.

3. If X is a set then $\mathcal{T} = \{\emptyset\} \cup \{U \subset X \mid U^c \text{ is finite}\}$ gives a topology on X.

Generalizing the case of metric spaces, we call the sets in $\mathcal{T}$, **open** sets, and we call a set **closed** if its complement is open. If a set is both open and closed then we say it is **clopen**. A set $A \subset X$ is a G_δ**-set** if A is the intersection of countably many open sets, and a set $B \subset X$ is an F_σ**-set** if B is the countable union of closed sets.

If $A \subset X$, then the closure $\overline{A}$ of A is the intersection of all closed sets containing A, and hence is the smallest closed set containing A. The interior A^o of A is the union of all open sets contained in A. The difference $\overline{A} \setminus A^o$ is the **boundary** of A and denoted by ∂A. If $\overline{A} = X$ then A is **dense** and if $A^o = \emptyset$ then A is **nowhere dense**. A topological space $(X, \mathcal{T})$ is **separable** if it has a countable dense subset.

Given two topologies $\mathcal{T}_1$ and $\mathcal{T}_2$ on X such that $\mathcal{T}_1 \subset \mathcal{T}_2$, we say that $\mathcal{T}_1$ is **weaker** (or **coarser**) than $\mathcal{T}_2$, and $\mathcal{T}_2$ is **stronger** (or **finer**) than $\mathcal{T}_1$. Thus, the trivial topology is the coarsest topology, while the discrete topology is the finest topology. If $\mathcal{E} \subset 2^X$, then the intersection of all topologies on X which contain $\mathcal{E}$ is clearly a topology and is denoted by $\mathcal{T}(\mathcal{E})$. It is called the topology generated by $\mathcal{E}$. For example:

1. if $K = \mathbb{R}^n$, or $\mathbb{C}^n$, then the **Zariski topology** on K is the weakest topology such that the zero set $\{k \in K \mid p(k) = 0\}$ of any polynomial p is closed.

2. the **Sorgenfrey line** $\mathbb{R}_l$ is the space $\mathbb{R}$, together with the topology generated by all half-open intervals $[a, b)$.

3. the **Moore plane** is the (closed) upper half plane $\Gamma = \{(x, y) \in \mathbb{R}^2 \mid y \geq 0\}$, together with the topology generated by Euclidean open sets, and all sets of the form $\{(x_0, 0)\} \cup (O \setminus \{(x, 0) \mid x \in \mathbb{R}\})$ where O is an open neighborhood of x_0 in the Euclidean sense.

4. If $(X, \leq)$ is a linearly ordered set, the **order topology** on X is the topology generated by the open sets $\{x \in X \mid a < x < b\}$ for all pair $(a, b) \in X^2$ such that $a < b$.

If $x \in X$ then a **neighborhood** of x is a set $A \subset X$ so that $x \in O \subset A$ for some open set $O \in \mathcal{T}$. A point $x \in X$ is an **accumulation point** (or **condensation point**, or **limit point**) of a set E if every neighborhood of x has nonempty intersection with $E \setminus \{x\}$. A **neighborhood base** for $x \in X$ is a family $\{O_i\}_{i \in I}$ of open neighborhoods of x such that for any open neighborhood O of x there is some O_i so that $O_i \subset O$. A **base** for the topology $\mathcal{T}$ is a family $\{O_i\}_{i \in I}$ of open sets which contains a neighborhood base for any point $x \in X$.

Proposition 3.1.1. *If $\mathcal{E} \subset 2^X$ then $\mathcal{T}(\mathcal{E})$ consists of all unions of finite intersections of $\mathcal{E}$.*

Proof. If we let $\mathcal{T}$ denote the unions of all finite intersections of $\mathcal{E}$ then we must show that $\mathcal{T}$ is a topology. Clearly $\mathcal{T}$ is closed under arbitrary unions. Suppose $U_1, \ldots, U_n \in \mathcal{T}$. Then we may write $U_i = \cup_{j \in J_i} O_{j,i}$ where $O_{j,i}$ is a finite intersection of sets in $\mathcal{E}$. Therefore $\cap_{i=1}^n U_i = \cup_{j_1 \in J_1, \ldots, j_n \in J_n} (\cap_{i=1}^n O_{j_i, i}) \in \mathcal{T}$, so that $\mathcal{T}$ is also closed under finite intersections. ■

A topological space $(X, \mathcal{T})$ is **first countable** if each point has a countable neighborhood base which is countable. $(X, \mathcal{T})$ is **second countable** if it has a countable base.

A topological space is:

1. T_1 if $\{x\}$ is closed for each point $x \in X$;

2. **Hausdorff** (or T_2) if for each $x \neq y$, there exist disjoint open sets $U, V \in \mathcal{T}$, such that $x \in U$ and $y \in V$;

3. **regular** (or T_3) if it is T_1 and for each closed set $A \subset X$ and $x \in A^c$ there exist disjoint open sets U, V with $x \in U$ and $A \subset V$;

4. **normal** (or T_4) if it is T_1, and for any disjoint closed sets $A, B \subset X$ there are disjoint open sets $U, V \subset X$ so that $A \subset U$, and $B \subset V$.

We leave it to the reader to check the implications $T_4 \implies T_3 \implies T_2 \implies T_1$.

3.1.1 Exercises

Exercise 3.1.2. Show that a metric space X is separable, if and only if X is second countable.

Exercise 3.1.3. Show that in a first countable space, singletons $\{x\}$ are G_δ.

Exercise 3.1.4. Prove that every metric space is normal and first countable.

Exercise 3.1.5. Prove that a metric space is separable if and only if it is second countable.

Exercise 3.1.6. Let $X = \mathbb{R}$ and let $\mathcal{T}$ be the family of all sets of the form $U \cup (V \cap \mathbb{Q})$ where U and V are open sets in the usual sense. Show that $\mathcal{T}$ gives a topology on $\mathbb{R}$ which is Hausdorff but not regular.

Exercise 3.1.7. Suppose (X, d) is a metric space. Show that closed subsets of X are G_δ.

Exercise 3.1.8. Let (X, d) be a metric space and consider the bounded metric $d'(x, y) = \frac{d(x,y)}{1+d(x,y)}$. Show that (X, d') describes the same topology on X.

Exercise 3.1.9 (Fréchet)**.** Let $A, B \subset \mathbb{R}$ be two countable dense sets, show that there is a homeomorphism $\theta : \mathbb{R} \to \mathbb{R}$ so that $\theta(A) = B$. Hint: Use Exercise 1.1.27.

3.2 Continuous maps

Let $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ be two topological spaces. A map $f : X \to Y$ is **continuous** if $f^{-1}(U)$ is open for any open set $U \subset Y$. Note that this agrees with our terminology for metric spaces. We say that f is **open** if $F(U)$ is open for all U open. We say that f is a **homeomorphism** if it is bijective, continuous, and open.

Proposition 3.2.1. *Suppose $\mathcal{E}$ generates the topology on Y, then $f : X \to Y$ is continuous if and only if $f^{-1}(U)$ is open for each $U \in \mathcal{E}$.*

Proof. If f is continuous then we trivially have that $f^{-1}(U)$ is open for each $U \in \mathcal{E}$. conversely, if $f^{-1}(U)$ is open for each $U \in \mathcal{E}$, then as the inverse image of a function distributes over unions and intersections it follows that $f^{-1}(O)$ is open whenever O is a union of finite intersections of sets in $\mathcal{E}$. By Proposition 3.1.1 every open set is of this form and hence f is continuous. ∎

A directed set is a set A, together with a binary relation $\leq$ such that

1. $x \leq x$ for all $x \in A$.
2. if $x \leq y$ and $y \leq z$ then $x \leq z$.
3. for each $x, y \in A$ there exists some $z \in A$ so that $x \leq z$ and $y \leq z$.

A net in X is a function $f : A \to X$ from a nonempty directed set A into X. We usually prefer to think of a net as being indexed by A and so we write this as $\{f_\alpha\}_{\alpha\in A}$, and we sometimes abuse notation by identifying a net with its image, so that we might say "let $\{x_\alpha\}_{\alpha\in A} \subset X$ be a net". Note that sequences are just nets when the index set is $\mathbb{N}$, nets were first introduced by Moore and Smith in 1922 as a generalization of sequences. A net $\{x_\alpha\}_{\alpha\in A} \subset X$ has a limit $x \in X$ if for every open neighborhood U of x there exists $\alpha \in A$ so that $x_\beta \in U$ for all $\beta \geq \alpha$. If x is a limit of a net $\{x_\alpha\}_{\alpha\in A}$ then we say that this net is convergent and we write $\lim_{\alpha\to\infty} x_\alpha = x$. In general, limits need not be unique, however, it's easy to see that if X is Hausdorff then limits must be unique.

Proposition 3.2.2. *Suppose X and Y are topological spaces and $f : X \to Y$, then f is continuous if and only if for any convergent net $\{x_\alpha\}_{\alpha\in A}$ such that $x = \lim_{\alpha\to\infty} x_\alpha$, we have that $\{f(x_\alpha)\}_{\alpha\in A}$ is also convergent and $\lim_{\alpha\to\infty} f(x_\alpha) = f(x)$. Moreover, if X is first countable, then one may consider only sequences rather than nets.*

Proof. Suppose first that f is continuous and $\{x_\alpha\}_{\alpha\in A}$ is a net such that $x = \lim_{\alpha\to\infty} x_\alpha$. If we fix an open neighborhood U of $f(x)$ then $f^{-1}(U)$ is an open neighborhood of x and since $x = \lim_{\alpha\to\infty} x_\alpha$ there then exists $a \in A$ so that $x_\beta \in f^{-1}(U)$ for all $\beta \geq a$. Therefore, $f(x_\beta) \in U$ for all $\beta \geq a$ and hence $\lim_{\alpha\to\infty} f(x_\alpha) = f(x)$.

Conversely, suppose that for any net $\{x_\alpha\}_{\alpha\in A}$ such that $x = \lim_{\alpha\to\infty} x_\alpha$, we have that $\{f(x_\alpha)\}_{\alpha\in A}$ is also convergent and $\lim_{\alpha\to\infty} f(x_\alpha) = f(x)$.

Conversely, suppose that f is not continuous and let U be an open set in Y such that $f^{-1}(U)$ is not open in X. Therefore there exists a point $x \in f^{-1}(U)$ so that $f^{-1}(U)$ contains no open neighborhood of x. We let A denote the set of open neighborhoods of x, and note that this is a directed set when ordered by reverse inclusion. For each $O \in A$ we take $x_O \in O \setminus f^{-1}(U)$. Then we have $\lim_{O\to\infty} x_O = x$. However, $f(x_O) \notin U$ for each $O \in A$ and hence $\{f(x_O)\}_{O\in A}$ does not converge to $f(x)$. ∎

If X is a set and $\{f_i : X \to Y_i\}_{i\in I}$ is a family of maps from X into topological spaces Y_i then there is a unique weakest topology on X making each of the maps f_i continuous. We call this topology the weak topology on X generated by $\{f_i\}_{i\in I}$. For example if $\{X_i\}_{i\in I}$ is a family of topological spaces then we may endow $\prod_{i\in I} X_i$ with the weak topology generated by the coordinate maps $\pi_i : \prod_{j\in I} X_j \to X_i$, we always consider $\prod_{i\in I} X_i$ with this topology unless otherwise stated.

Proposition 3.2.3. *If X_i is Hausdorff for each $i \in I$ then $\prod_{i\in I} X_i$ is Hausdorff.*

Proof. Suppose $x, y \in \prod_{i\in I} X_i$ such that $x \neq y$. Then for some coordinate $i \in I$ we have $\pi_i(x) \neq \pi_i(y)$. Since X_i is Hausdorff there exists disjoint open neighborhoods O and U of $\pi_i(x)$ and $\pi_i(y)$ respectively, then $\pi_i^{-1}(O)$ and $\pi_i^{-1}(U)$ give disjoint open neighborhoods of x and y respectively. ∎

Proposition 3.2.4. *If Y is a topological space and $f : Y \to \prod_{i\in I} X_i$, then f is continuous if and only if $\pi_i \circ f$ is continuous for each $i \in I$.*

Proof. Since composition of continuous maps are continuous we see that if f is continuous then $\pi_i \circ f$ is continuous for each $i \in I$. Conversely, suppose $\pi_i \circ f$ is continuous for each $i \in I$. Then for any $i \in I$ and open set $O \subset X_i$ we have that $f^{-1}(\pi_i^{-1}(O))$ is open. Since these sets generate the topology on $\prod_{i \in I} X_i$ we then have that f is continuous. ■

Proposition 3.2.5. *A net* $\{x_\alpha\}_{\alpha \in A}$ *converges to* $x \in \prod_{i \in I} X_i$ *if and only if for each* $i \in I$ *the net* $\{\pi_i(x_\alpha)\}_{\alpha \in A}$ *converges to* $\pi_i(x)$.

Proof. If $\{x_\alpha\}_{\alpha \in A}$ is a net which converges to x. Then for each $i \in I$ we have that $\{\pi_i(x_\alpha)\}_{\alpha \in A}$ converges to $\pi_i(x)$ by Proposition 3.2.2. Conversely, if for each $i \in I$ we have $\{\pi_i(x_\alpha)\}_{\alpha \in A}$ converges to $\pi_i(x)$, then for each $i_1, \ldots, i_n \in I$ and $O_1, \ldots, O_n$ open neighborhoods of $\pi_{i_1}(x), \ldots, \pi_{i_n}(x)$ respectively, we have that there exists $a \in A$ so that $\pi_{i_k}(x_\alpha) \in O_k$ for each $\alpha \geq a$. Since sets of the form $\cap_{k=1}^n \pi_{i_k}^{-1}(O_k)$ form a base for the topology it then follows that $\{x_\alpha\}_{\alpha \in A}$ converges to x. ■

In the case when each X_i is equal to some fixed space X, then we are considering the function space X^I, and the topology we are considering is the **topology of pointwise convergence**; a net $\{f_\alpha\}_{\alpha \in A}$ converges to $f : I \to X$ if and only if for each $i \in I$ we have $\lim_{\alpha \to \infty} f_\alpha(i) = f(i)$.

If X is a topological space then we denote by $C_b(X)$ the space of all continuous functions with bounded image. We consider the **uniform norm** of f as

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

The function $d(f, g) = \|f - g\|_\infty$ gives a metric on $C_b(X)$, which we call the uniform metric.

Proposition 3.2.6. *The space* $C_b(X)$ *is a Banach algebra when endowed with the uniform metric, and pointwise operations.*

Proof. The arguments in Propositions 1.3.1 and 1.3.1 when X is a metric space work equally well here. ■

If $K \subset \mathbb{C}$ is closed, then we denote by $C_b(X; K)$ the subspace of all functions which take values in K. It is easy to see that $C_b(X; K)$ is a closed subspace of $C_b(X)$ in the uniform norm.

Lemma 3.2.7. *Let* X *be a normal space. Suppose that* A *and* B *are disjoint closed sets in* X, *and let* $D = \{k2^{-n} \mid n \geq 1 \text{ and } 0 < k < 2^n\}$ *be the set of dyadic rationals in* $(0, 1)$. *There is a family* $\{U_r \mid r \in D\}$ *of open sets in* X *such that* $A \subset U_r \subset B^c$ *for all* $r \in D$ *and* $\overline{U_r} \subset U_s$ *for* $r < s$.

Proof. Set $U_0 = A$ and $U_1 = W^c$. As X is normal there exist disjoint open sets V and W such that $A \subset V$ and $B \subset W$. We set $U_{1/2} = V$, so that $A \subset U_{1/2} \subset \overline{U_{1/2}} \subset W^c \subset B^c$. We now select U_r for $r = k2^{-n}$ by induction on n. Suppose $N \geq 2$ and we have chosen U_r for $r = k2^{-n}$, for $0 < k < 2^n$, and

$1 \leq n < N$. Then for each $r = (2j+1)2^{-N}$, $0 \leq j < 2^{N-1}$, we have that $\overline{U_{j2^{1-N}}}$ and $(U_{(j+1)2^{1-N}})^c$ are disjoint closed sets and so as above we may choose U_r so that

$$A \subset \overline{U_{j2^{1-N}}} \subset U_r \subset \overline{U_r} \subset U_{(j+1)2^{1-N}} \subset B^c.$$

■

Lemma 3.2.8 (Urysohn's Lemma)**.** *Let X be a normal space. If A and B are disjoint closed sets in X, then there exists $f \in C(X; [0,1])$ such that $f_{|A} = 0$, and $f_{|B} = 1$.*

Proof. Let $\{U_r\}_{r \in D}$ be as in the previous lemma. Set $U_1 = X$ and for $x \in X$ define $f(x) = \inf\{r \mid x \in U_r\}$. Since $A \subset U_r$ for $0 < r < 1$ we have $f_{|A} = 0$. We also have $f_{|B} = 1$ and clearly $f(X) \subset [0,1]$, so all that remains is to show that f is continuous. Towards this end note that $f(x) < a$ if and only if $x \in U_r$ for some $r < \alpha$, thus $f^{-1}((-\infty, a)) = \cup_{r<\alpha} U_r$ is open. Also, $f(x) > a$ if and only if $x \notin U_r$ for some $r > a$, and hence if and only if $x \notin \overline{U_s}$ for some $s > a$ (since $\overline{U_s} \subset U_r$ for $s < r$). Thus $f^{-1}((a, \infty)) = \cup_{s>\alpha}(\overline{U_x})^c$ is open. Since half-lines generate the topology on $\mathbb{R}$ it follows that f is continuous. ■

A space X is **Tychonoff** (or $T_{3^1/_2}$) if it is T_1 and for every closed set A and point $x \in A^c$, there exists a continuous function $f \in C(X; [0,1])$ so that $f_{|A} = 0$, and $f(x) = 1$. Note that by Urysohn's lemma we have that $T_4 \implies T_{3^1/_2}$. It's also easy to check that $T_{3^1/_2} \implies T_3$.

Theorem 3.2.9 (The Tietze Extension Theorem)**.** *Let X be a normal space. If A is a closed subset of X and $f : A \to \mathbb{R}$ is continuous, then there exists $F : X \to \mathbb{R}$ continuous, such that $F_{|A} = f$. Moreover, if f is bounded then we may choose F so that $\|F\|_\infty = \|f\|_\infty$.*

Proof. We first consider the case when f is bounded. We may assume f is nonconstant. If we set $a = \inf f(A)$ and $b = \inf f(A)$, then by replacing f with $(f-a)/(b-a)$ we may assume that $f : X \to [0,1]$. We will inductively construct a sequence of continuous functions $g_n : X \to [0,1]$ so that

$$f(x) - \sum_{i=0}^{n} \frac{2^i}{3^{i+1}} g_i(x) \in [0, (2/3)^{n+1}]$$

for all $x \in A$. Then Proposition 3.2.6 shows that $F = \sum_{i=1}^{\infty} g_i$ defines a continuous function with $\|F\| \leq 1$, such that F agrees with f on A.

To construct g_0 we set $E = f^{-1}([0, 1/3])$ and $F = f^{-1}([2/3, 1])$. Then E and F are disjoint closed sets and so by Urysohn's lemma there exists $g_0 : X \to [0,1]$ continuous so that $g_0(x) = 0$ for $x \in E$ and $g_0(x) = 1$ for $x \in F$. We then have $f(x) - \frac{1}{3} g_0(x) \in [0, 2/3]$ for all $x \in A$. Now suppose $g_0, \ldots, g_{n-1}$ have been constructed so that

$$\tilde{f}(x) = f(x) - \sum_{i=0}^{n-1} \frac{2^i}{3^{i+1}} g_i(x) \in [0, (2/3)^n]$$

for all $x \in A$. Then as above we set $E = \tilde{f}^{-1}([0, 2^n/3^{n+1}])$ and set $F = \tilde{f}^{-1}([2^{n+1}/3^{n+1}, (2/3)^n]$, and we take $g_n : X \to [0,1]$ continuous so that $g_n(x) = 0$ if $x \in E$ and $g_n(x) = 1$ if $x \in F$. Then, we have

$$f(x) - \sum_{i=0}^{n} \frac{2^i}{3^{i+1}} g_i(x) = \tilde{f}(x) - 2^n/3^{n+1} g_n(x) \in [0, (2/3)^{n+1}],$$

finishing the induction step.

For the case when f is not bounded we take a homeomorphism $\theta : \mathbb{R} \to (-1/2, 1/2)$ and consider $\theta \circ f : A \to (-1/2, 1/2)$. Then from above there exists a continuous function $F : X \to [-1/2, 1/2]$ so that F agrees with $\theta \circ f$ on A. We let $E = F^{-1}(\{-1/2, 1/2\})$. Then E is a closed set which is disjoint from A and so by Urysohn's lemma there exists $g : X \to [0,1]$ so that $g_{|E} = 0$ and $g_{|A} = 1$. Then $\tilde{F} = gF$ also agrees with $\theta \circ f$ on A and sastisfies $\tilde{F} : X \to (-1/2, 1/2)$. Therefore $\theta^{-1} \circ \tilde{F}$ gives the desired continuous function. ■

Corollary 3.2.10. *If X is normal, $A \subset X$ is closed, and $f \in C_b(A)$, then there exists $F \in C_b(X)$ such that $F_{|A} = f$, and $\|F\|_\infty = \|f\|_\infty$.*

Proof. We assume $f \neq 0$. By considering the real and imaginary parts separately, it then follows from the previous theorem that there exists a bounded continuous function F_0 such that F_0 agrees with f on A. We let $E = \{x \in X \mid |F_0(x)| \geq \|f\|_\infty\}$. Then

$$h(x) = \begin{cases} \|f\|_\infty / |F_0(x)| & \text{if } x \in E; \\ 1 & \text{if } x \notin E; \end{cases}$$

gives a continuous function and hF_0 agrees with f on A and satisfies $\|hF_0\|_\infty = \|f\|_\infty$. ■

3.2.1 Exercises

A topological space is **disconnected** if there exists nonempty disjoint open sets U, V which cover X; otherwise X is connected. A subset $E \subset X$ is connected or disconnected if this is the case in the relative topology.

Exercise 3.2.11. (a) Show that if $\{A_i\}_{i \in I}$ is a family of connected subsets such that $\cap_{i \in I} A_i \neq \emptyset$ then $\cup_{i \in I} A_i$ is connected.

(b) Show that if $A \subset X$ is connected then $\overline{A}$ is also connected.

(c) Show that every point $x \in X$ is contained in a unique maximal connected subset of X, and this subset is closed. (This is the **connected component** of x).

A topological space is **totally disconnected** if $\{x\}$ is the connected component of x, for each $x \in X$.

Exercise 3.2.12. Show that the continuous image of a connected set is connected.

A topological space $(X, \mathcal{T})$ is **arc-connected** if for each $x, y \in X$ there exists a continuous function $f : [0, 1] \to X$ so that $f(0) = x$ and $f(1) = y$.

Exercise 3.2.13. Show that arc-connected spaces are connected. Also, find an example of a connected space which is not arc-connected.

3.3 Compact spaces

Generalizing the case for metric spaces, a topological space $(X, \mathcal{T})$ is **compact** if every open cover has a finite subcover. A subset $E \subset X$ is compact if it is compact with respect to the relative topology. We say a subset $E \subset X$ is **precompact** if $\overline{E}$ is compact.

A family of subsets $\mathcal{F}$ of X has the **finite intersection property** if for any $F_1, \ldots, F_n \in \mathcal{F}$, with $n \geq 1$ we have $\cap_{i=1}^n F_i \neq \emptyset$.

Proposition 3.3.1. *A topological space X is compact if and only if, whenever $\mathcal{F}$ is a non-empty family of closed subsets which has the finite intersection property then we have $\cap_{F \in \mathcal{F}} F \neq \emptyset$.*

Proof. By contraposition a space X is compact if and only if whenever $\mathcal{G}$ is a family of open sets which does not have a finite subfamily covering X, then $\mathcal{G}$ itself does not cover X. Taking complements of the sets in $\mathcal{G}$, then gives the criterion for compactness above. ∎

Proposition 3.3.2. *Suppose X and Y are topological spaces with X compact. If $f : X \to Y$ is continuous then $f(X)$ is compact.*

Proof. Suppose $\mathcal{G}$ is an open cover for $f(X)$, then $\{f^{-1}(O) \mid O \in \mathcal{G}\}$ is an open cover for X and hence has a finite subcover $f^{-1}(O_1), \ldots, f^{-1}(O_n)$. It then follows that $O_1, \ldots, O_n$ is a finite subcover of $f(X)$. Hence, $f(X)$ is compact. ∎

By the previous proposition, any continuous map from a compact space to $\mathbb{C}$ is bounded, thus for compact spaces we write $C(X)$ for $C_b(X)$.

Proposition 3.3.3. *A closed subset of a compact space is compact. Also, a compact subset of a Hausdorff space is closed.*

Proof. First, suppose X is compact, and $F \subset X$ is closed, if $\mathcal{G}$ is an open cover of F, then $\mathcal{G} \cup \{F^c\}$ is an open cover for X and hence by compactness there is a finite subcover $\mathcal{G}_0$, then $\mathcal{G}_0 \setminus \{F^c\}$ is a finite subcover of $\mathcal{G}$ which covers F, showing that F is compact.

Next, suppose X is Hausdorff and $F \subset X$ is compact. Fix $x_0 \notin F$. Since X is Hausdorff, for each $x \in X$ there exist disjoint open neighborhoods O_x, and G_x of x and x_0 respectively. We have that $\{O_x\}_{x \in F}$ covers F and so by compactness there is a finite subcover $\{O_{x_1}, \ldots, O_{x_n}\}$. If we set $U = \cap_{i=1}^n G_{x_i}$ then we have that U is open and is disjoint from $F \subset \cup_{i=1}^n O_{x_i}$. Thus, $x \notin \overline{F}$, and as x was arbitrary it then follows that $F = \overline{F}$ is closed. ■

Corollary 3.3.4. *Suppose X and Y are topological spaces with X compact and Y Hausdorff, and suppose $f : X \to Y$ is a continuous bijection. Then f is a homeomorphism.*

Proof. Suppose $O \subset X$ is open. By Proposition 3.3.3 we have that O^c is compact. By Proposition 3.3.2 we then have that $f(O^c)$ is compact and hence closed. Since f is a bijection we then have that $f(O) = f(O^c)^c$ is open. Thus, f^{-1} is continuous and hence f is a homeomorphism. ■

Proposition 3.3.5. *Suppose X is a Hausdorff topological space, and $E, F \subset X$ are disjoint compact subsets, then there exist disjoint open sets $U, V \subset X$ so that $E \subset U$ and $F \subset V$.*

Proof. We first consider the case when E is a singleton $E = \{x\}$. Since X is Hausdorff, for each $y \in F$ there exists disjoint open sets O_y and V_y so that $y \in O_y$ and $x \in V_y$. We then have that $\{O_y\}_{y \in F}$ forms an open cover of F and by compactness there exists a finite subcover $\{O_{y_i}\}_{i=1}^n$. If we set $U = \cup_{i=1}^n O_{y_i}$ and $V = \cap_{i=1}^n V_{y_i}$, then U and V are disjoint open sets such that $F \subset U$ and $x \in V$.

We now consider the general case. From above, for each $y \in E$ there exist disjoint open sets O_y and V_y so that $y \in O_y$ and $F \subset V_y$. Again by compactness there exists a finite collection $\{O_{y_i}\}_{i=1}^n$ which covers E. Then $U = \cup_{i=1}^n O_{y_i}$ and $V = \cap_{i=1}^n V_{y_i}$ are disjoint open sets and we have $E \subset U$, while $F \subset V$. ■

Corollary 3.3.6. *A compact Hausdorff space X is normal.*

Proof. This follows directly from Propositions 3.3.3 and 3.3.5. ■

Theorem 3.3.7 (Tychonoff's Theorem)**.** *If $\{X_i\}_{i \in I}$ is a family of compact topological spaces, then $\prod_{i \in I} X_i$ is also compact.*

Proof. Suppose $\mathcal{F}$ is a family of closed subsets of $\prod_{i \in I} X_i$ with the finite intersection property. By Zorn's lemma there exists a maximal family $\mathcal{E}$ of (not necessarily closed) subsets with the finite intersection property such that $\mathcal{F} \subset \mathcal{E}$. Note that $\mathcal{E}$ itself must then be closed under finite intersections, and if $E \in \mathcal{E}$ and $E \subset F$, then $F \in \mathcal{E}$.

For each $i \in I$ the family $\{\overline{\pi_i(E)} \mid E \in \mathcal{E}\}$ has the finite intersection property and hence by compactness we have $\cap_{E \in \mathcal{E}} \overline{\pi_i(E)} \neq \emptyset$. Take x_i a point in this intersection. We let x be the point in $\prod_{i \in I} X_i$ whose ith coordinate is x_i.

We claim that all neighborhoods of x are contained in $\mathcal{E}$. To prove this it is enough to show that the neighborhoods of the form $\cap_{k=1}^n \pi_{i_k}^{-1}(E_{i_k})$ are contained in $\mathcal{E}$, and since $\mathcal{E}$ is closed under finite intersections it is then enough to show that neighborhoods of the form $\pi_i^{-1}(E_i)$ are contained in $\mathcal{E}$. To see this note that since $x_i \in \overline{\pi_i(E)}$ for any $E \in \mathcal{E}$ it follows that $\pi_i^{-1}(E_i) \cap E \neq \emptyset$ for all

$E \in \mathcal{E}$. This then shows that $\mathcal{E} \cup \{\pi_i^{-1}(E_i)\}$ has the finite intersection property and hence $\pi_i^{-1}(E_i) \in \mathcal{E}$ by maximality of $\mathcal{E}$. Thus, arbitrary neighborhoods of x are contained in $\mathcal{E}$ and hence have non-trivial intersection with an arbitrary set $E \in \mathcal{E}$. Thus, $x \in \overline{E}$ for all $E \in \mathcal{E}$, and hence $x \in \cap_{E\in\mathcal{E}}\overline{E} \subset \cap_{F\in\mathcal{F}}F$, showing that $\prod_{i\in I} X_i$ is compact. ■

If X is a Banach space, then the weak*-topology on X^* is defined to be the coarsest topology so that the maps $X^* \ni \varphi \mapsto \varphi(x)$ are continuous for each $x \in X$.

Theorem 3.3.8 (The Banach-Alaoglu theorem)**.** *Let X be a Banach space. Then the closed unit ball in X^* is compact in the weak*-topology.*

Proof. Let $D = \prod_{x\in X} \overline{B}(\|x\|, 0)$. Since closed balls in Euclidean space are compact, it then follows from Tychonoff's theorem that D is compact. We let K denote the closed unit ball in X^* and consider the map $\pi : K \to D$ where $\pi(\varphi)$ has coordinates $\pi(\varphi)_x = \varphi(x)$. Note that since φ is in the unit ball we have $|\pi(\varphi_x)| = |\varphi(x)| \leq \|x\|$, so that π is well defined. Also note that π is injective since if $\pi(\varphi) = \pi(\psi)$ then for each point $x \in X$ we have $\varphi(x) = \psi(x)$.

If $\{\varphi_\alpha\}_{\alpha\in A}$ is a net in K then $\{\varphi_\alpha\}_{\alpha\in A}$ converges to φ if and only if for each $x \in X$ we have $\varphi_\alpha(x) \to \varphi(x)$, and this is also if and only if $\pi(\varphi_\alpha) \to \pi(\varphi)$ in D. Therefore π defines a homeomorphism from D onto its image. K is therefore compact if and only if the image of π is closed.

Note that D consists of functions from X to $\mathbb{K}$ (where $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$), and as these functions take x to an element in the ball $\overline{B}(\|x\| < x)$ it follows that they are bounded functions. Thus, the image of π consists of those functions which are linear, i.e.

$$\pi(K) = \cap_{x_1,x_2\in X,\alpha\in\mathbb{K}}\{f \in D \mid f_{x_1+\alpha x_2} = f_{x_1} + \alpha f_{x_2}\}.$$

As an intersection of closed sets is closed it then follows that $\pi(K)$ is closed. ■

A property $\mathcal{P}$ of topological spaces is said to hold **locally** for a space X if each $x \in X$ has a neighborhood which satisfies $\mathcal{P}$. For example a locally compact space X is one in which each point $x \in X$ has a compact neighborhood. Euclidean spaces $\mathbb{R}^n$ are examples of locally compact spaces.

A subset $\mathcal{F} \subset C_b(X)$ is **equicontinuous at** $x \in X$ if for each $\varepsilon > 0$ there is a neighborhood U of x such that $|f(y) - f(x)| < \varepsilon$ for all $u \in U$ and $f \in \mathcal{F}$. $\mathcal{F}$ is **equicontinuous** if it is equicontinuous at each point. Also, $\mathcal{F}$ is **pointwise bounded** if $\{f(x) \mid f \in \mathcal{F}\}$ is bounded for each $x \in X$.

Theorem 3.3.9 (The Arzelà-Ascoli Theorem)**.** *Let X be a compact Hausdorff space. If $\mathcal{F} \subset C(X)$ is equicontinuous and pointwise bounded, then $\mathcal{F}$ is totally bounded in the uniform metric, and $\mathcal{F}$ is precompact.*

Proof. Fix $\varepsilon > 0$. Since $\mathcal{F}$ is equicontinuous, for each $x \in X$ there exists an open neighborhood O_x of x such that $|f(y) - f(x)| < \varepsilon/4$ for all $f \in \mathcal{F}$, and $y \in O_x$. The family $\{O_x\}_{x\in X}$ is an open cover, and since X is compact it is covered by a finite collection $O_{x_1}, \ldots, O_{x_n}$.

Set $K = \sup_{f \in \mathcal{F}, 1 \leq k \leq n} |f(x_k)|$. Since $\mathcal{F}$ is pointwise bounded we have $K < \infty$. We cover the closed ball $\overline{B(K,0)} \subset \mathbb{C}$ with finitely many $\varepsilon/4$ balls $B(\varepsilon/4, z_1), \ldots, B(\varepsilon/4, z_m)$.

We consider the finite set

$$F = \{\phi : \{1, \ldots, n\} \to \{1, \ldots, m\} \mid \text{there exists } f \in \mathcal{F} \text{ such that } f(x_i) \in B(\varepsilon/4, z_{\phi(i)}) \text{ for all } 1 \leq i \leq n\},$$

and for each $\phi \in F$, we choose $f_\phi \in \mathcal{F}$ which realizes the fact that $\phi \in F$.

If $f \in \mathcal{F}$, then as $B(\varepsilon/4, z_1), \ldots, B(\varepsilon/4, z_m)$ cover $\overline{B(K,0)}$, there exists a function $\phi \in F$ so that $f(x_i) \in B(\varepsilon/4, z_{\phi(i)})$, for $1 \leq i \leq n$. If $y \in O_{x_i}$, we then have

$$\begin{aligned} |f(y) - f_\phi(y)| &\leq |f(y) - f(x_i)| + |f(x_i) - f_\phi(x_i)| + |f_\phi(x_i) - f_\phi(y)| \\ &< \varepsilon/4 + \varepsilon/2 + \varepsilon/4 = \varepsilon. \end{aligned}$$

Thus $\|f - f_\phi\|_\infty < \varepsilon$, showing that $\mathcal{F}$ is covered by finitely many ε-balls, and is therefore totally bounded. It then follows that $\mathcal{F}$ is precompact in the uniform norm by the Heine-Borel property. ■

A topological space is **σ-compact** if it is a countable union of compact subsets.

Lemma 3.3.10. *Suppose X is a σ-compact, locally compact Hausdorff space. Then there is a sequence $\{U_n\}_{n=1}^\infty$ of precompact open sets such that $U_n \subset U_{n+1}$ for each $1 \leq n < \infty$, and $\cup_{n=1}^\infty U_n = X$.*

Proof. We have $X = \cup_{n=1}^\infty F_n$ where F_n are compact sets. Since X is locally compact each point $x \in F_n$ has a precompact open neighborhood O_x, then $\{O_x\}_{x \in F_n}$ covers F_n and hence has a finite subcover $O_{x_1}, \ldots, O_{x_k}$. Setting $V_n = \cup_{i=1}^k O_{x_i}$, we then have that V_n is open, precompact, and $F_n \subset V_n$ for each $1 \leq n < \infty$. Setting $U_n = \cup_{m=1}^n V_n$ then produces the desired sequence. ■

Note that if $\{U_n\}_{n=1}^\infty$ are as in the previous lemma and $F \subset X$ is compact, then $\{U_n\}_{n=1}^\infty$ covers F and hence by compactness there exists a finite subcover. However, since $\{U_n\}_{n=1}^\infty$ is an increasing union it then follows that $F \subset U_n$ for some $1 \leq n < \infty$.

Theorem 3.3.11. *Let X be a σ-compact, locally compact Hausdorff space. If $\{f_n\}_{n=1}^\infty$ is a sequence which is equicontinuous and pointwise bounded, then there exists a continuous function $f : X \to \mathbb{C}$ and a subsequence of $\{f_n\}_{n=1}^\infty$ which converges to f uniformly on compact sets.*

Proof. We write $X = \cup_{n=1}^\infty F_n$ where each $F_n = \overline{U_n}$ as in the previous lemma. Set $F_0 = \emptyset$, $g_0 = \emptyset$, and $\{f_n^0\}_{n=1}^\infty = \{f_n\}_{n=1}^\infty$. For $n \geq 1$ we inductively choose $g_k : F_k \to \mathbb{C}$, and subsequences $\{f_n^k\}_{n=1}^\infty$ of $\{f_n^{k-1}\}_{n=1}^\infty$ so that $g_{k|F_{k-1}} = g_{k-1}$ as follows: We suppose that g_k and $\{f_n^k\}_{n=1}^\infty$ have already been chosen for $0 \leq k < \infty$. Restricting $\{f_n^k\}_{n=1}^\infty$ to F_{k+1} we have an equicontinuous, pointwise bounded family and hence this is precompact in the uniform norm by the Arzelà-Ascoli Theorem. Therefore there exists a subsequence $\{f_n^{k+1}\}_{n=1}^\infty$ such that $\{f_n^{k+1}\}_{n=1}^\infty$ converges uniformly on F_{k+1} to a continuous function $g_{k+1} : F_{k+1} \to \mathbb{C}$. As $\{f_n^{k+1}\}_{n=1}^\infty$ is a subsequence of $\{f_n^k\}_{n=1}^\infty$ we have that $g_{k+1|F_k} = g_k$.

We may then define $g : X \to \mathbb{C}$ by $g(x) = g_k(x)$ for $x \in F_k$. Then g is a well defined continuous function. If we consider the diagonal subsequence $\{f_n^n\}_{n=1}^\infty$ then we have that $\{f_n^n\}_{n=1}^\infty$ converges to g uniformly on F_k for any $1 \leq k < \infty$. Since any compact set is covered by some F_k it follows that $\{f_n^n\}_{n=1}^\infty$ converges to g uniformly on compact sets. ■

3.3.1 Exercises

Exercise 3.3.12. Show that a metric space X is compact if and only if every continuous real valued function on X is bounded.

Exercise 3.3.13. Let $(X, \mathcal{T})$ be a compact Hausdorff space. Show that if $\mathcal{T}'$ is any weaker topology then $(X, \mathcal{T}')$ is not Hausdorff. Show that if $\mathcal{T}'$ is any stronger topology then $(X, \mathcal{T}')$ is not compact.

Let X be a locally compact topological space, and fix a point $\omega \notin X$. On the space $\tilde{X} = X \cup \{\omega\}$ we define a new topology whose open sets consist of the open sets in X, together with the compliments (in $\tilde{X}$) of compact subsets of X. The space $\tilde{X}$ is called the **one-point compactification** of X.

Exercise 3.3.14. Show that $\tilde{X}$ is a compact Hausdorff space and that the relative topology from $X \subset \tilde{X}$ agrees with the topology on X.

Let X be a locally compact topological space, a function $f : X \to \mathbb{C}$ is said to **vanish at infinity** if for every $\varepsilon > 0$, there exists a compact set $K \subset X$ so that $|f(x)| < \varepsilon$ for all $x \in K^c$. We denote by $C_0(X)$ the space of all continuous functions which vanish at infinity.

Exercise 3.3.15. Show that $C_0(X)$ is a closed subspace of $C_b(X)$.

Exercise 3.3.16 (Compare this with Exercise 3.1.3)**.** Suppose X is a compact Hausdorff space such that singletons $\{x\}$ are G_δ.

1. For each $x \in X$ find a countable open cover $\mathcal{O}$ of $X \setminus \{x\}$ so that $x \notin \overline{O}$ for all $O \in \mathcal{O}$.
2. Show that X is first countable.

If (V, E) is a graph, and $k \in \mathbb{N}$, a k**-coloring** of the graph (V, E) is an assignment $f \in \{1, 2, \ldots, k\}^V$ such that for all $(v, w) \in E$ we have $f(v) \neq f(w)$.

Exercise 3.3.17. Prove the De Bruijn-Erdös theorem: If (V, E) is a graph such that a k-coloring exists for every finite subgraph, then a k-coloring exists for (V, E). Hint: For each finite subgraph (V_0, E_0) consider

$$F_{(V_0,E_0)} = \{f \in \{1, \ldots, k\}^V \mid f_{|V_0} \text{ gives a } k - \text{coloring of } (V_0, E_0)\},$$

then show that the family of all such $F_{(V_0,E_0)}$ has the finite intersection property. (This approach is due to Gottschalk.)

3.4 The Stone-Weierstrass Theorem

A subset $A \subset C_b(X)$ (resp. $C_b(X;\mathbb{R})$) is an **algebra** if it is a complex (resp. real) vector subspace such that $fg \in A$ for all $f, g \in A$. A is said to **separate points** if for all $x \neq y$ there exists $f \in A$ such that $f(x) \neq f(y)$. A subset $A \subset C_b(X;\mathbb{R})$ is a **lattice** if it is a real vector subspace such that $f \vee g = \max\{f, g\} \in A$, and $f \wedge g = \min\{f, g\} \in A$ for all $f, g \in A$.

Lemma 3.4.1. *For any $\varepsilon > 0$ there is a polynomial p on $\mathbb{R}$ such that $p(0) = 0$ and $||x| - p(x)| < \varepsilon$ for $x \in [-1, 1]$.*

Proof. Consider the Maclaurin series $1 - \sum_{n=1}^{\infty} c_n t^n$ for $(1-t)^{1/2}$. This series converges absolutely and uniformly on $[-1, 1]$ and its sum is $(1-t)^{1/2}$. Therefore, given any $\varepsilon > 0$ we may take a suitable partial sum to obtain a polynomial q so that $|(1-t)^{1/2} - q(t)| < \varepsilon/2$ for $t \in [-1, 1]$. Setting $r(x) = q(1-x^2)$, we then obtain a polynomial r such that $||x| - r(x)| < \varepsilon/2$ for $x \in [-1, 1]$. If we set $p(x) = r(x) - r(0)$, then p is a polynomial such that $p(0) = 0$ and $||x| - p(x)| < \varepsilon$ for all $x \in [-1, 1]$. ∎

Proposition 3.4.2. *Let X be a topological space. If $A \subset C_b(X;\mathbb{R})$ is a closed subalgebra, then A is a lattice.*

Proof. As $f \vee g = \frac{1}{2}(f + g + |f - g|)$ and $f \wedge g = \frac{1}{2}(f + g - |f - g|)$ it is enough to show that $|f| \in A$ whenever $f \in A$. Suppose $f \in A$ and $\varepsilon > 0$ is given. We may assume $f \neq 0$. Since $f/\|f\|_\infty$ maps into $[-1, 1]$ it follows from the previous lemma that there exists a polynomial p on $\mathbb{R}$ so that $p(0) = 0$ and $|(p \circ f)/\|f\|_\infty - |f|/\|f\|_\infty| < \varepsilon$. Since $p(0) = 0$ it follows that p has 0 for its constant coefficient, thus since A is an algebra we have $p \circ f \in A$, and since $\varepsilon > 0$ was arbitrary it then follows that $|f|/\|f\|_\infty \in A$ and hence also $|f| \in A$. ∎

Lemma 3.4.3. *Let X be a compact space. Suppose $A \subset C_b(X;\mathbb{R})$ is a closed lattice and $f \in C(X;\mathbb{R})$. If for every $x, y \in X$ there exists $g \in A$ so that $g(x) = f(x)$ and $g(y) = f(y)$, then $f \in A$.*

Proof. Fix $\varepsilon > 0$. For each x, y take $g_{x,y} \in A$ so that $g_{x,y}(x) = f(x)$ and $g_{x,y}(y) = f(y)$. Let $U_{x,y} = \{z \in X \mid f(z) < g_{x,y}(z) + \varepsilon\}$. Fix y; then $\{U_{x,y}\}_{x \in X}$ is an open cover of X and so there is a finite subcover $\{U_{x_i,y}\}_{i=1}^n$. Set $g_y = \max\{g_{x_1,y}, \ldots, g_{x_n,y}\} \in A$. Then $f < g_y + \varepsilon$ on X and $f(y) = g_y(y)$ so that in some neighborhood V_y of y we have that $f > g_y - \varepsilon$. We then have that $\{V_y\}_{y \in X}$ covers X and so there is a finite subcover $\{V_{y_j}\}_{j=1}^m$. Set $g = \min\{g_{y_1}, \ldots, g_{y_m}\} \in A$. Then $\|f - g\|_\infty < \varepsilon$, $g \in A$, and since $\varepsilon > 0$ was arbitrary we then have $f \in A$. ∎

Theorem 3.4.4 (The Stone-Weierstrass Theorem)**.** *Let X be a compact Hausdorff space. If $A \subset C(X;\mathbb{R})$ is a closed algebra which separates points then either $A = C(X;\mathbb{R})$ or else there exists $x_0 \in X$ such that $A = \{f \in C(X;\mathbb{R}) \mid f(x_0) = 0\}$.*

Proof. We first consider the case when X is a two point set $\{x, y\}$, so that $C(\{x, y\})$ is two dimensional. If A is two dimensional then we are done. Also, since A separates points we have a function $f \in A$ with $f(x) \neq f(y)$, so that we may assume A is one dimensional. Then $f^2 = cf$ for some $c \in \mathbb{R}$, so that $f(x)$ and $f(y)$ distinct roots of the polynomial $t^2 - ct$. It then follows that either $f(x) = 0$ in which case $A = \{g \in C(\{x, y\}) \mid g(x) = 0\}$, or else $f(y) = 0$ in which case $A = \{g \in C(\{x, y\}) \mid g(y) = 0\}$.

We now consider the general case. Suppose $x, y \in X$, with $x \neq y$. Considering the restriction map from A we obtain an algebra $A_{x,y} \subset C(\{x, y\}; \mathbb{R})$. Note that since A separates points so does $A_{x,y}$. If there exists $x_0 \in X$ so that $f(x_0) = 0$ for every $f \in A$, then as A separates points there can be at most one such x_0 and from above we then have $A_{x,y} = C(\{x, y\})$ whenever $x_0 \notin \{x, y\}$. It then follows from Proposition 3.4.2 and Lemma 3.4.3 that $A = \{f \in C(X) \mid f(x_0) = 0\}$. Otherwise $A_{x,y} = C(\{x, y\})$ for all $x, y \in X$ in which case it again follows from Proposition 3.4.2 and Lemma 3.4.3 that $A = C(X)$. ■

Corollary 3.4.5. *Let $K \subset \mathbb{R}^n$ be compact, and $f \in C(K; \mathbb{R})$, then for every $\varepsilon > 0$ there exists a polynomial $p : \mathbb{R}^n \to \mathbb{R}$ so that $|f(k) - p(k)| < \varepsilon$ for all $k \in K$.*

Proof. Since the space of polynomials forms an algebra which contains the constant functions and separate points it follows that this space is dense in $C(K; \mathbb{R})$ by the Stone-Weierstrass theorem. ■

Theorem 3.4.6 (The Complex Stone-Weierstrass Theorem)**.** *Let X be a compact Hausdorff space. If $A \subset C(X)$ is a closed algebra which is closed under complex conjugation and separates points then either $A = C(X)$ or else there exists $x_0 \in X$ such that $A = \{f \in C(X; \mathbb{R}) \mid f(x_0) = 0\}$.*

Proof. Since $\operatorname{Re} f = (f + \overline{f})/2$ and $\operatorname{Im} f = (f - \overline{f})/2i$ it follows that the set of real and imaginary parts of functions in A is an algebra $A_{\mathbb{R}}$ in $C(X; \mathbb{R})$. Moreover it is easy to see that this separates points and hence the Stone-Weierstrass theorem applies. Since $A = \{f + ig \mid f, g \in A_{\mathbb{R}}\}$ the complex version then follows. ■

3.4.1 Exercises

Exercise 3.4.7. Suppose X and Y are compact Hausdorff spaces and $f \in C(X \times Y)$. Show that for all $\varepsilon > 0$ there exist $g_1, \ldots, g_n \in C(X)$ and $h_1, \ldots, h_n \in C(Y)$ so that $|f(x, y) - \sum_{i=1}^n g_i(x) h_i(y)| < \varepsilon$ for all $(x, y) \in X \times Y$.

Exercise 3.4.8. Let X be a compact Hausdorff space. An **ideal** in $C(X)$ is a subalgebra $I \subset C(X)$, such that $fg \in I$ whenever $f \in C(X)$ and $g \in I$.

1. If $I \subset C(X)$ is an ideal, let $h(I) = \{x \in X \mid f(x) = 0 \text{ for all } f \in I\}$, the **hull** of I. Show that $h(I)$ is closed.

2. If $A \subset X$, let $k(A) = \{f \in C(X) \mid f(x) = 0 \text{ for all } x \in A\}$, the **kernel** of A. Show that $k(A)$ is a closed ideal in $C(X)$ which is closed under conjugation.

3. Show that $k(h(I)) = \overline{I}$ for any ideal $I \subset C(X)$ which is closed under conjugation, and $h(k(A)) = \overline{A}$ for any subset $A \subset X$.

Given a topological space X and an equivalence relation R on X, we let X/R denote the set of equivalence classes and we consider $q : X \to X/R$ the quotient map $q(x) = [x]$. We endow X/R with the weakest topology so that q is continuous.

Exercise 3.4.9. Let X be a compact Hausdorff space.

1. Show that X/R is Hausdorff if and only if R is a closed subset of $X \times X$.

2. If $R \subset X \times X$ is closed, consider $A_R = \{f \circ q \mid f \in C(X/R)\}$. Show that A_R is a closed subalgebra of $C(X)$ which contains the constant functions and is closed under complex conjugation.

3. Show that $R \mapsto A_R$ gives a bijection between equivalence relations on X which are closed in $X \times X$, and closed subalgebras of $C(X)$ which contain the constant functions and are closed under complex conjugation.

3.5 The Stone-Čech compactification

Let X be a topological space. A map $\chi : C_b(X) \to \mathbb{C}$ is a **homomorphism** if it is linear, and satisfies $\chi(fg) = \chi(f)\chi(g)$ for $f, g \in C_b(X)$, and $\chi(1) = 1$. We denote by $\sigma(C_b(X))$ the space of all such homomorphisms and we endow this with the topology of pointwise convergence inherited from $\mathbb{C}^{C_b(X)}$. Note that if $\varphi, \chi \in \sigma(C_b(X))$ then $\varphi = \chi$ if and only if $\ker(\varphi) = \ker(\chi)$.

Lemma 3.5.1. *$\sigma(C_b(X))$ is compact.*

Proof. Suppose $\varphi \in \sigma(C_b(X))$, and $f \in C_b(X)$ with $\|f\|_\infty \leq 1$. We claim that $|\varphi(f)| \leq 1$. If this were not the case then the function $g(x) = \varphi(f) - f(x)$ would satisfy $|g(x)| \geq |\varphi(f)| - 1$ for each $x \in X$ and hence the function $h(x) = \frac{1}{g(x)}$ would be in $C_b(X)$. However, we would then have $1 = \varphi(1) = \varphi(g)\varphi(h) = 0$ a contradiction.

Also, restricting to $C_b(X)_1 = \{f \in C_b(X) \mid \|f\| \leq 1\}$, gives the same topology of pointwise convergence and hence we may view $\sigma(C_b(X))$ as a subspace of D^B where D is the closed unit disc in $\mathbb{C}$. Since D^B is compact it is then enough to show that $\sigma(C_b(X))$ is a closed subspace.

Suppose therefore that $\{\varphi_\alpha\}_{\alpha \in A}$ is a net of homomorphisms which converge pointwise to a function $\varphi : C_b(X) \to \mathbb{C}$. As addition, scalar multiplication are continuous in $\mathbb{C}$ it then follows that φ is linear, and as multiplication is jointly continuous in $\mathbb{C}$ it follows that $\varphi(fg) = \varphi(f)\varphi(g)$ for all $f, g \in C_b(X)$. Therefore φ is a homomorphism. ∎

Theorem 3.5.2 (Stone). *Let X be a topological space. For each $x \in X$ denote by $\beta_x : C_b(X) \to \mathbb{C}$ the homomorphism given by $\beta_x(f) = f(x)$, then $X \ni x \mapsto \beta_x \in \sigma(C(X))$ is a continuous map with dense image which satisfies the universal property that if $\pi : X \to K$ is any continuous map into a compact Hausdorff space K, then there exists a unique continuous map $\beta_\pi : \sigma(C(X)) \to K$, such that for $x \in X$ we have $\pi(x) = \beta_\pi(\beta_x)$. In particular, if X is a compact Hausdorff space then β is a homeomorphism.*

Proof. If $\{x_i\} \subset X$ is a net such that $x_i \to x$, then for any $f \in C_b(X)$ we have $\beta_{x_i}(f) = f(x_i) \to f(x) = \beta_x(f)$, hence $\beta_{x_i} \to \beta_x$. Thus, $x \mapsto \beta_x$ is continuous. To show that this map has dense image we suppose by way of contradiction that $\varphi \in \sigma(C_b(X))$ is not in the closure of $\beta(X)$, and set $I = \ker(\varphi)$.

If $\psi \in \overline{\beta(X)}$, then there exists $f_\psi \in I$ such that $f_\psi \not\in \ker(\psi)$. Hence, for some $c_\psi > 0$, and an open neighborhood O_ψ of ψ, we have that $|\psi'(f)| > c_\psi$ for all $\psi' \in O_\psi$. As $\overline{\beta(X)}$ is compact we may take a finite subcover of the cover $\{O_\psi\}_{\psi \in \overline{\beta(X)}}$. Thus, we obtain $f_1, \ldots, f_n \in I$, and $c > 0$ such that $\sum_{i=1}^n \psi(|f|^2) > c$ for all $\psi \in \overline{\beta(X)}$. In particular we have $\sum_{i=1}^n |f|^2(x) = \beta_x(\sum_{i=1}^n |f|^2) > c$, for all $x \in X$. Thus, if we consider the function $g(x) = 1/\left(\sum_{i=1}^n |f|^2\right)$, then $g \in C_b(X)$ and we have $fg = 1$. We would then have $1 = \varphi(1) = \varphi(fg) = \varphi(f)\varphi(g) = 0$, a contradiction. Thus, we must have that $\overline{\beta(X)} = \sigma(C_b(X))$.

If X is a compact Hausdorff space then β is surjective since the image is dense and compact. Moreover, β is injective since $C_b(X)$ separates points. Hence, β is a homoeomorphism, being a continuous bijection between compact Hausdorff spaces.

In general, to see that $\beta : X \to \sigma(C_b(X))$ satisfies the above universal property, suppose that K is a compact Hausdorff space and $\pi : X \to K$ is continuous. We then obtain a continuous map $\pi^* : C(K) \to C_b(X)$ given by $\pi^*(f)(x) = f(\pi(x))$. Thus, we obtain the continuous map $\tilde{\pi} : \sigma(C_b(X)) \to \sigma(C(K))$ by $\tilde{\pi}(\varphi)(g) = \varphi(\pi^*(g))$. Since K is compact and Hausdorff we have established above that $\beta^K : K \to \sigma(C_b(K))$ is a homeomorphism. Thus, we obtain a continuous map $\beta_\pi : \sigma(C_b(X)) \to K$ by setting $\beta_\pi = \beta^{K^{-1}} \circ \tilde{\pi}$. If $x \in X$, and $g \in C(K)$ then we compute directly

$$\tilde{\pi}(\beta_x)(\varphi)(g) = \beta_x(\pi^*(g)) = \pi^*(g)(x) = g(\pi(x)) = \beta^K_{\pi(x)}(g).$$

Hence, $\beta_\pi(\beta_x) = \pi(x)$. ■

If X is a topological space, then the **Stone-Čech compactification** of X consists of a compact Hausdorff space βX, together with a continuous map $\beta : X \to \beta X$, which satisfies the universal property given in the previous theorem. If follows easily that, up to homeomorphism, this is uniquely defined by its universal property. The previous theorem shows that βX exists and may be identified with $\sigma(C_b(X))$. The following easy consequence (implicit already in Tychonoff's work) was obtained independently by Čech using different methods:

Corollary 3.5.3 (Stone, Čech). *Let X be a topological space, then $\beta : X \to \beta X$ is a homeomorphism onto its image if and only if X is a Tychonoff space.*

Proof. From the previous theorem we have that $\beta : X \to \beta X$ is continuous. Since X is Tychonoff we have, in particular, that $C_b(X)$ separates points, and it then follows that β is injective. Thus, we just need to show that β is an open map into $\beta(X)$. Suppose that $F \subset X$ is closed and $x \in X \setminus F$. As X is Tychonoff there exists $f : X \to [0,1]$ continuous so that $f_{|F} = 0$, and $f(x) = 1$. Thus, we have $\beta_x(f) = 1$ while $\beta_y(f) = 0$ for all $y \in F$, and hence $\beta_x \notin \overline{\beta(F)}$. Since $x \notin F$ was arbitrary it follows that $\overline{\beta(F)} \cap \beta(X) = \beta(F)$, and hence $\beta(F)$ is closed in $\beta(X)$. As β is injective, taking complements shows that β is an open map into $\beta(X)$. ■

As a subspace of a Tychonoff space is again Tychonoff, and compact Hausdorff spaces are normal and hence Tychonoff by Corollary 3.3.6, the previous corollary gives the following characterization of Tychonoff spaces.

Corollary 3.5.4. *A topological space X is Tychonoff if and only if X is homeomorphic to a subspace of a compact Hausdorff space.*

Considering the one-point compactification gives the following:

Corollary 3.5.5. *Locally compact spaces are Tychonoff.*

Theorem 3.5.6 (The Tietze Extension Theorem for Tychonoff spaces)**.** *Let X be a Tychonoff space, $K \subset U \subset X$, with K compact, and U open. If $f \in C(K)$ then there exists $F \in C_b(X)$, with $\|F\|_\infty = \|f\|_\infty$, such that $F_{|K} = f$, and $F_{|U^c} = 0$.*

Proof. Since X is Tychonoff, the map $\beta : X \to \beta X$ is a homeomorphism onto its image. Thus $\beta(K) \subset \beta X$ is compact, and there exists $V \subset \beta X$ open such that $\beta(K) \subset \beta(U) = V \cap \beta X \subset V$. We consider the function $g : \beta(K) \cup V^c \to \mathbb{C}$ by setting $g(\beta(k)) = f(k)$ for $k \in K$, and $g(x) = 0$ for $x \in V^c$. Since $\beta(U)$ and V^c are disjoint closed sets, and βX is normal, they can be separated so that they are both clopen in the relative topology. Thus, g is continuous and by the Tietze Extension Theorem for compact Hausdorff spaces there is then a continuous function $G \in C(\beta X)$, with $\|G\|_\infty = \|g\|_\infty = \|f\|_\infty$ so that $G_{|\beta(K)} = g$, and $G_{|V^c} = 0$. Taking $F = G \circ \beta$ then gives the desired function. ■

Lemma 3.5.7. *If X is normal and second countable then there exists a countable family $\mathcal{F} \subset C_b(X;[0,1])$ which separates points.*

Proof. Let $\mathcal{E}$ be a countable base for X. For each $U, V \in \mathcal{E}$ such that $\overline{U} \subset V$ we may use Urysohn's lemma to construct a continuous function $f_{U,V} : X \to [0,1]$ so that $f_{|U} = 0$ and $f_{|V^c} = 1$. If we let $\mathcal{F}$ be the collection of all such $f_{U,V}$ and claim that $\mathcal{F}$ separates points. Indeed, if $x, y \in X$ with $x \neq y$, then as X is normal there exist disjoint closed neighborhoods E and F of x and y respectively. Then there must exist $U, V \in \mathcal{E}$ neighborhoods of x and y respectively such that $U \subset E$ and $V \subset F$. We then have that $f_{U,V}(x) = 0$ while $f_{U,V}(y) = 1$. ■

Proposition 3.5.8. *Every second countable normal space is homeomorphic to a subspace of the* **Hilbert cube** $[0,1]^{\mathbb{N}}$

Proof. Suppose X is normal and second countable. Then by Lemma 3.5.7 there is a countable family $\mathcal{F} \subset C_b(X)$ which separates points in $\mathcal{F}$. Consider the evaluation map $e : X \to [0,1]^{\mathcal{F}}$ given by $e(x)(f) = f(x)$. Then this is continuous and since $\mathcal{F}$ separates points it is injective. Since $[0,1]^{\mathcal{F}}$ is compact e extends to a continuous map $\beta_e : \beta X \to [0,1]^{\mathcal{F}}$. If $F \subset X$ is closed then $\overline{F} \subset \beta X$ is compact and satisfies $\overline{F} \cap X = F$. As β_e is continuous we have that $\beta_e(\overline{F})$ is compact, and hence $e(F) = \beta_e(\overline{F}) \cap e(X)$ is closed in $e(X)$. Thus, the map e onto its image preserves closed sets and hence e is a homeomorphism of X onto its image in $[0,1]^{\mathcal{F}}$. ■

Theorem 3.5.9 (The Urysohn Metrization Theorem)**.** *Every second countable normal space is metrizable.*

Proof. The Hilbert cube is metrizable. Indeed, the explicit metric $d(f,g) = \sum_{n=1}^{\infty} 2^{-n}|f(n) - g(n)|$ is easily seen to give the topology on $[0,1]^{\mathbb{N}}$. Since subspaces of metrizable spaces are again metrizable, the result then follows from Proposition 3.5.8 ■

3.5.1 Exercises

Exercise 3.5.10. Let X be a compact Hausdorff space. Show that X is a second countable if and only if $C(X)$ is separable.

Exercise 3.5.11. Suppose that a topological space X has a countable basis of clopen sets, show that X embedds into $\{0,1\}^{\mathbb{N}}$.

Exercise 3.5.12. Let X and Y be compact Hausdorff spaces and suppose $\phi : C(X) \to C(Y)$ is a (unital) homomorphism, i.e., ϕ is complex linear, $\phi(1) = 1$, and $\phi(fg) = \phi(f)\phi(g)$ for all $f, g \in C(X)$. Show that there exists a unique continuous map $\pi : Y \to X$ so that $\phi(f) = f \circ \pi$ for all $f \in C(X)$. Moreover, show that π is bijective if and only if ϕ is bijective.

3.6 The property of Baire

A topological space X is **completely metrizable** if there is a complete metric on X which gives the topology.

Proposition 3.6.1. *A G_δ subset A of a completely metrizable space X is completely metrizable in the relative toplogy.*

Proof. Suppose that d is a complete metric giving the topology on X.

We consider first the case when A is open. In this case we may consider the metric d_1 on A given by $d_1(x,y) = d(x,y) + \left| \frac{1}{d(x,A^c)} - \frac{1}{d(y,A^c)} \right|$. Then it is easy to check that d_1 is a complete metric on A which gives the relative topology.

Next suppose that $A = \cap_{n \in \mathbb{N}} O_n$ where each $O_n \subset X$ is open. for each $n \in \mathbb{N}$ we let d_n be a complete metric on O_n which gives the relative topology on O_n. Replacing $d_n(x,y)$ with $\frac{d_n(x,y)}{1+d_n(x,y)}$ we assume that $d_n(x,y) < 1$ for all $x, y \in O_n$.

We define the metric $\tilde{d}$ by $\tilde{d}(x,y) = \sum_{n=1}^{\infty} 2^{-n} d_n(x,y)$. It is then easy to see that $\tilde{d}$ gives a complete metric on A which gives the relative topology. ■

Theorem 3.6.2 (Kuratowski's Extension Theorem)**.** *Let X be a Hausdorff space and $A \subset X$ a dense subset of X. Suppose that Y is a completely metrizable space and $f : A \to Y$ is continuous, then there exists a continuous extension $\tilde{f} : B \to Y$ where $B \subset X$ is G_δ with $A \subset B$.*

Proof. We fix a complete metric d on Y. For $x \in X$ we set

$$\mathrm{osc}_f(x) = \inf\{\mathrm{diam} f(U \cap A) \mid U \text{ an open neighborhood of } x\}.$$

Then $B_n = \{x \in X \mid \mathrm{ocs}_f(x) < 1/n\}$ is open for each $n \in \mathbb{N}$, and hence $B = \cap_n B_n = \{x \in X \mid \mathrm{osc}_f(x) = 0\}$ is G_δ. Since f is continuous on A we have $A \subset B$.

We define $\tilde{f} : B \to Y$ by $\tilde{f}(x) = \lim_{\alpha\to\infty} f(x_\alpha)$ where $\{x_\alpha\}_\alpha \subset A$ is any net such that $x_\alpha \to x$. Since $B = \{x \in X \mid \mathrm{osc}_f(x) = 0\}$ and Y is a complete space it follows easily that $\tilde{f}$ is a well defined continuous extension of f. ■

Corollary 3.6.3. *Let X be a Hausdorff space and $A \subset X$ a dense subset such that A is completely metrizable, then A is a G_δ-set in X.*

Proof. If we consider the identity map on A then from Kuratowski's theorem there exists a G_δ-set $G \subset X$ with $A \subset G$ and a continuous extension $\tilde{f} : G \to A \subset G$. Since A is dense in G and $\tilde{f}$ agrees with the identity on A it then follows that $\tilde{f}$ is the identity map, hence $A = G$. ■

Corollary 3.6.4. *A subspace F of a completely metrizable space X is completely metrizable if and only if F is G_δ.*

Proof. Replacing X with $\overline{F}$, we may assume that F is dense. The result then follows from Proposition 3.6.1 and Corollary 3.6.3. ■

A **Polish space** is a topological space X which is separable and completely metrizable.

Corollary 3.6.5. *A topological space X is Polish if and only if X is homeomorphic to a G_δ subset of a second countable compact Hausdorff space.*

Proof. If X is Polish, then Proposition 3.5.8 shows that X is homeomorphic to a subset of a second countable compact Hausdorff space, and Corollary 3.6.3 shows that this subset must be G_δ.

Conversely, second countable compact Hausdorff spaces are completely metrizable by Urysohn's Metrization Theorem, hence if X is a G_δ subset then X is completely metrizable by Corollary 3.6.4. ■

A topological space X is **Čech-complete** if it is homeomorphic to a G_δ-subset of a compact Hausdorff space.

Corollary 3.6.6. *Let X be a completely metrizable space, then X is Čech-complete.*

Proof. Let $\beta : X \to \beta X$ be the Stone-Čech compactification of X. As X is Tychonoff, β is a homeomorphism onto its image. Corollary 3.6.3 shows that the image must be a G_δ-set. ■

Note that, by considering the one point compactification, any locally compact Hausdorff space is also Čech-complete.

Let X be a topological space. A subset $A \subset X$ is **meager** if it is a countable union of nowhere dense sets. A subset $B \subset X$ is **comeager** (or **residual**) if its complement is meager. We say that X is a **Baire space** if every comeager set is dense. Equivalently, X is a Baire space if whenever $\{O_n\}_{n\in\mathbb{N}}$ is a sequence of open dense sets, we have that $\cap_n O_n$ is dense. The following lemma is left to the reader.

Lemma 3.6.7. *Let X be a Baire space, and $Y \subset X$ a dense G_δ-subset, then Y is a Baire space.*

Theorem 3.6.8 (Baire)**.** *Čech-complete spaces are Baire.*

Proof. By the previous lemma it is enough to show that compact Hausdorff spaces are Baire. Thus, suppose X is a compact Hausdorff space and $\{O_n\}_{n\in\mathbb{N}}$ is a sequence of open dense sets. Let U be any non-empty open set in X. We now inductively define a decreasing sequence of closed sets $\{F_n\}_n$, and non-empty open sets $\{G_n\}_n$ such that $G_n \subset F_n \subset U \cap O_1 \cap \cdots O_n$: Since O_1 is dense, we have $O_1 \cap U \neq \emptyset$. Let F_1 be a non-empty closed subset of $O_1 \cap U$ with non-empty interior G_n. Now suppose $F_1, \ldots, F_n$ and $G_1, \ldots, G_n$ have been constructed. Since O_{n-1} is dense we have $O_{n-1} \cap G_n$ is non-empty and hence we may take F_{n+1} to be any closed subset of $O_{n-1} \cap G_n$ with non-empty interior G_{n+1}.

By construction we then have $\cap_n F_n \subset U \cap \cap_n O_n$, and by compactness we have that $\cap_n F_n$ is not empty. Since U was an arbitrary non-empty open subset it follows that $\cap_n O_n$ is dense in X. ■

3.6.1 Exercises

Note that from Corollary 3.6.5 the space of irrationals $\mathbb{R} \setminus Q$ with its subspace topology is Polish, even though the usual metric is far from complete. The next two exercises give an explicit way to see this.

Exercise 3.6.9. Suppose $\{X_n\}_{n=1}^\infty$ is a sequence of completely metrizable spaces. Show that $\prod_{n=1}^\infty X_n$ is completely metrizable. Moreover, show that $\prod_{n=1}^\infty X_n$ is separable if each X_n is separable.

Exercise 3.6.10. Show that $\mathbb{R} \setminus Q$ and **the Baire space** $\mathbb{N}^\mathbb{N}$ are homeomorphic. Hint: Consider continued fraction expansions.

Exercise 3.6.11. Let X be a compact Hausdorff space and suppose $|X| = \infty$. Show that, as a complex vector space, $C(X)$ has no countable basis.

Exercise 3.6.12. Suppose $f_n : \mathbb{R} \to \mathbb{R}$ are continuous, and $f : \mathbb{R} \to \mathbb{R}$ such that $f_n(x) \to f(x)$ for each $x \in \mathbb{R}$.

1. Show that if I is an open interval of positive length then $f^{-1}(I) \cap \overline{f^{-1}(I)^c}$ is F_σ and nowhere dense.

2. Show that if f is not continuous at a point x then there exists an open interval I with rational endpoints such that $x \in f^{-1}(I) \cap \overline{f^{-1}(I)^c}$.

3. Show that f is continuous on a dense set of points in $\mathbb{R}$.

Exercise 3.6.13. Show that there exists a function $f \in C([0,1])$ so that f is not monotone on any interval of positive length.

Exercise 3.6.14 ([hg])**.** Let $f : \mathbb{R} \to \mathbb{R}$ be an infinitely differentiable function such that at each point $x \in \mathbb{R}$ there is a derivative $f^{(n)}$ so that $f^{(n)}(x) = 0$. Let

$$Y = \{x \in \mathbb{R} \,|\, f_{|O} = p_{|O} \text{ for some polynomial } p \text{ and some neighborhood } O \text{ of } x\},$$

and let $X = Y^c$. Suppose that $X \neq \emptyset$. For each $n \geq 0$ let $S_n = \{x \in \mathbb{R} \mid f^{(n)}(x) = 0\}$.

(a) Show that X is a closed set without isolated points.

(b) Show that there exists an interval (a,b) such that $\emptyset \neq (a,b) \cap X \subset S_n$.

(c) Reach a contradiction by showing that $f^{(n)}(x) = 0$ for all $x \in (a,b)$.

(d) Conclude that, in fact, $X = \emptyset$, and deduce from this that f agrees with a polynomial on $\mathbb{R}$.

Exercise 3.6.15. For each $n, m \in \mathbb{N}$ let

$$A_{n,m} = \left\{ f \in C([0,1]) \mid \text{ there exists } x \in [0,1] \text{ such that } \left| \frac{f(t) - f(x)}{t - x} \right| \leq n \text{ if } 0 < |x - t| < \frac{1}{m} \right\}.$$

1. Show that if $f \in C([0,1])$ is differentiable at some point in $[0,1]$ then $f \in A_{n,m}$ for some $n, m \in \mathbb{N}$.

2. Show that $A_{n,m}$ is closed in $C([0,1])$.

3. Show that $A_{n,m}$ is nowhere dense in $C([0,1])$.

4. Show that the set of functions $f \in C([0,1])$ which are nowhere differentiable is dense in $C([0,1])$.

3.7 Cantor spaces

A **Cantor space** is a non-empty compact Hausdorff space without isolated points and having a countable basis consisting of clopen sets. For example, suppose $\{F_n\}_{n=1}^{\infty}$ is a sequence of finite sets with $|F_n| \geq 2$, which we consider as discrete topological spaces, then $X = \prod_{n=1}^{\infty} F_n$ is a Cantor space. Indeed, by Tychonoff's theorem X is non-empty compact Hausdorff. A countable basis of clopen sets are given by finite intersections of sets of the form $\pi_n^{-1}(E_n)$ where $E_n \subset F_n$ is a non-empty set. It is also easy to see that since $|F_n| \geq 2$, X has no isolated points.

Given a set A, we let $A^{<\mathbb{N}}$ denote the set of finite sequences in A, i.e., $A^{<\mathbb{N}}$ consists of the empty set, together with the disjoint union of A^n, for $n \in \mathbb{N}$. If $s = (s_1, \ldots, s_n) \in A^{<\mathbb{N}}$, and $k \in A$, we denote by $s\hat{\ }k$ the sequence $(s_1, \ldots, s_n, k)$. If $s \in A^{\mathbb{N}}$, and $n \in \{0\} \cup \mathbb{N}$ then we denote by $s|n$ the sequence which consists of the first n entires of s.

A **Cantor scheme** on a set X is a family $\{A_s\}_{s \in \{0,1\}^{<\mathbb{N}}}$ of subsets of X such that

1. $A_{s\hat{\ }0} \cap A_{s\hat{\ }1} = \emptyset$, for $s \in \{0,1\}^{<\mathbb{N}}$,
2. $A_{s\hat{\ }i} \subset A_s$, for $s \in \{0,1\}^{<\mathbb{N}}$.

If (X, d) is a metric space and we additionally have $\lim_{n\to\infty} \mathrm{diam}(A_{s|n}) = 0$ for any $s \in \{0,1\}^{\mathbb{N}}$ then we say that $\{A_s\}_{s \in \{0,1\}^{<\mathbb{N}}}$ has **vanishing diameter**.

Lemma 3.7.1. *Suppose (X, d) is a complete metric space and we have a Cantor scheme $\{A_s\}_{s \in \{0,1\}^{<\mathbb{N}}}$ which has vanishing diameter and such that A_s is nonempty open for each $s \in \{0,1\}^{<\mathbb{N}}$. Then for each $s \in \{0,1\}^{<\mathbb{N}}$ there is a unique element $f(s)$ in $\cap_{n\in\mathbb{N}} \overline{A_{s|n}}$, and the map $f : \{0,1\}^{\mathbb{N}} \to X$ is a continuous embedding.*

If, moreover, we have that $A_\emptyset = X$ and $A_s = A_{s\hat{\ }0} \cup A_{s\hat{\ }1}$, for $s \in \{0,1\}^{<\mathbb{N}}$, then f is a homeomorphism.

Proof. Since $\{A_s\}_{s \in \{0,1\}^{<\mathbb{N}}}$ has vanishing diameter and (X, d) is complete it follows that f is well defined. Moreover, since $A_{s\hat{\ }0} \cap A_{s\hat{\ }1} = \emptyset$, for $s \in \{0,1\}^{<\mathbb{N}}$, it follows that f is injective.

A sequence $\{s_n\}_{n=1}^{\infty} \subset \{0,1\}^{\mathbb{N}}$ converges to a point s if and only if for each $k \in \mathbb{N}$ we have $s_n|k = s|k$ for large enough n, thus it follows that $f(s_n) \in A_{s|k}$ for large enough n, since k is arbitrary we then have $\lim_{n\to\infty} f(s_n) = f(s)$. Therefore f is continuous.

If $A_\emptyset = X$ and $A_s = A_{s\hat{\ }0} \cup A_{s\hat{\ }1}$, for $s \in \{0,1\}^{<\mathbb{N}}$, then it follows that f is surjective and hence a homeomorphism since $\{0,1\}^{\mathbb{N}}$ is compact. ■

Theorem 3.7.2 (Brouwer)**.** *Any two Cantor spaces are homeomorphic.*

Proof. Let C be a Cantor space. By Lemma 3.7.1, to prove the theorem it is enough to produce a Cantor scheme $\{A_s\}_{s \in \{0,1\}^{<\mathbb{N}}}$ which has vanishing diameter and satisfies

1. $A_\emptyset = X$;

2. A_s is open nonempty;

3. $A_s = A_{s^\frown 0} \cup A_{s^\frown 1}$, for $s \in \{0,1\}^{<\mathbb{N}}$.

We construct $\{A_s\}_{s\in\{0,1\}^{<\mathbb{N}}}$ as follows: Fix a compatible metric d on X and decompose X as a disjoint union $X = X_1 \cup X_2 \cup \cdots \cup X_n$ so that each X_i is nonempty clopen and has diameter at most $1/2$. Using the notation $a^k = aa\ldots a$ (k times), we define the sets $A_{0^k\frown 1} = X_{k+1}$, for $0 \leq k < n-1$, and $A_{0^k} = X_{i+1} \cup \cdots \cup X_n$, for $0 \leq k < n$. We now repeat this process within each X_i, using sets of diameter at most $1/3$. We then continue this process by induction. ■

Proposition 3.7.3. *Let X be a nonempty Polish space without isolated points, then there exists an embedding of $\{0,1\}^{\mathbb{N}}$ into X.*

Proof. Let d be a compatible metric on X. By Lemma 3.7.1 it is enough to produce a Cantor scheme $\{U_s\}_{s\in\{0,1\}^{<\mathbb{N}}}$ in X such that

1. U_s is open nonempty for each s;

2. $\mathrm{diam}(U_s) \leq 2^{-\mathrm{length}(s)}$;

3. $U_{s^\frown i} \subset U_s$, for $s \in \{0,1\}^{<\mathbb{N}}$.

We construct such a scheme by induction on the length. Let $U_\emptyset$ be any open nonempty set with diameter at most 1. Given U_s, as X has no isolated points there exist distinct points $x, y \in U_s$. We then let $U_{s^\frown 0}$ and $U_{s^\frown 1}$ be disjoint open neighborhoods in U_s of x and y respectively so that each has diameter at most $2^{-\mathrm{length}(s)-1}$. ■

Theorem 3.7.4 (The Cantor-Bendixson theorem)**.** *Let X be a Polish space. Then X can be written uniquely as $P \cup C$, where P has no isolated points, and C is countable open.*

Proof. We let P denote the set of points $x \in X$ such that any neighborhood of x has uncountably many points, and we let $C = P^c$. If $\{O_n\}_{n\in\mathbb{N}}$ is a countable open basis, then C is the union of all countable O_n, hence, C is countable and open. Each neighborhood in X of each point in P is uncountable, and since C is countable, this also holds for each neighborhood in P, thus P has no isolated points.

For uniqueness, suppose that $X = Q \cup D$ where D is countable open and Q has no isolated points. Since D is countable open we clearly have that $D \subset C$. If $x \in C \setminus D$ were isolated as C is open we would have that x is also isolated in Q, however, Q has no isolated points and hence we conlude that $C \setminus D$ also has no isolated points. By Proposition 3.6.1 we then have that $C \setminus D$ is a countable Polish space without isolated points. Proposition 3.7.3 then shows that $C \setminus D = \emptyset$ and hence $C = D$, and $P = Q$. ■

Lemma 3.7.5. *Suppose $A \subset \{0,1\}^{\mathbb{N}}$ is nonempty closed, then there exists a continuous map $f : \{0,1\}^{\mathbb{N}} \to A$ so that f is the identity on A.*

Proof. For each $s \in \{0,1\}^{\mathbb{N}}$ we define $f(s)$ to be the point in A which has the longest common initial segment with s. It's easy to see that f is well defined, and if $s, t \in \{0,1\}^{\mathbb{N}}$ and $k \in \mathbb{N}$ such that $s|k = t|k$ then $f(s)|k = f(t)|k$. Since a sequence $\{s_n\}_{n=1}^{\infty} \subset \{0,1\}^{\mathbb{N}}$ converges to a point s if and only if for each $k \in \mathbb{N}$ we have $s_n|k = s|k$ for large enough n, it then follows that f is continuous. ■

In Proposition 3.5.8 we saw that every compact metric has a continuous injective map into the Hilbert cube. The following result gives a nice complement to this result.

Theorem 3.7.6 (The Hausdorff-Alexandroff theorem)**.** *Every nonempty compact metric space X is a continuous image of the Cantor space.*

Proof. Set $\mathcal{C} = \{0,1\}^{\mathbb{N}}$. We first prove the theorem in the case when X is the Hilbert cube. Note that the map $f(x) = \sum_{n=0}^{\infty} x_n 2^{-n-1}$ maps $\mathcal{C}$ continuously onto $[0,1]$, hence $\mathcal{C}^{\mathbb{N}}$ maps continuously onto $[0,1]^{\mathbb{N}}$. Since $\mathcal{C}^{\mathbb{N}}$ is homeomorphic to $\mathcal{C}$ we are done.

We now consider the general case. Since X is a compact metric space, Proposition 3.5.8 shows that we may assume $X \subset [0,1]^{\mathbb{N}}$. From above we know that there is a continuous surjection $g : \mathcal{C} \to [0,1]^{\mathbb{N}}$. Then $g^{-1}(X) \subset \mathcal{C}$ is closed and from Lemma 3.7.5 there is a continuous surjection $h : \mathcal{C} \to g^{-1}(X)$. The map $g \circ h$ then gives a continuous surjection of $\mathcal{C}$ onto X. ■

3.7.1 Exercises

Exercise 3.7.7. Give an explicit homoemorphism between the Cantor space $\{0,1\}^{\mathbb{N}}$ and the usual Cantor set $C \subset [0,1]$.

A **Souslin scheme** on a set X is a family $\{A_s\}_{s \in \mathbb{N}^{<\mathbb{N}}}$ of subsets of X. A **Lusin scheme** on X is a Souslin scheme such that

1. $A_{s\hat{}i} \cap A_{s\hat{}j} = \emptyset$, for $s \in \mathbb{N}^{<\mathbb{N}}$, $i \neq j$.
2. $A_{s\hat{}i} \subset A_s$, for $s \in \mathbb{N}^{<\mathbb{N}}$.

If (X,d) is a metric space and we additionally have $\lim_{n\to\infty} \mathrm{diam}(A_{s|n}) = 0$ for any $s \in \{0,1\}^{\mathbb{N}}$ then we say that $\{A_s\}_{s \in \mathbb{N}^{<\mathbb{N}}}$ has **vanishing diameter**. In this case we let $D = \{s \mid \cap_{n\in\mathbb{N}} A_{s|n} \neq \emptyset\}$, and for $s \in D$ we define $f(s) \in X$ so that $\{f(s)\} = \cap_{n\in\mathbb{N}} A_{s|n}$. The map $f : D \to X$ is the **associated map**.

Exercise 3.7.8. Suppose (X,d) is a metric space and we have a Souslin scheme $\{A_s\}_{s \in \mathbb{N}^{<\mathbb{N}}}$ which has vanishing diameter, and associated map $f : D \to X$.

1. Show that f is continuous.
2. Show that f is open if each A_s is open and $A_s \subset \cup_{n\in\mathbb{N}} A_{s\hat{}n}$, for all $s \in \mathbb{N}^{<\mathbb{N}}$.

3. Show that f is injective if $\{A_s\}_{s\in\mathbb{N}^{<\mathbb{N}}}$ is a Lusin scheme.

4. Show that f is surjective if $A_\emptyset = X$, and $A_s = \cup_{n\in\mathbb{N}} A_{s\hat{\ }n}$, for all $s \in \mathbb{N}^{<\mathbb{N}}$.

Exercise 3.7.9. Show that every nonempty Polish space without isolated points X is the continuous image of the Baire space $\mathbb{N}^{\mathbb{N}}$.

Exercise 3.7.10. Let (X, d) be a complete metric space such that every compact subset of X has empty interior. Show that for each nonempty open set $A \subset X$, there exists $\varepsilon > 0$, so that if $A \subset \cup_{n\in\mathbb{N}} B_n$ and $\operatorname{diam}(B_n) < \varepsilon$, then $B_n \neq \emptyset$ for infinitely many $n \in \mathbb{N}$.

Exercise 3.7.11. Prove the Alexandrov-Urysohn Theorem: If X is a Polish space which has a countable basis of clopen sets and such that any compact subset of X has empty interior, then X is homeomorphic to $\mathbb{N}^{\mathbb{N}}$.

Exercise 3.7.12. Show that if (X, d) is a nonempty countable metric space without isolated points, then there is an embedding $F : X \to \mathbb{N}^{\mathbb{N}}$ which has dense image. Hint: Find a Lusin scheme on X so that the associated map $f : D \to X$ is open, and bijective, with D dense, and then let $F = f^{-1}$.

Exercise 3.7.13. Prove Sierpiński's Theorem: Let (X, d) be a nonempty countable metric space without isolated points, then X is homeomorphic to $\mathbb{Q}$ with its usual topology. Hint: Combine Exercises 3.1.9, 3.6.10, and 3.7.12.

Exercise 3.7.14. Consider $\mathbb{Q}^2$ with the **lexicographical ordering** $(q, r) \leq (s, t)$ if and only if either $q < s$, or $q = s$ and $r \leq t$. Show that $\mathbb{Q}^2$ with the corresponding order topology is homeomorphic to $\mathbb{Q}^2$ with its usual product topology.

Exercise 3.7.15 (Benyamini)**.** Show that there exists $f \in C_b(\mathbb{R})$ such that given any doubly infinite sequence $\{y_n\}_{n\in\mathbb{Z}}$ there is a point $t \in \mathbb{R}$ so that $y_n = f(t + n)$ for $n \in \mathbb{Z}$. Hint: If $C \subset [0, 1/2]$ is homeomorphic to the Cantor set, first construct a continuous surjection $f : C \to [0, 1]^{\mathbb{Z}}$. Then define $g : \cup_{n\in\mathbb{Z}} C + n \to \mathbb{R}$ by $g(t + n) = \pi_n(f(t))$ for $t \in C$. Then extend g to a continuous function in $C_b(\mathbb{R})$.

3.8 Standard Borel spaces

A **standard Borel space** is a measurable space $(X, \mathcal{M})$ so that $\mathcal{M}$ is the Borel σ-algebra for some Polish topology on X. A **standard measure space** is a σ-finite measure space $(X, \mathcal{M}, \mu)$ whose underlying measurable structure $(X, \mathcal{M})$ is a standard Borel space. If, in addition, we have $\mu(X) = 1$, then we say that $(X, \mathcal{M}, \mu)$ is a **standard probability space**. A **Lebesgue space** is a standard probability space which has no atoms, e.g., $X = [0, 1]$ with its Borel structure and Lebesgue measure.

Lemma 3.8.1. *Let X be a Polish space with topology $\mathcal{T}$, and suppose $A \subset X$ closed, then $\mathcal{T} \cup \{A\}$ is again a Polish topology.*

Proof. Since X is Polish there exists a complete metric d on X such that (X, d) gives the topology $\mathcal{T}$ on X. Replacing d with $\frac{d(x,y)}{1+d(x,y)}$ we may assume that $\mathrm{diam}_d(X) \leq 1$.

Since $A \subset X$ is closed, d restricts to a complete metric on A, and from Proposition 3.6.1 we have a complete metric d_1 on A^c, which satisfies $\mathrm{diam}_{d_1}(A^c) \leq 1$, and gives the topological structure to A^c. We may then define a metric on X by

$$\tilde{d}(x, y) = \begin{cases} d(x, y) & \text{if } x, y \in A, \\ d_1(x, y) & \text{if } x, y \in A^c, \\ 1 & \text{otherwise.} \end{cases}$$

Then $\tilde{d}$ is a complete metric on X, and the corresponding topology is $\mathcal{T} \cup \{A\}$. ∎

Lemma 3.8.2. *Let X be a Polish space with topology $\mathcal{T}$, and suppose that $\{\mathcal{T}_n\}_{n=1}^{\infty}$ is a sequence of Polish topologies on X such that $\mathcal{T} \subset \mathcal{T}_n$ for each $n \in \mathbb{N}$, then the topology $\mathcal{T}_\infty$ generated by $\cup_{n\in\mathbb{N}}\mathcal{T}_n$ is again a Polish topology on X. Moreover, if $\mathcal{B}(\mathcal{T}_n) = \mathcal{B}(\mathcal{T})$ for all $n \in \mathbb{N}$, then $\mathcal{B}(\mathcal{T}_\infty) = \mathcal{B}(\mathcal{T})$.*

Proof. Let $X_n = X$ for $n \in \mathbb{N}$. Consider the map $\varphi : X \to \prod_{n=1}^{\infty} X_n$ given by $\varphi(x) = (x, x, \ldots)$. Then φ gives a homeomorphism between $(X, \mathcal{T}_\infty)$ and $\varphi(X) \subset \prod_{n=1}^{\infty} X_n$. Thus, to show that $(X, \mathcal{T}_\infty)$ is Polish, it is enough to show that $\varphi(X) \subset \prod_{n=1}^{\infty} X_n$ is closed. Suppose $(x_n) \notin \varphi(X)$, then for some $i < j$ we have $x_i \neq x_j$. We let U and V be disjoint open sets in $\mathcal{T}$ (hence also in $\mathcal{T}_i$ and $\mathcal{T}_j$) so that $x_i \in U$ and $x_j \in V$, then

$$(x_n) \in \pi_i^{-1}(U) \cap \pi_j^{-1}(V) \subset \varphi(X)^c.$$

Since Polish spaces are separable, given any set $\mathcal{G}$ which generates the topology, we have that any open set is a countable union of finite intersections in $\mathcal{G}$. Thus, if $\mathcal{G}$ is in a given σ-algebra $\mathcal{M}$, then this σ-algebra contains all Borel sets.

If $\mathcal{T}_n \subset \mathcal{B}(\mathcal{T})$, then $\cup_{n\in\mathbb{N}}\mathcal{T} \subset \mathcal{B}(\mathcal{T})$ and this generates the Polish topology $\mathcal{T}_\infty$. Thus, from the remark above we have that $\mathcal{B}(\mathcal{T}_\infty) \subset \mathcal{B}(\mathcal{T})$. ∎

Theorem 3.8.3. *Let X be a Polish space, and $\{E_n\}_{n\in\mathbb{N}}$ a countable collection of Borel subsets, then there exists a finer Polish topology on X with the same Borel structure, such that for each $n \in \mathbb{N}$, E_n is clopen in this new topology.*

Proof. We first consider the case of a single Borel subset $E \subset X$. We let $\mathcal{A}$ denote the set of subsets which satisfy the conclusion of the theorem and we let $\mathcal{B}$ be the σ-algebra of Borel subsets of X.

Lemma 3.8.1 shows that $\mathcal{A}$ contains all closed subsets of (X, d). It is also clear that $\mathcal{A}$ is closed under taking complements. Thus, to conclude that $\mathcal{B} \subset \mathcal{A}$ it is then enough to show that $\mathcal{A}$ is closed under countable intersections. If $A_n \in \mathcal{A}$, and $\mathcal{T}_n$ are finer Polish topologies on X, with Borel structure $\mathcal{B}$, such that A_n is clopen in $\mathcal{T}_n$ for each $n \in \mathbb{N}$, then by Lemma 3.8.2 there is a finer Polish topology $\mathcal{T}_\infty$ which generates $\mathcal{B}$ and such that A_n is clopen in $\mathcal{T}_\infty$, for each

$n \in \mathbb{N}$. We then have that $\cap_{n\in\mathbb{N}} A_n$ is closed in $\mathbb{T}_\infty$, and hence by Lemma 3.8.1 we have that $\cap_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Having established the result for a single Borel set E, we may then apply Lemma 3.8.1 to obtain the result for a sequence of Borel sets $\{E_n\}_{n\in\mathbb{N}}$. ■

Corollary 3.8.4. *Let $(X, \mathcal{B})$ be a standard Borel space, and $E \in \mathcal{B}$ a Borel subset, then $(E, \mathcal{B}|_E)$ is a standard Borel space.*

Proof. By the previous theorem we may assume X is Polish and $E \subset X$ is clopen, and hence Polish. We then have that $\mathcal{B}|_E$ is the associated Borel structure on E and hence $(E, \mathcal{B}|_E)$ is standard. ■

Corollary 3.8.5. *Let X be a standard Borel space, Y a Polish space, and $f : X \to Y$ a Borel map, then there exists a Polish topology on X which generates the same Borel structure and such that f is continuous with respect to this topology.*

Proof. Let $\{E_n\}$ be a countable basis for the topology on Y. By Theorem 3.8.3 there exists a Polish topology on X which generates the same Borel structure and such that $f^{-1}(E_n)$ is clopen for each $n \in \mathbb{N}$. Hence, in this topology f is continuous. ■

Lemma 3.8.6. *Let X be a Polish space, then there exists a Souslin scheme $\{E_s\}_{s\in\mathbb{N}^{<\mathbb{N}}}$ consisting of Borel subsets such that the the following conditions are satisfied:*

(*i*) *$E_\emptyset = X$.*

(*ii*) *For each $s \in \mathbb{N}^{<\mathbb{N}}$, $E_s = \sqcup_{k\in\mathbb{N}} E_{s\hat{\ }k}$.*

(*iii*) *For each $s \in \mathbb{N}^{\mathbb{N}}$ the set $\cap_{n\in\mathbb{N}} \overline{E_{s|n}}$ consists of at most one element.*

(*iv*) *For each $s \in \mathbb{N}^{\mathbb{N}}$, $\cap_{n\in\mathbb{N}} \overline{E_{s|n}} = \{x\} \neq \emptyset$ if and only if $E_{s|n} \neq \emptyset$ for all $n \in \mathbb{N}$, and in this case for any sequence $x_n \in E_{s|n}$ we have $x_n \to x$.*

Proof. Let d be a complete metric on X which generates the Polish topology on X, and such that X has diameter at most 1. We will inductively construct $\{E_s\}_{s\in\mathbb{N}^{<\mathbb{N}}}$ so that for $s \in \mathbb{N}^n$ the diameter of E_s is at most 2^{-n}. First, we set $E_\emptyset = X$. Now suppose E_s has been constructed for each $s \in \{\emptyset\} \cup_{n=1}^k \mathbb{N}^n$. If $s \in \mathbb{N}^k$, let $\{x_n\}_{n\in\mathbb{N}}$ be a countable dense subset of E_s (note that any subspace of a separable metric space is again separable).

We define $E_{s\hat{\ }i} = E_s \cap (B_{2^{-k-1}}(x_i) \setminus \cup_{j<i} B_{2^{-k-1}}(x_j))$, where $B_r(x)$ denotes the open ball of radius r centered at x. It is then easy to see that for each $s \in \mathbb{N}^{<\mathbb{N}}$, we have $E_s = \sqcup_{k\in\mathbb{N}} E_{s\hat{\ }k}$. Moreover, for each $s \in \mathbb{N}^{\mathbb{N}}$, we have that $E_{s|n}$ has diameter at most 2^{-n}, hence $\cap_{n\in\mathbb{N}} \overline{E_{s|n}}$ contains at most one element. Finally, if $\overline{E_{s|n}} \neq \emptyset$ for all $n \in \mathbb{N}$, then as the diameter of $E_{s|n}$ converges to 0, it follows from completeness, that there exists $x \in \cap_{n\in\mathbb{N}} \overline{E_{s|n}}$, and for each sequence $x_n \in E_{s|n}$ we have $x_n \to x$. ■

If X is a standard Borel space and $A, B \subset X$ are disjoint, then we say that A and B are **Borel separated** if there exists a Borel subset $E \subset X$ such that $A \subset E$, and $B \subset X \setminus E$.

Lemma 3.8.7. *Let X be a standard Borel space and suppose that $A = \cup_{n\in\mathbb{N}} A_n$, and $B = \cup_{m\in\mathbb{N}} B_m$, are such that A_n and B_m are Borel separated for each $n, m \in \mathbb{N}$, then A and B are Borel separated.*

Proof. Suppose $E_{n,m}$ is a Borel subset which separates A_n and B_m for each $n, m \in \mathbb{N}$. Then $E = \cup_{n\in\mathbb{N}} \cap_{m\in\mathbb{N}} E_{n,m}$ separates A and B. ■

If X is a Polish space, a subset $E \subset X$ is **analytic** if there exits a Polish space Y and a continuous function $f : Y \to X$ such that $E = f(Y)$. Note that it follows from Corollary 3.8.5 that if $f : Y \to X$ is Borel then $f(Y)$ is analytic. In particular, it follows that all Borel sets are analytic. If X is a standard Borel space then a subset $E \subset X$ is **analytic** if it is analytic for some (and hence all) Polish topologies on X which give the Borel structure.

Theorem 3.8.8 (The Lusin Separation Theorem)**.** *Let X be a standard Borel space, and $A, B \subset X$ two disjoint analytic sets, then A and B are Borel separated.*

Proof. We may assume that X is a Polish space, and that there are Polish spaces Y_1, and Y_2, and continuous functions $f_i : Y_i \to X$ such that $A = f_1(Y_1)$ and $B = f_2(Y_2)$.

Let $\{E_s\}_{s\in\mathbb{N}^{<\mathbb{N}}}$ (resp. $\{F_s\}_{s\in\mathbb{N}^{<\mathbb{N}}}$) be a Souslin scheme for Y_1 (resp. Y_2) which satisfies the conditions in Lemma 3.8.6. If A and B are not Borel separated then by Lemma 3.8.7 we may recursively define sequences $s, r \in \mathbb{N}^{\mathbb{N}}$ such that $f_1(E_{s|n})$ and $f_2(F_{r|n})$ are not Borel separated for each $n \in \mathbb{N}$. In particular, we have that $E_{s|n}$ and $F_{r|n}$ are non-empty for each $n \in \mathbb{N}$, hence there exists $a \in Y_1$, $b \in Y_2$ such that $\cap_{n\in\mathbb{N}}\overline{E_{s|n}} = \{a\}$, $\cap_{n\in\mathbb{N}}\overline{F_{r|n}} = \{b\}$.

If $V, W \subset X$ are disjoint open subsets with $f_1(a) \in V$, and $f_2(b) \in W$, then by continuity of f_i, for large enough n we have $f_1(E_{x|n}) \subset V$, and $f_2(F_{y|n}) \subset W$. Hence V separates $E_{x|n}$ from $F_{y|n}$ for large enough n, a contradiction. ■

Corollary 3.8.9. *If X is a standard Borel space then a subset $E \subset X$ is Borel if and only if both E and $X \setminus E$ are analytic.*

Corollary 3.8.10. *let X be a standard Borel space, and let $\{A_n\}_{n\in\mathbb{N}}$ be a sequence of disjoint analytic subsets, then there exists a sequence $\{E_n\}_{n\in\mathbb{N}}$ of disjoint Borel subsets such that $A_n \subset E_n$ for each $n \in \mathbb{N}$.*

Proof. It is easy to see that the countable union of analytic sets is analytic. Hence, by Lusin's separation theorem we may inductively define a sequence of Borel subsets $\{E_n\}_{n\in\mathbb{N}}$ such that $A_n \subset E_n$, while $(\cup_{k>n} A_k) \cup (\cup_{k<n} E_k) \subset X \setminus A_n$. ■

Theorem 3.8.11 (Lusin-Souslin)**.** *Let X and Y be standard Borel spaces, and $f : X \to Y$ an injective Borel map, then $f(X)$ is Borel, and f implements an isomorphism of standard Borel spaces between X and $f(X)$.*

Proof. We first show that $f(X)$ is Borel. By Corollary 3.8.5 we may assume that X and Y are Polish spaces and f is continuous. Let $\{E_s\}_{s\in\mathbb{N}^{<\mathbb{N}}}$ be a Souslin scheme for X which satisfies the conditions of Lemma 3.8.6. Then $\{f(E_s)\}_{s\in\mathbb{N}^{<\mathbb{N}}}$ gives a Souslin scheme of analytic sets for Y, and since f is injective it follows that for each $s\in\mathbb{N}^n$ we have that $\{f(E_{s\hat{}k})\}_{k\in\mathbb{N}}$ are pairwise disjoint. Thus, by Corollary 3.8.10 there exist pairwise disjoint Borel subsets $\{Y_{s\hat{}k}\}_{k\in\mathbb{N}}$ such that $f(E_{s\hat{}k})\subset Y_{s\hat{}k}$ for each $k\in\mathbb{N}$.

We inductively define a new Souslin scheme $\{C_s\}_{s\in\mathbb{N}^{<\mathbb{N}}}$ for Y by setting $C_\emptyset=Y$, and $C_{s\hat{}k}=C_s\cap\overline{f(E_{s\hat{}k})}\cap Y_{s\hat{}k}$ for all $s\in\mathbb{N}^{<\mathbb{N}}$, and $k\in\mathbb{N}$. Then for each $s\in\mathbb{N}^{<\mathbb{N}}$ we have that C_s is Borel, and also

$$f(E_s)\subset C_s\subset\overline{f(E_s)}.$$

We claim that $f(X)=\cap_{k\in\mathbb{N}}\cup_{s\in\mathbb{N}^k}C_s$, from which it then follows that $f(X)$ is Borel.

If $y\in f(X)$, then let $x\in X$ be such that $f(x)=y$. There exists $s\in\mathbb{N}^{\mathbb{N}}$ such that $x\in\cap_{k\in\mathbb{N}}E_{s|k}$, and hence $y\in\cap_{k\in\mathbb{N}}f(E_{s|k})$. Thus, $y\in\cap_{k\in\mathbb{N}}C_{s|k}\subset\cap_{k\in\mathbb{N}}\cup_{s\in\mathbb{N}^k}C_s$. Conversely, if $y\in\cap_{k\in\mathbb{N}}\cup_{s\in\mathbb{N}^k}C_s$, then there exists $s\in\mathbb{N}^{\mathbb{N}}$ such that $y\in C_{s|k}\subset\overline{f(E_{s|k})}$ for each $k\in\mathbb{N}$. Hence $E_{s|k}\neq\emptyset$ for each $k\in\mathbb{N}$ and thus $\cap_{k\in\mathbb{N}}\overline{E_{s|k}}=\{x\}$ for some $x\in X$. We must then have that $f(x)=y$, since if this were not the case there would exists an open neighborhood U of $f(x)$ such that $y\notin\overline{U}$. By continuity of f we would then have that $f(E_{s|k})\subset U$ for large enough k, and hence $y\in\cap_{k\in\mathbb{N}}\overline{f(E_{s|k})}\subset\overline{U}$, a contradiction.

Having established that $f(X)$ is Borel, the rest of the theorem follows easily. We have that f gives a bijection from X to $f(X)$ which is Borel, and if $E\subset X$ is Borel, then from Corollary 3.8.4 and the argument above we have that $f(E)$ is again Borel. Thus, f^{-1} is a Borel map. ∎

Corollary 3.8.12. *Suppose X and Y are standard Borel spaces such that there exists injective Borel maps $f:X\to Y$, and $g:Y\to X$, then X and Y are isomorphic as standard Borel spaces.*

Proof. Suppose $f:X\to Y$, and $g:Y\to X$ are injective Borel maps. From Theorem 3.8.11 we have that f and g are Borel isomorphisms onto their image and hence we may apply an argument used for the Cantor-Schröder-Bernstein theorem. Specifically, if we set $B=\cup_{n\in\mathbb{N}}(f\circ g)^n(Y\setminus f(X))$, and we set $A=X\setminus g(B)$, then we have $g(B)=X\setminus A$, and

$$f(A)=f(X)\setminus(f\circ g)(B)=Y\setminus((Y\setminus f(X))\cup(f\circ g)(B))=Y\setminus B.$$

Hence if we define $\theta:X\to Y$ by $\theta(x)=\begin{cases} f(x) & \text{if } x\in A,\\ g^{-1}(x) & \text{if } x\in Y\setminus A=g(B),\end{cases}$ then we have that θ is a bijective Borel map whose inverse is also Borel. ∎

Theorem 3.8.13 (Kuratowski)**.** *Any two uncountable standard Borel spaces are isomorphic. In particular, two standard Borel spaces X and Y are isomorphic if and only if they have the same cardinality.*

Proof. Let X be an uncountable standard Borel space, we'll show that X is isomorphic as Borel spaces to the Polish space $\mathcal{C} = 2^{\mathbb{N}}$. Note that by Corollary 3.8.12 it is enough to show that there exist injective Borel maps $f : X \to \mathcal{C}$, and $g : \mathcal{C} \to X$. Note that such a map g exists by Proposition 3.7.3 and the Cantor-Bendixson theorem, so we only need to construct f.

To construct f, fix a metric d on X such that d gives the Borel structure to X and such that the diameter of X is at most 1. Let $\{x_n\}$ be a countable dense subset of (X, d), and define $f_0 : X \to [0,1]^{\mathbb{N}}$ by $(f_0(x))(n) = d(x, x_n)$. The function f_0, is clearly injective and continuous, thus to construct f it is enough to construct an injective Borel map from $[0,1]^{\mathbb{N}}$ to $\mathcal{C}$, and since $\mathcal{C}^{\mathbb{N}}$ is homeomorphic to $\mathcal{C}$, it is then enough to construct an injective Borel map from $[0,1]$ to $\mathcal{C}$, and this is easily done. For example, if $y \in [0,1)$ then we may consider its dyadic expansion $y = \sum_{k=1}^{\infty} b_k 2^{-k}$, where in the case when y is a dyadic rational we take the expansion such that b_k is eventually 0. Then it is easy to see that $[0,1) \ni y \mapsto \{b_k\}_k \in \mathcal{C}$ gives an injective function which is continuous except at the countable family of dyadic rational, hence is Borel. We may then extend this map to $[0,1]$ by sending 1 to $(1,1,1,\cdots) \in \mathcal{C}$. ■

Theorem 3.8.14 (von Neumann)**.** *Let $(X, \mathcal{M}, \mu)$ be a standard probability space which has no atoms, then there exists an isomorphism of standard Borel spaces $\theta : X \to [0,1]$ so that $\theta_* \mu$ is Lebesgue measure on $[0,1]$.*

Proof. By Kuratowski's theorem we may assume that $(X, \mathcal{M})$ is $[0,1]$ with its Borel σ-algebra. We then consider $f : [0,1] \to [0,1]$ the cumulative distribution function $f(x) = \mu([0,x])$. Then f is a monotone nondecreasing function which is continuous on the right. Since $f(x) - \lim_{t \to x^-} f(t) = \mu(\{x\}) = 0$ we see that f is continuous. Moreover, $f(0) = 0$, and $f(1) = 1$ so f is surjective. For each $y \in [0,1]$ we have that $f^{-1}(\{y\})$ is a closed set and since f is monotone this must be a closed interval. We let M denote the set of y's so that $f^{-1}(\{y\})$ is an interval of positive length. Then M is countable and the set $N = f^{-1}(M)$ has μ-measure zero. $g = f_{|X \setminus N}$ then gives a Borel isomorphism from $X \setminus N$ to $[0,1] \setminus M$ such that $g_* \mu$ is Lebesgue measure on $[0,1] \setminus M$.

We fix $C \subset [0,1]$ an uncountable Borel set with Lebesgue measure zero, e.g., C the usual Cantor set. Then $B = g^{-1}(C)$ has μ-measure zero and is again uncountable. Thus, $\tilde{N} = N \cup B$, and $\tilde{M} = M \cup C$ are both uncountable standard Borel spaces by Corollary 3.8.4 and so by Kuratowski's theorem there is a Borel isomorphism $h : \tilde{N} \to \tilde{M}$.

We then define $\theta : X \to [0,1]$ by setting $\theta(x) = \begin{cases} g(x) & \text{if } x \in X \setminus \tilde{M}; \\ h(x) & \text{if } x \in \tilde{M}. \end{cases}$

Then θ is a Borel isomorphism and $\theta_* \mu$ is Lebesgue measure. ■

If $(X, \mathcal{M}, \mu)$ is a measure space, then we may consider on $\mathcal{M}$ the equivalence relation $\sim$ given by $E \sim F$ if $\mu(E \Delta F) = 0$. We denote the set of equivalence classes by $\widehat{\mathcal{M}} = \mathcal{M}/\sim$. We transfer the operations of complements, countable unions and countable intersections to $\widehat{\mathcal{M}}$ respectively as as $[E]' = [E^c]$, $\vee_{n=1}^{\infty}[E_n] = [\cup_{n=1}^{\infty} E_n]$, and $\wedge_{n=1}^{\infty}[E_n] = [\cup_{n=1}^{\infty} E_n]$. Note that these are well defined operations. We also denote by $\hat{\mu}$ the function on $\widehat{\mathcal{M}}$ given by $\hat{\mu}([E]) = \mu(E)$, where $[E]$ denotes the equivalence class of E. If $(Y, \mathcal{N}, \nu)$ is another measure space, then a **measure algebra homomorphism** is a map $\alpha : \widehat{\mathcal{N}} \to \widehat{\mathcal{M}}$ such that for $E, E_1, E_2, \ldots \in \mathcal{N}$ we have

1. $\alpha([\emptyset]) = [\emptyset]$;
2. $\alpha([E]') = \alpha([E])'$;
3. $\alpha(\vee_{n=1}^{\infty}[E_n]) = \vee_{n=1}^{\infty}\alpha([E_n])$;
4. $\hat{\mu}(\alpha([E])) = \hat{\nu}([E])$.

As an example, if we have a measure preserving map $\theta : X \to Y$, then we obtain a measure algebra homomorphism $\hat{\theta}$ by setting $\hat{\theta}([E]) = [\theta^{-1}(E)]$.

Theorem 3.8.15 (von Neumann)**.** *Let $(X, \mathcal{M}, \mu)$, and $(Y, \mathcal{N}, \mu)$ be probability spaces without atoms such that Y is standard. Suppose $\alpha : \widehat{\mathcal{N}} \to \widehat{\mathcal{M}}$ is a measure algebra homomorphism, then there exists a measurable map $\theta : X \to Y$ so that $\alpha = \hat{\theta}$.*

Proof. By Theorem 3.8.14 we may assume that $Y = [0, 1]$, endowed with the Borel σ-algebra and Lebesgue measure.

For each rational $r \in \mathbb{Q} \cap [0, 1]$ we choose a measurable set $X_r \subset X$ such that $[X_r] = \alpha([[0, r)]) = \alpha([[0, r]])$. We may assume that $X_0 = \emptyset$, and $X_1 = X$. By replacing X_r with $\cup_{s \in \mathbb{Q} \cap [0,1], s<r} X_s$ we may further assume that $X_r = \cup_{s \in \mathbb{Q} \cap [0,1], s<r} X_s$, and hence $X_r \subset X_s$ for all $r, s \in \mathbb{Q} \cap [0, 1]$ with $r \leq s$.

We define $\pi : X \to Y$, by $\pi(x) = \inf\{r \in \mathbb{Q} \cap [0, 1] \mid x \in X_r\}$. Note that for each $t \in [0, 1]$ we have $\pi^{-1}([0, t)) = \cup_{r \in \mathbb{Q} \cap [0,1], r<t} X_r$ and hence π is a measurable map, which satisfies

$$\hat{\pi}([[0, t)]) = \hat{\pi}([\cup_{r \in \mathbb{Q} \cap [0,1], r<t} X_r]) = \hat{\pi}([X_t]) = \alpha([[0, t)])$$

for all $t \in \mathbb{Q} \cap [0, 1]$. Using the properties of a measure algebra homomorphism it then follows that $\hat{\pi}([E]) = \alpha([E])$ for any set E in the σ-algebra generated by $\{[0, t) \mid t \in \mathbb{Q} \cap [0, 1]\}$. As this generates the Borel σ-algebra $\mathcal{N}$ it follows that $\alpha = \hat{\pi}$. ■

3.8.1 Exercises

Exercise 3.8.16. Let $(X, \mathcal{M}, \mu)$, and $(Y, \mathcal{N}, \mu)$ be probability spaces such that Y is standard, and suppose $\theta, \phi : X \to Y$ are measure preserving maps. Show that $\hat{\theta} = \hat{\phi}$ if and only if θ and ϕ agree almost everywhere.

Exercise 3.8.17. Suppose $(X, \mathcal{M}, \mu)$, and $(Y, \mathcal{N}, \mu)$ are standard probability spaces without atoms, and $\alpha : \widehat{\mathcal{N}} \to \widehat{\mathcal{M}}$ is a measure algebra homomorphism. Show that if α is surjective, then there exists an isomorphism of standard Borel spaces $\theta : X \to Y$ such that $\hat{\theta} = \alpha$.

Chapter 4

Differentiation and integration

4.1 The Lebesgue differentiation theorem

4.1.1 Vitali's covering lemma

If (X, d) is a metric space, $B \subset X$ is a (open or closed) ball in X, and $c > 0$, then we denote by cB the (open or closed respectively) ball in X having the same center as B, and having radius satisfy $\text{rad}(cB) = c\text{rad}(B)$.

Lemma 4.1.1 (Vitali's covering lemma)**.** *Let (X, d) be a metric space and let $\mathcal{F}$ be a collection of balls in X, having positive radii, such that*

$$\sup\{\text{rad}(B) \mid B \in \mathcal{F}\} < \infty,$$

then there exists a pairwise disjoint subcollection $\mathcal{G} \subset \mathcal{F}$ such that for all $B \in \mathcal{F}$ there exists $C \in \mathcal{G}$ with $B \cap C \neq \emptyset$, and with $B \subset 5C$.

Proof. Set $R = \sup\{\text{rad}(B) \mid B \in \mathcal{F}\}$ and partition $\mathcal{F}$ as $\mathcal{F} = \cup_{n=0}^{\infty} \mathcal{F}_n$ where $B \in \mathcal{F}_n$ if and only if $\text{rad}(B) \in (2^{-n-1}R, 2^{-n}R]$. We inductively define subcollections $\mathcal{G}_n \subset \mathcal{F}_n$ as follows: Set $\mathcal{H}_0 = \mathcal{F}_0$. By Zorn's lemma there exists $\mathcal{G}_0 \subset \mathcal{F}_0$, a maximal (with respect to inclusion) pairwise disjoint family. Having defined $\mathcal{G}_0, \ldots, \mathcal{G}_{n-1}$ we let

$$\mathcal{H}_n = \{C \in \mathcal{F}_n \mid C \cap B = \emptyset \text{ for all } B \in \cup_{k=0}^{n-1} \mathcal{G}_k\},$$

and we again use Zorn's lemma to find $\mathcal{G}_n \subset \mathcal{H}_n$ which is a maximal pairwise disjoint family. We set $\mathcal{G} = \cup_{n=0}^{\infty} \mathcal{G}_n$, and note that $\mathcal{G}$ is pairwise disjoint.

If $B \in \mathcal{F}$, then $B \in \mathcal{F}_n$ for some $n \geq 0$. Either $B \notin \mathcal{H}_n$, in which case there exists $C \in \cup_{k=0}^{n-1} \mathcal{G}_k$ such that $B \cap C \neq \emptyset$, or else $B \in \mathcal{H}_n$, in which case by maximality of $\mathcal{G}_n$ there exists $C \in \mathcal{G}_n$ so that $B \cap C \neq \emptyset$. Since $B \in \mathcal{F}_m$ we have $\text{rad}(B) \leq 2^{-n}R$, and so in either case there exists $C \in \mathcal{G}$ with $B \cap C \neq \emptyset$ and with $\text{rad}(C) > 2^{-n-1}R \geq \frac{1}{2}\text{rad}(B)$, hence $B \subset 5C$. ■

If $E \subset \mathbb{R}^d$, and $\mathcal{V}$ is a collection of closed balls in $\mathbb{R}^d$, then we say that $\mathcal{V}$ is a **Vitali covering** of E if for each $x \in E$, and $\varepsilon > 0$ there exists $B \in \mathcal{V}$ with $x \in B$ such that $\mathrm{rad}(B) < \varepsilon$.

Theorem 4.1.2 (Vitali's covering theorem)**.** *Let $E \subset \mathbb{R}^d$ be a Borel set with finite Lebesgue measure, and suppose $\mathcal{V}$ is a Vitali covering of E, then there exists a pairwise disjoint (hence countable) subcollection $\mathcal{G} \subset \mathcal{V}$ so that*

$$\lambda(E \setminus \cup\{C \mid C \in \mathcal{G}\}) = 0.$$

Proof. By considering the subcollection of $\mathcal{V}$ consisting of balls with radius at most 1 we may assume that $\sup\{\mathrm{rad}(B) \mid B \in \mathcal{V}\} \leq 1$. By Vitali's covering lemma there then exists a pairwise disjoint subcollection $\mathcal{G} \subset \mathcal{V}$ such that for every $B \in \mathcal{V}$ there exists $C \in \mathcal{G}$ with $B \cap C \neq \emptyset$ and with $B \subset 5C$.

Fix $r > 0$ and set $Z = (E \setminus \cup\{C \mid C \in \mathcal{G}\}) \cap B(r, 0)$. We let $\tilde{\mathcal{G}}$ denote the subcollection consisting of balls which intersect $B(r, 0)$, and hence are contained in $B(r+2, 0)$. Partition $\tilde{\mathcal{G}}$ as $\tilde{\mathcal{G}} = \cup_{n=0}^\infty \mathcal{G}_n$ where $C \in \mathcal{G}_n$ if and only if $\mathrm{rad}(C) \in (2^{-n-1}, 2^{-n}]$. Since $\tilde{\mathcal{G}}$ is pairwise disjoint we then have

$$\sum_{n=0}^\infty \sum_{C \in \mathcal{G}_n} \lambda(C) = \lambda(\cup_{C \in \tilde{\mathcal{G}}} C) \leq \lambda(B(r+2, 0)) < \infty.$$

Fix $\varepsilon > 0$. Then there exists $N \geq 0$ so that $\sum_{n=N}^\infty \sum_{B \in \mathcal{G}_n} \lambda(B) < \varepsilon$. Set $K = \cup_{n=0}^{N-1} \cup_{C \in \mathcal{G}_n} C$ which is compact as it is a finite union of closed balls. If $z \in Z$, then $z \notin K$ and as $\mathcal{V}$ is a Vitali covering of E there then exists $B \in \mathcal{V}$ such that $z \in B$, $B \subset B(r, 0)$, and $B \cap K = \emptyset$. As $B \in \mathcal{V}$ and $B \subset B(r, 0)$ there exists $C \in \tilde{\mathcal{G}}$ so that $B \cap C \neq \emptyset$ and $z \in B \subset 5C$. Since $B \cap C \neq \emptyset$ we must have $C \not\subset K$ and hence $C \in \cup_{n=N}^\infty \mathcal{G}_n$. Since $z \in Z$ was arbitrary we have then shown that

$$Z \subset \cup_{n=N}^\infty \cup_{C \in \mathcal{G}_n} 5C,$$

hence

$$\lambda(Z) \leq \sum_{n=N}^\infty \sum_{C \in \mathcal{G}_n} \lambda(5C) < 5^d \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary we conclude that $\lambda(Z) = 0$, and since $r > 0$ was arbitrary we then conclude that $\lambda(E \setminus \cup\{C \mid C \in \mathcal{G}\}) = 0$. ■

4.1.2 The Lebesgue differentiation theorem

A function $f : \mathbb{R}^d \to \mathbb{C}$ is **locally integrable** if $\int_K |f| \, d\lambda < \infty$ for any compact set $K \subset \mathbb{R}^d$. We let $L^1_{\mathrm{loc}}(\mathbb{R}^d)$ denote the space of all locally integrable functions.

Theorem 4.1.3 (The Lebesgue differentiation theorem)**.** *Let $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$, then for almost every $x \in \mathbb{R}^d$ we have*

$$f(x) = \lim_{r \to 0} \frac{1}{\lambda(B(r, x))} \int_{B(r,x)} f \, d\lambda.$$

Note that Lebesgue's differentiation theorem is obvious when f is continuous. Our strategy to prove the theorem in general will be to approximate f in $L^1(\mathbb{R}^d)$ by a continuous function. It then becomes necessary to control the the size of the set on which the averages $\frac{1}{\lambda(B(r,x))}\int_{B(r,x)} |f|\,d\lambda$ can be large in terms of the L^1-norm of the function f. This is achieved by the following lemma:

Lemma 4.1.4. *Suppose $f \in L^1(\mathbb{R}^d)$ with compact support, and for each $x \in \mathbb{R}^d$ set*

$$\tilde{f}(x) = \limsup_{r\to 0} \frac{1}{\lambda(B(r,x))}\int_{B(r,x)} |f|\,d\lambda, \tag{4.1}$$

then for $\alpha > 0$ we have

$$\lambda(\{x \in \mathbb{R}^d \mid \tilde{f}(x) > \alpha\}) \leq \frac{1}{\alpha}\|f\|_1.$$

Proof. Since f has compact support, so does $\tilde{f}$ and hence the set

$$E = \{x \in \mathbb{R}^d \mid \tilde{f}(x) > \alpha\}$$

has finite measure. Let $\mathcal{V}$ be the collection of closed balls B such that

$$\frac{1}{\lambda(B)}\int_B |f|\,d\lambda > \alpha.$$

From the definition of $\tilde{f}$ we see that $\mathcal{V}$ is a Vitali covering of E. By Vitali's covering theorem there exists a pairwise disjoint family $\mathcal{G}$, so that $\lambda(E\setminus \cup_{C\in\mathcal{G}} C) = 0$. Hence,

$$\lambda(E) \leq \sum_{C\in\mathcal{G}} \lambda(C) \leq \sum_{C\in G} \frac{1}{\alpha}\int_C |f|\,d\lambda \leq \frac{1}{\alpha}\|f\|_1.$$

■

Proof of Theorem 4.1.3. First note that if the theorem holds for $1_{B(n,0)}f$ for each $n \in \mathbb{N}$ then the theorem also holds for f, therefore we may assume that $f \in L^1(\mathbb{R}^d)$ with compact support. The set of continuous functions with compact support is dense in $L^1(\mathbb{R}^d)$. Indeed, it is enough to approximate characteristic functions, which can be easily done by combining the regularity of Lebesgue measure (Corollary 2.3.8) with Urysohn's lemma. Let g_n be a sequence of continuous functions with compact support such that $\|f - g_n\|_1 \to 0$.

For $x \in \mathbb{R}^d$ and $n \in \mathbb{N}$ we then have

$$\limsup_{r\to 0}\left|\frac{1}{B(r,x)}\int_{B(r,x)} f\,d\lambda - f(x)\right| \leq \limsup_{r\to 0}\left|\frac{1}{B(r,x)}\int_{B(r,x)} (f - g_n)\,d\lambda\right| + \limsup_{r\to 0}\left|\frac{1}{B(r,x)}\int_{B(r,x)} g_n\,d\lambda - g_n(x)\right| + |g_n(x) - f(x)|.$$

Since g_n is continuous the second term on the right vanishes. Thus,

$$\limsup_{r\to 0}\left|\frac{1}{B(r,x)}\int_{B(r,x)} f\,d\lambda - f(x)\right| \le \widetilde{f-g_n}(x) + |g_n(x)-f(x)|,$$

where $\widetilde{f-g_n}$ is defined as in (4.1).

If $\varepsilon > 0$ and we let E be the set of points where the left hand side is greater than ε then we have

$$E \subset \{x\in\mathbb{R}^d \mid \widetilde{f-g_n}(x) > \varepsilon/2\} \cup \{x\in\mathbb{R}^d \mid |g_n(x)-f(x)| > \varepsilon/2\}.$$

Using Lemma 4.1.4 together with Chebyshev's inequality then gives

$$\lambda(E) \le \frac{2}{\varepsilon}\|g_n - f\|_1 + \frac{2}{\varepsilon}\|g_n - f\|_1.$$

Taking $n\to\infty$ shows $\lambda(E) = 0$, and the result follows. ■

Corollary 4.1.5 (The Lebesgue density theorem)**.** *Suppose $E\subset\mathbb{R}^d$ is a Borel set, then for almost every $x \in E$ we have*

$$\lim_{r\to 0}\frac{\lambda(B(r,x)\cap E)}{\lambda(B(r,x))} = 1. \tag{4.2}$$

Proof. This follows immediately from the Lebesgue differentiation theorem by considering the characteristic function $f = 1_E$. ■

Points where (4.2) holds are called **points of density** for the set E. In light of the triangle inequality for integration, the following gives a slight improvement on Lebesgue's differentiation theorem.

Theorem 4.1.6. *Let $f\in L^1_{\mathrm{loc}}(\mathbb{R}^d)$, then for almost every $x\in\mathbb{R}^d$ we have*

$$\lim_{r\to 0}\frac{1}{\lambda(B(r,x))}\int_{B(r,x)} |f(y)-f(x)|\,d\lambda(y). \tag{4.3}$$

Proof. For each rational number q let Z_q denote the set of points where the formula

$$\lim_{r\to 0}\frac{1}{\lambda(B(r,x))}\int_{B(r,x)} |f(y)-q|\,d\lambda(y) = |f(x)-q|$$

does not hold. Since $x\mapsto |f(x)-q|$ is locally integrable it follows from Lebesgue's differentiation theorem that $\lambda(Z_q) = 0$ for every $q\in\mathbb{Q}$. If we set $Z = \cup_{q\in\mathbb{Q}} Z_q$ then we have $\lambda(Z) = 0$. For any $x\in\mathbb{R}^d$, $q\in\mathbb{Q}$, and $r > 0$ we have

$$\frac{1}{\lambda(B(r,x))}\int |f(y)-f(x)|\,d\lambda(y) \le \frac{1}{\lambda(B(r,x))}\int |f(y)-q|\,d\lambda(y) + |q-f(x)|.$$

Therefore if $x\notin Z$ we have

$$\limsup_{r\to 0}\frac{1}{\lambda(B(r,x))}\int |f(y)-f(x)|\,d\lambda(y) \le 2|f(x)-q|,$$

for every $q\in\mathbb{Q}$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$ the result then follows. ■

Any point $x \in \mathbb{R}^d$ which satisfies (4.3) is called a **Lebesgue point** of f.

Theorem 4.1.7. *Let μ be a complex Borel measure on $\mathbb{R}^d$, let $\mu = \mu_{ac} + \mu_s$ be its Lebesgue decomposition so that $\mu_{ac} \ll \lambda$ and $\mu_s \perp \lambda$, and let $f = \frac{d\mu_{ac}}{d\lambda}$ be the Radon-Nikodym derivative. Then for λ-almost every $x \in \mathbb{R}^d$ we have*

$$\lim_{r\to 0} \frac{\mu(B(r,x))}{\lambda(B(r,x))} = f(x).$$

Proof. In light of Lebesgue's density theorem it is enough to show that when $\mu \perp \lambda$ we have $\lim_{r\to 0} \frac{\mu(B(r,x))}{\lambda(B(r,x))} = 0$ for λ-almost every $x \in \mathbb{R}^d$. Note also that since $\left|\frac{\mu(B(r,x))}{\lambda(B(r,x))}\right| \le \frac{|\mu|(B(r,x))}{\lambda(B(r,x))}$ it is enough to consider the case when μ is a finite positive measure.

Let A be a Borel set so that $\mu(A) = \lambda(A^c) = 0$. Fix $r > 0$ and set

$$F = \left\{ x \in A \cap [-r, r] \mid \limsup_{r\to 0} \frac{\mu(B(r,x))}{\lambda(B(r,x))} > 1/r \right\}.$$

Fix $\varepsilon > 0$, and take O an open set so that $A \subset O$ and $\mu(O) < \varepsilon$. Let $\mathcal{V}$ consist of all closed balls $B \subset O$ so that $\frac{\mu(B)}{\lambda(B)} > r$. Then $\mathcal{V}$ is a Vitali cover of $F \cap A$ and hence by Vitali's covering theorem there is a pairwise disjoint subcollection $\mathcal{G}$ so that $\lambda(F \setminus (\cup_{C\in\mathcal{G}} C)) = 0$.

Therefore, $\lambda(F) \le \sum_{C\in\mathcal{G}} \lambda(C) \le r \sum_{C\in\mathcal{G}} \mu(C) \le r\mu(O) < r\varepsilon$. Since $\varepsilon > 0$ was arbitrary we conclude that $\lambda(F) = 0$, and since $r > 0$ was arbitrary the result follows. ∎

4.1.3 Exercises

A net of measurable sets $\{S_\alpha\}_{\alpha\in I}$ is said to **shrink regularly** to x if

1. the diameter of S_α tends to 0, and
2. there exists $K > 0$ so that for all $\alpha \in I$, if B is the smallest ball with center x containing S_α, then $\lambda(B) \le K\lambda(S_\alpha)$.

Exercise 4.1.8. If $\{S_\alpha\}_{\alpha\in I}$ shrinks regularly to x, and if x is a Lebesgue point of $f \in L^1_{\text{loc}}(\mathbb{R}^d)$, then

$$\lim_{\alpha\to\infty} \frac{1}{\lambda(S_\alpha)} \int_{S_\alpha} |f(y) - f(x)|\, d\lambda(y) = 0.$$

Exercise 4.1.9 (The Lebesgue density topology on $\mathbb{R}$)**.** For a Lebesgue measurable set $A \subset \mathbb{R}$, set $D(A) = \{x \in \mathbb{R} \mid \lim_{x\in I, \text{diam}(I)\to 0} \frac{\lambda(A\cap I)}{\lambda(I)} = 1\}$. We define a topology on $\mathbb{R}$ by letting the open sets be all measurable sets A such that $A \subset D(A)$. Give a description of meager sets and use this to show that $\mathbb{R}$ with the Lebesgue density topology is a Baire space.

4.2 Functions of bounded variation

Lemma 4.2.1. *Suppose $f : \mathbb{R} \to \mathbb{R}$ is monotone, then f is continuous except at countably many points.*

Proof. Let $D \subset \mathbb{R}$ denote the set of points where f is discontinuous. As f is monotone it can only have jump discontinuities, so that for each $x \in D$ we have $\lim_{h\to 0^-} f(x) < \lim_{h\to 0^+} f(x)$. We let $D_n = \{x \in \mathbb{R} \mid \lim_{h\to 0^-} f(x) < \lim_{h\to 0^+} f(x) + 1/n\}$. If $F \subset D_n \cap [a,b]$ is finite then we have $f(b) - f(a) \geq \sum_{x\in F} \lim_{h\to 0^+} f(x) - \lim_{h\to 0^-} f(x) > |F|/n$. Therefore $|D_n \cap [a,b]| < \infty$ for each $n \geq 0$ and $a, b \in \mathbb{Q}$. We then have that D is countable. ■

Suppose f is a complex function which is defined in a neighborhood of a point $x_0 \in \mathbb{R}$, the **Dini numbers** associated to f at x_0 are

- $D_1 f(x_0) = \limsup_{h\to 0^+} \frac{f(x_0+h)-f(x_0)}{h}$.
- $D_2 f(x_0) = \liminf_{h\to 0^+} \frac{f(x_0+h)-f(x_0)}{h}$.
- $D_3 f(x_0) = \limsup_{h\to 0^-} \frac{f(x_0+h)-f(x_0)}{h}$.
- $D_4 f(x_0) = \liminf_{h\to 0^+} \frac{f(x_0+h)-f(x_0)}{h}$.

We say that f is **differentiable at** x_0 if $D_1 f(x_0) = D_2 f(x_0) = D_3 f(x_0) = D_4 f(x_0)$ and in this case this common value is the **derivative** $f'(x_0)$.

Theorem 4.2.2. *Suppose $f : [a,b] \to \mathbb{R}$ is monotone increasing, then f is almost everywhere (with respect to Lebesgue measure) differentiable, f' is measurable, and we have*

$$\int_a^b f' \, d\lambda \leq f(b) - f(a).$$

Proof. We obviously have $0 \leq D_2 f \leq D_1 f$ and $0 \leq D_4 f \leq D_3 f$, thus it is enough to show that at almost every point we have $D_1 \leq D_4$ and $D_3 \leq D_2$. We will show that $D_1 \leq D_4$ almost everywhere. A similar argument will apply for $D_3 \leq D_2$.

It is enough to show that for all $r, s \in \mathbb{Q}$ with $r > s > 0$ the set

$$A = A_{r,s} = \{x \in (a,b) \mid D_1 f(x) > r > s > D_4 f(x)\}$$

has measure zero.

Fix $\varepsilon > 0$ and take $O \subset (a,b)$ open so that $A \subset O$ and $\lambda(O \setminus A) \leq \varepsilon$. Let $\mathcal{V}$ denote the collection of intervals $[x-h, x]$ with $h > 0$ so that $[x-h, x] \subset O$ and

$$\frac{f(x-h) - f(x)}{-h} < s. \tag{4.4}$$

As $D_4 f(x) < s$ for $x \in A$ it follows that $\mathcal{V}$ gives a Vitali covering of A. By Vitali's covering theorem there exists $\{[x_i - h_i, x_i]\}_i \subset \mathcal{V}$ pairwise disjoint so that $\lambda((\cup_i [x_i - h_i, x_i]) \setminus A) = 0$. Since (4.4) holds we have

$$\sum_i f(x_i) - f(x_i - h_i) < s \sum_i h_i = s\lambda(\cup_i [x_i - h_i, x_i]) \leq s\lambda(O) \leq s\lambda(A) + s\varepsilon. \tag{4.5}$$

Set $B = A \cap (\cup_i (x_i - h_i, x_i))$ and note that $\lambda(B) = \lambda(A)$. Let $\mathcal{W}$ be the collection of intervals $[y, y+k]$ so that $[y, y+k] \subset (x_i - h_i, x_i)$ for some i, and such that

$$\frac{f(y+k) - f(y)}{k} > r. \tag{4.6}$$

As $D_1(x) > r$ for all $x \in B \subset A$ it follows that $\mathcal{W}$ is a Vitali covering of B. Again by Vitali's covering theorem there exists $\{[y_j, y_j + k_j]\}_j \subset \mathcal{W}$ pairwise disjoint so that $\lambda((\cup_j [y_j, y_j + k_j]) \setminus B) = 0$. Since (4.6) holds we have

$$\sum_j f(y_j + k_j) - f(y_j) > r \sum_j k_j = r\lambda(\cup_j [y_j, y_j + k_j]) \geq r\lambda(A). \tag{4.7}$$

If J_i denotes the collection of j's so that $[y_j, y_j + k_j] \subset [x_i - h_i, x_i]$ then as f is monotone increasing we have

$$\sum_{j \in J_i} f(y_j + k_j) - f(y_j) \leq f(x_i) - f(x_i - h_i).$$

Summing over all $i \in I$ and using (4.5) and (4.7) then gives

$$r\lambda(A) \leq \sum_{j \in J} f(y_j + k_j) - f(y_j) \leq \sum_{i \in I} f(x_i) - f(x_i - h_i) \leq s\lambda(A) + s\varepsilon.$$

As $\varepsilon > 0$ was arbitrary this then shows $r\lambda(A) \leq s\lambda(A)$. Since $r > s$ we conclude that $\lambda(A) = 0$.

We have therefore established that the derivative f' exists almost everywhere. We let $D \subset (a, b)$ denote the set of points where f' exists. Then D is clearly measureable and if we set $f_k : D \to [0, \infty)$ as $f_k(x) = \frac{f(x+k^{-1}) - f(x)}{k^{-1}}$, then f_k are measurable and $f_k \to f'$ pointwise which shows that f' is measurable. If $a < c < d < b$ and f is continuous at c and d then by Fatou's lemma we have

$$\begin{aligned}\int_c^d f' \, d\lambda &\leq \liminf_{k \to \infty} \int_c^d f_k \, d\lambda \\ &= \liminf_{k \to \infty} \left(k \int_d^{d+k^{-1}} f \, d\lambda - k \int_{c-k^{-1}}^c f \, d\lambda \right) \\ &= f(d) - f(c) \leq f(b) - f(a).\end{aligned}$$

By Lemma 4.2.1 we may take limits $c \to a$ and $d \to b$, which then gives

$$\int_a^b f' \, d\lambda \leq f(b) - f(a).$$

■

If $f : \mathbb{R} \to \mathbb{C}$ and $\Gamma = \{x_0, x_1, \ldots, x_n\}$ with $x_0 < x_1 < \cdots < x_n$, then consider the sum

$$S_\Gamma(f) = \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|.$$

The **variation of** f is then defined as

$$V(f) = \sup_\Gamma S_\Gamma(f)$$

If $V(f) < \infty$ then we say that f is a function of **bounded variation**. If $I \subset \mathbb{R}$ is an interval then we set $V_I(f) = \sup_{\Gamma \subset I} S_\Gamma(f)$, and say that f is a function of bounded variation on I.

For example, if $f : \mathbb{R} \to \mathbb{R}$ is monotone then f is of bounded variation on any bounded interval $[a, b]$ and we have $V_{[a,b]}(f) = |f(b) - f(a)|$, if we moreover have $\lim_{x \to -\infty} |f(x)| < \infty$ and $\lim_{x \to \infty} |f(x)| < \infty$ then f is of bounded variation on all of $\mathbb{R}$. Another example is given by $f = 1_{\{x\}}$ in which case we have $V(f) = 2$. Recall that a function is Lipschitz continuous with Lipschitz constant C if for all $x, y \in \mathbb{R}$ we have

$$|f(x) - f(y)| \leq C|x - y|.$$

It is easy to see that a Lipschitz function is of bounded variation on any bounded interval $[a, b]$, and we have

$$V_{[a,b]}(f) \leq C(b - a).$$

Proposition 4.2.3. *If f, g are functions of bounded variation on an interval I, then f and g are bounded, and both $f + g$ and fg are of bounded variation on I. Moreover, we have*

$$V_I(f + g) \leq V_I(f) + V_I(g); \qquad V_I(fg) \leq \|f\|_\infty V_I(g) + \|g\|_\infty V_I(f).$$

Also, if f is of bounded variation and $\|1/f\|_\infty < \infty$, then $1/f$ is also of bounded variation and we have

$$V_I(1/f) \leq \|1/f\|_\infty^2 V_I(f).$$

Proof. If $f(x_n) \to \infty$, then taking $\Gamma_n = \{x_1, \ldots, x_n\}$ (unordered) its easy to see that we have $S_{\Gamma_n}(f) \to \infty$. Therefore functions of bounded variation are bounded. Using the triangle inequality it is easy to see that for any finite partition Γ we have

$$S_\Gamma(f + g) \leq S_\Gamma(f) + S_\Gamma(g);$$

$$S_\Gamma(fg) \leq \|f\|_\infty S_\Gamma(g) + \|g\|_\infty S_\Gamma(f);$$

$$S_\Gamma(1/f) \leq \|1/f\|_\infty^2 S_\Gamma(f).$$

Taking suprema then gives the result. ■

Note that of course any scalar multiple of function of bounded variation is again of bounded variation, and we also have $V_I(\overline{f}) = V_I(f)$. Therefore it follows that f is of bounded variation on an interval I if and only if its real and imaginary parts are of bounded variation on I.

If $\Gamma = \{x_0, \ldots, x_n\}$ is a partition and f is real valued we set

$$P_\Gamma(f) = \sum_{i=1}^{n} (f(x_i) - f(x_{i-1}))_+;$$

$$N_\Gamma(f) = \sum_{i=1}^{n} (f(x_i) - f(x_{i-1}))_-.$$

Note that

$$P_\Gamma(f) + N_\Gamma(f) = S_\Gamma(f), \tag{4.8}$$

while

$$P_\Gamma(f) - N_\Gamma(f) = f(b) - f(a). \tag{4.9}$$

The **positive (resp. negative) variation** of f on an interval I is given by

$$P(f) = \sup_{\Gamma \subset I} P_\Gamma(f) \qquad (\text{resp. } N(f) = \sup_{\Gamma \subset I} N_\Gamma(f)\ ).$$

Proposition 4.2.4. *If any of the three of $P(f), N(f), V(f)$ are finite then all three must be finite, and in this case we have*

$$P(f) + N(f) = V(f); \qquad P(f) - N(f) = f(b) - f(a).$$

Proof. From (4.8) we see that $P(f), N(f) \leq V(f)$, so that if $V(f)$ is finite then so is $P(f)$ and $N(f)$. Also, if either of $P(f)$ or $N(f)$ is finite then from (4.9) we see that they both must be finite, and from (4.8) we see that $V(f) \leq P(f) + N(f)$ so that $V(f)$ is also finite. If we take a sequence of partitons Γ_n^1, so that $P_{\Gamma_n^1}(f) \to P(f)$, and Γ_n^2, so that $N_{\Gamma_n^2}(f) \to N(f)$ then setting $\Gamma_n = \Gamma_n^1 \cup \Gamma_n^2$ we see that $P_{\Gamma_n}(f) \to P(f)$ and $N_{\Gamma_n}(f) \to N(f)$. From (4.9) we then deduce that $P(f) - N(f) = f(b) - f(a)$, and from (4.8) we deduce that $P(f) + N(f) \leq V(f)$. ∎

Theorem 4.2.5 (Jordan decomposition)**.** *If $f : [a, b] \to \mathbb{R}$ is of bounded variation then there exist monotone increasing functions $f_1, f_2 : [a, b] \to [0, \infty)$ so that $f = f_1 - f_2$.*

Proof. Since f is of bounded variation the restriction of f to $[a, x]$ is also of bounded variation for any $a < x \leq b$. We set $f_1(x) = P_{[a,x]}(f)$ and set $f_2(x) = N_{[a,x]}(f) - c$. The functions f_1 and f_2 are increasing and from the previous proposition we have $f_1(x) - f_2(x) = f(x)$. ∎

The previous theorem can also be easily extended to unbounded intervals.

Corollary 4.2.6. *Let $f : \mathbb{R} \to \mathbb{C}$ be a function of bounded variation, then the derivative f' exists almost everywhere, and is integrable.*

Proof. This follows from the Jordan decomposition and Theorem 4.2.2. ∎

4.2.1 Exercises

Exercise 4.2.7. There is a continuous function $f : \mathbb{R} \to \mathbb{R}$ which is not of bounded variation on any interval of positive length.

Exercise 4.2.8. There is a function of bounded variation $f : \mathbb{R} \to \mathbb{R}$ which is not monotone on any interval of positive length.

4.3 Absolutely continuous and singular functions

Let I be an interval in the real line. A function $f : I \to \mathbb{C}$ is **absolutely continuous** if for all $\varepsilon > 0$ there exists $\delta > 0$ so that whenever $\{(a_i, b_i)\}_i$ is a pairwise disjoint collection of subintervals of I which satisfy $\sum_i (b_i - a_i) < \delta$ we have $\sum_i |f(b_i) - f(a_i)| < \varepsilon$. A function $f : I \to \mathbb{C}$ is **singular** if it its derivative exists and equals 0 almost everywhere.

Lemma 4.3.1. *Suppose $g : [a, b] \to \mathbb{C}$ is integrable, then $f(x) = \int_a^x g \, d\lambda$ is absolutely continuous.*

Proof. Since $|g|\lambda \ll \lambda$ this follows easily from Proposition 2.7.1. ∎

Lemma 4.3.2. *If f is absolutely continuous then f is of bounded variation on any compact interval.*

Proof. Since f is absolutely continuous there exists δ so that if $\{(a_i, b_i)\}_i$ is a pairwise disjoint collection of subintervals of I which satisfy $\sum_i (b_i - a_i) < \delta$ we have $\sum_i |f(b_i) - f(a_i)| < 1$. Then we have $V_{[a,b]}(f) \leq 1$ for any interval $[a, b] \subset I$ which satisfies $b - a < \delta$. If we take $a_1 < b_1 = a_2 < \cdots < b_N$ so that $I = \cup_{i=1}^N [a_i, b_i]$ and such that $b_i - a_i < \delta$ then we must have $V_I(f) = \sum_{i=1}^N V_{[a_i,b_i]}(f) \leq N$. ∎

Corollary 4.3.3. *If f is absolutely continuous then f is differentiable almost everywhere.*

Lemma 4.3.4. *If $f : I \to \mathbb{C}$ is absolutely continuous and singular then f is constant.*

Proof. Take $a, b \in I$, with $a < b$. Let $E = \{x \in (a, b) \in I \mid f'(x) = 0\}$, and fix $\varepsilon > 0$. Since f is singular we have $\lambda(E) = b - a$. Since f is absolutely continuous on I there exists $\delta > 0$ so that whenever $\{(a_i, b_i)\}_i$ is a pairwise disjoint collection of subintervals of I which satisfy $\sum_i (b_i - a_i) < \delta$ then we have $\sum_i |f(b_i) - f(a_i)| < \varepsilon$.

Let $\mathcal{V}$ denote the collections of intervals $[a_0, b_0] \subset (a, b)$ so that

$$|f(b_0) - f(a_0)| < (b_0 - a_0)\varepsilon/(b - a). \tag{4.10}$$

Then $\mathcal{V}$ is a Vitali covering of E and hence by Vitali's covering theorem there exists a finite pairwise disjoint collection of intervals $\{[a_i, b_i]\}_{i=1}^N \subset \mathcal{V}$ so that

$$\lambda(E \setminus (\cup_i [a_i, b_i])) < \delta.$$

Rearranging we may assume that $a_1 < b_1 \leq a_2 < b_2 \leq \cdots < b_n$. From (4.10) we have

$$\sum_{i=1}^{N} |f(b_i) - f(a_i)| \leq \varepsilon \sum_{i=1}^{N} (b_i - a_i)/(b-a) \leq \varepsilon. \tag{4.11}$$

If we set $b_0 = a$ and $a_{N+1} = b$ we then have

$$\sum_{i=0}^{N} (b_i - a_{i+1}) = \lambda(\cup_{i=0}^{N}(b_i, a_{i+1})) = \lambda([a,b] \setminus (\cup_i [x_i, x_i + h_i])) < \delta.$$

Hence

$$\sum_{i=0}^{N} |f(b_i) - f(a_{i+1})| < \varepsilon. \tag{4.12}$$

Combining (4.11) and (4.12) with the triangle inequality then gives

$$|f(b) - f(a)| \leq \sum_{i=1}^{N} |f(b_i) - f(a_i)| + \sum_{i=0}^{N} |f(b_i) - f(a_{i+1})| < 2\varepsilon.$$

As $\varepsilon > 0$ was arbitrary we then have $f(b) = f(a)$, and as $a < b$ was arbitrary it then follows that f is constant. ∎

Theorem 4.3.5. *Let $f : [a,b] \to \mathbb{C}$ be a function of bounded variation, then*

$$f_{ac}(x) = \int_a^x f' \, d\lambda$$

is absolutely continuous and $f - f_{ac}$ is singular. Moreover, if $f = g_{ac} + g_s$ where g_{ac} is absolutely continuous and g_s is singular then there exists a constant $c \in \mathbb{C}$ so that $g_{ac} = f_{ac} + c$ and $g_s = f_s - c$.

Proof. Since f is of bounded variation f' exists almost everywhere and is integrable, therefore f_{ac} is well defined. Moreover, by Lebesgue's differentiation theorem we have that f_{ac} is differentiable almost everywhere and we have $f'_{ac} = f'$ almost everywhere, hence $f - f_{ac}$ is singular.

$f = g_{ac} + g_s$ where g_{ac} is absolutely continuous and g_s is singular, then $f'_{ac} = f' = g'_{ac}$ almost everywhere and hence $f_{ac} - g_{ac}$ is an absolutely continuous function which is also singular, hence constant by the previous lemma. ∎

Corollary 4.3.6. *A function $f : [a,b] \to \mathbb{C}$ is absolutely continuous if and only if f' exists almost everywhere, is integrable, and we have*

$$f(x) - f(a) = \int_a^x f' \, d\lambda$$

for each $x \in [a,b]$.

4.3.1 Exercises

Exercise 4.3.7. Every absolutely continuous function is uniformly continuous.

Exercise 4.3.8. The Cantor function is monotone, uniformly continuous, but not absolutely continuous.

Exercise 4.3.9. Every Lipschitz continuous function is absolutely continuous.

Exercise 4.3.10. The sum and product of two absolutely continuous function on a compact interval remains absolutely continuous.

Exercise 4.3.11. Let μ be a complex Borel measure on $\mathbb{R}$ and set $f(x) = \mu((-\infty, x])$. Then f is of bounded variation. Moreover, f is absolutely continuous if and only if $\mu \ll \lambda$, and f is singular if and only if $\mu \perp \lambda$.

Chapter 5

L^p spaces

Suppose (X, μ) is a measure space, and $0 < p < \infty$. We denote by $L^p(X, \mu)$ the collection of all measurable functions $f \in \mathcal{M}(X, \mu)$ such that $|f|^p \in L^1(X, \mu)$. We identify two functions if they agree almost everywhere. Given $f \in L^p(X, \mu)$ we set

$$\|f\|_p = \left(\int |f|^p \, d\mu \right)^{1/p}.$$

We will almost exclusively be interested in the case when $p \geq 1$. When $p \geq 1$ we will show that $L^p(X, \mu)$ is a vector space, and $\| \cdot \|_p$ gives a complete norm on $L^p(X, \mu)$.

Throughout this chapter we will use the convention $\frac{1}{\infty} = 0$.

5.1 Hölder's and Minkowski's inequalities

Lemma 5.1.1 (Young's inequality)**.** *Suppose* $0 \leq a, b < \infty$*, and* $1 \leq p, q \leq \infty$ *with* $\frac{1}{p} + \frac{1}{q} = 1$*, then*

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Proof. Set $t = \frac{1}{p}$, so that $1 - t = \frac{1}{q}$. As logarithm is concave we have

$$\log(ta^p + (1-t)b^q) \geq t \log(a^p) + (1-t)\log(b^q) = \log(a) + \log(b) = \log(ab).$$

Exponentiating then gives the inequality. ■

Theorem 5.1.2 (Hölder's inequality)**.** *Let* (X, μ) *be a measure space and suppose* $1 \leq p, q \leq \infty$ *with* $\frac{1}{p} + \frac{1}{q} = 1$*, then for* $f \in L^p(X, \mu)$ *and* $g \in L^q(X, \mu)$ *we have* $fg \in L^1(X, \mu)$ *and*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q.$$

Proof. We assume that neither f nor g is essentially 0 since otherwise the inequality is trivial. Applying Young's inequality to $a = \frac{|f(x)|}{\|f\|_p}$ and $b = \frac{|g(x)|}{\|g\|_q}$ gives

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

Since $\|f\|_p^p = \int |f|^p \, d\mu$ and $\|g\|_q^q = \int |g|^q \, d\mu$, integrating the right hand side gives $\frac{1}{p} + \frac{1}{q} = 1$. It then follows that $fg \in L^1(X, \mu)$ and integrating the left hand side gives

$$\frac{\|fg\|_1}{\|f\|_p\|g\|_q} \leq 1,$$

so that Hölder's inequality follows. ∎

Theorem 5.1.3. *Let (X, μ) be a measure space and suppose $1 \leq p, q \leq \infty$ with $p < \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$, then for any $f \in L^p(X, \mu)$ we have*

$$\|f\|_p = \sup_{\|g\|_q=1} \int |fg| \, d\mu.$$

Proof. Set $L = \sup_{\|g\|_q=1} \int |fg| \, d\mu$. Then by Hölder's inequality we have $L \leq \|f\|_p$. If f is essentially 0 then the result is trivial so we assume that f is not essentially 0. Set

$$g_0 = |f|^{p-1}\overline{\text{sgn} f}.$$

If $q = \infty$ then $\|g_0\|_\infty = 1$, otherwise $\|g_0\|_q^q = \int |f|^{(p-1)q} \, d\mu = \int |f|^p \, d\mu = \|f\|_p^p$, so that $\|g_0\|_q = \|f\|_p^{p-1}$. If we set $g = \frac{g_0}{\|f\|_p^{p-1}}$ then $\|g\|_q = 1$. Hence,

$$L \geq \int fg \, d\mu = \frac{\int |f|^p \, d\mu}{\|f\|_p^{p-1}} = \|f\|_p.$$

∎

Note that by Lemma 2.7.7, the previous theorem holds in the case $p = \infty$ if and only if (X, μ) is semifinite. In the σ-finite case we also have the inequality even if $f \notin L^p(X, \mu)$:

Theorem 5.1.4 (Minkowski's inequality)**.** *Let (X, μ) be a measure space and suppose $1 \leq p \leq \infty$, then $L^p(X, \mu)$ is a vector space and $\|\cdot\|_p$ gives a norm on $L^p(X, \mu)$, i.e., for $f, g \in L^p(X, \mu)$ we have $f + g \in L^p(X, \mu)$ and*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Proof. This is just the pointwise triangle inequality for $p = \infty$ so we consider only the case when $p < \infty$. Note first that from convexity of the function $t \mapsto t^p$ we have the pointwise inequality $|f + g|^p \leq 2^{p-1}(|f|^p + |g|^p)$, so that $f + g \in L^p(X, \mu)$.

By the previous theorem if we take $1 < q \leq \infty$ so that $\frac{1}{p} + \frac{1}{q} = 1$ then we have

$$\|f+g\|_p = \sup_{\|h\|_q=1} \int |(f+g)h| \, d\mu$$
$$\leq \sup_{\|h\|_q=1} \int |fh| \, d\mu + \sup_{\|k\|_q=1} \int |gk| \, d\mu = \|f\|_p + \|g\|_p.$$

■

Theorem 5.1.5. *Let (X, μ) be a measure space and $1 \leq p \leq \infty$. Then $L^p(X, \mu)$ is a vector space and $\| \cdot \|_p$ gives a complete norm on $L^p(X, \mu)$.*

Proof. We have already shown this for $p = \infty$, and so we may assume $p < \infty$. It's enough to show that every absolutely convergent series in $L^p(X, \mu)$ actually converges in $L^p(X, \mu)$. Suppose $\{f_n\}_{n=1}^\infty \subset L^p(X, \mu)$ such that $A = \sum_{n=1}^\infty \|f_n\|_p < \infty$. Set $G_n = \sum_{k=1}^n |f_k|$, and $G = \sum_{k=1}^\infty |f_k|$. Then by Minkowski's inequality we have $\|G_n\|_p \leq \sum_{k=1}^n \|f_k\|_p \leq A$. By the monotone convergence theorem we then have $\int G^p \, d\mu = \lim_{n \to \infty} \int G_k^p \, d\mu \leq A^p$. In particular, we have that the series $\sum_{n=1}^\infty f_n$ converges almost everywhere to a function F such that $|F| \leq G$.

Then $|F - \sum_{k=1}^n f_k|^p \leq (2G)^p \in L^1(X, \mu)$, and by the dominated convergence theorem we have

$$\left\| F - \sum_{k=1}^n f_k \right\|_p^p = \int \left| F - \sum_{k=1}^n f_k \right|^p d\mu \to 0.$$

Therefore the series $\sum_{n=1}^\infty f_n$ converges to F in $L^p(X, \mu)$. ■

The next proposition extends Theorem 5.1.3 to the case when $f \notin L^p(X, \mu)$.

Proposition 5.1.6. *Let (X, μ) be a σ-finite measure space and suppose $1 \leq p, q \leq \infty$ with $q < \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$. If $f \notin L^p(X, \mu)$ then we have*

$$\sup_{\|g\|_q \leq 1} \int |fg| \, d\mu = \infty.$$

Proof. We leave the case $p = \infty$ to the reader. For $1 \leq p < \infty$ we first consider the case when (X, μ) is finite, so that if $a < \infty$ and $E = \{x \in X \mid |f(x)| \leq a\}$ then $1_E f \in L^p(X, \mu)$ and hence $1_{E^c} f \notin L^p(X, \mu)$. Therefore, we may construct a pairwise disjoint sequence of measurable sets E_n of the form $E_n = \{x \in X \mid a < |f(x)| \leq b\}$ so that $1_{E_n} f \in L^p(X, \mu)$ for each n and $\|1_{E_n} f\|_p > 4^n$.

By Theorem 5.1.3 there then exists $g_n \in L^q(X, \mu)$ with $\|g_n\|_q \leq 1$ so that $\int_{E_n} |fg_n| \, d\mu \geq 4^n$. We set $g = \sum_{n=1}^\infty 2^{-n} g_n 1_{E_n}$, so that by Minkowski's inequality we have $\|g\|_q \leq 1$. For each $n \geq 1$ we then have

$$\int |fg| \, d\mu \geq \int_{E_n} 2^{-n} |fg_n| \, d\mu \geq 2^n.$$

Hence, $\int |fg|\, d\mu = \infty$.

Now suppose (X, μ) is σ-finite and suppose $X = \cup_n X_n$ where $\{X_n\}_n$ is an increasing sequence of finite measure subsets. If $f \notin L^p(X_n, \mu)$ for some n then the result follows from the finite case above. Otherwise we have that $f \in L^p(X_n, \mu)$ for each n and $\|1_{X_n} f\|_p \to \infty$. By Theorem 5.1.3 there then exists a sequence $g_n \in L^q(X_n, \mu)$ with $\|g_n\|_q$ so that $\int |f g_n|\, d\mu \to \infty$, completing the proof. ∎

5.1.1 Minkowski's integral inequality

Minkowski's inequality shows that the L^p-norm of a sum is dominated by the sum of their L^p-norms. Generalizing from sums to integrals gives the following:

Theorem 5.1.7 (Minkowski's integral inequality). *Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces, $1 \leq p < \infty$, and $F : X \times Y \to [0, \infty)$ is $\mathcal{M} \otimes \mathcal{N}$-measurable. Then*

$$\left(\int_Y \left(\int_X F(x, y)\, d\mu(x) \right)^p d\nu(y) \right)^{1/p} \leq \int_X \left(\int_Y F(x, y)^p\, d\nu(y) \right)^{1/p} d\mu(x).$$

Proof. If $g \in L^q(Y, \nu)$ then by Tonelli's theorem and Hölder's inequality we have

$$\begin{aligned} \int \left(\int F(x, y)\, d\mu(x) \right) |g(y)|\, d\nu(y) &= \iint F(x, y) |g(y)|\, d\nu(y) d\mu(x) \\ &\leq \|g\|_q \int \left(\int F(x, y)^p\, d\nu(y) \right)^{1/p} d\mu(x). \end{aligned}$$

Proposition 5.1.6 and Theorem 5.1.3 then gives the result. ∎

5.1.2 Exercises

Exercise 5.1.8. Suppose (X, μ) is a finite measure space, then for $1 \leq p \leq q \leq \infty$ we have $L^q(X, \mu) \subset L^p(X, \mu)$.

Exercise 5.1.9. Let (X, μ) be a measure space. If $0 < p < q < r \leq \infty$ then $L^q(X, \mu) \subset L^p(X, \mu) + L^r(X, \mu)$.

Exercise 5.1.10. Let (X, μ) be a measure space. If $0 < p < q < r \leq \infty$ then $L^p(X, \mu) \cap L^r(X, \mu) \subset L^q(X, \mu)$ and if $\lambda \in [0, 1]$ satisfies $\frac{1}{q} = \lambda \frac{1}{p} + (1 - \lambda) \frac{1}{q}$ then

$$\|f\|_q \leq \|f\|_p^\lambda \|f\|_r^{1-\lambda}.$$

Hint: Apply Holder's inequality with $|f|^{\lambda q} \in L^{p/\lambda q}(X, \mu)$ and $|f|^{(1-\lambda)q} \in L^{r/(1-\lambda)q}(X, \mu)$.

Exercise 5.1.11. If X is any set and $0 < p < q \leq \infty$ then $\ell^p(X) \subset \ell^q(X)$ and $\|f\|_q \leq \|f\|_p$.

Exercise 5.1.12. If (X, μ) is a measure space with $\mu(X) = 1$, and $0 < p < q \leq \infty$ then $L^q(X, \mu) \subset L^p(X, \mu)$ and $\|f\|_p \leq \|f\|_q$.

Exercise 5.1.13. Suppose (X, μ) is a measure space and $f \in L^p(X, \mu) \cap L^\infty(X, \mu)$ for some $p < \infty$ (hence $f \in L^q(X, \mu)$ for $q \geq p$), then $\|f\|_\infty = \lim_{q \to \infty} \|f\|_q$.

Exercise 5.1.14 (Chebyshev's inequality)**.** Let (X, μ) be a measure space, $0 < p < \infty$, and $f \in L^p(X, \mu)$. Then for any $\alpha > 0$ we have

$$\mu(\{x \in X \mid |f(x)| > \alpha\}) \leq \left(\frac{\|f\|_p}{\alpha} \right)^p.$$

Exercise 5.1.15. Let (X, μ) be a measure space and suppose $1 \leq p, q, r \leq \infty$ such that $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$. If $f \in L^p(X, \mu)$ and $g \in L^q(X, \mu)$ then $fg \in L^r(X, \mu)$ and

$$\|fg\|_r \leq \|f\|_p \|g\|_q.$$

Exercise 5.1.16. Let (X, μ) be a semifinite measure space, $1 \leq p \leq \infty$, and $g \in L^\infty(X, \mu)$. The operator $M_g : L^p(X, \mu) \to L^p(X, \mu)$ given by $M_g(f) = gf$ satisfies $\|M_g\|_{\mathcal{B}(L^p(X,\mu))} = \|g\|_\infty$.

5.2 The dual of L^p-spaces

Lemma 5.2.1. *Let (X, μ) be a finite measure space and suppose $1 \leq p, q \leq \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$. If $g \in L^1(X, \mu)$ then*

$$\|g\|_p = \sup_{f \in L^\infty(X,\mu), \|f\|_q \leq 1} \left| \int fg \, d\mu \right|.$$

Proof. Replacing f with $|f| \overline{\text{sgn } g}$ we see that

$$\sup_{f \in L^\infty(X,\mu), \|f\|_q \leq 1} \left| \int fg \, d\mu \right| = \sup_{f \in L^\infty(X,\mu), \|f\|_q \leq 1} \int |fg| \, d\mu.$$

If $f \in L^q(X, \mu)$, then setting $E_k = \{x \in X \mid |f|(x) \leq k\}$ we have $1_{E_k} f \in L^\infty(X, \mu)$ with $\|1_{E_k} f\|_q \leq \|f\|_q$ and by the monotone convergence theorem we have

$$\lim_{k \to \infty} \int |1_{E_k} fg| \, d\mu = \int |fg| \, d\mu.$$

Therefore,

$$\sup_{f \in L^\infty(X,\mu), \|f\|_q \leq 1} \int |fg| \, d\mu = \sup_{f \in L^q(X,\mu), \|f\|_q \leq 1} \int |fg| \, d\mu,$$

and the result then follows from Proposition 5.1.6. ∎

Theorem 5.2.2. *Let (X, μ) be a measure space and suppose $1 < p, q < \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$. For each $g \in L^q(X, \mu)$ let $\Xi_g \in L^p(X, \mu)^*$ be given by $\Xi_g(f) = \int fg\, d\mu$. Then Ξ defines an isometric isomorphism from $L^q(X, \mu)$ onto $L^p(X, \mu)^*$.*

Proof. First, note that by Theorem 5.1.3 we have that Ξ is a well defined and isometric map, thus we only need to show that it is surjective.

Suppose $\varphi \in L^p(X, \mu)^*$. We first consider the case when μ is finite so that all simple functions are in $L^p(X, \mu)$. For each measurable set $E \subset X$ we set $\nu(E) = \varphi(1_E)$. If $\{E_n\}_{n=1}^{\infty}$ is a pairwise disjoint sequence of measurable sets and $E = \cup_{n=1}^{\infty} E_n$ then we have $1_E = \sum_{n=1}^{\infty} 1_{E_n}$ where the series converges in L^p-norm since

$$\left\| \sum_{n=K}^{\infty} 1_{E_n} \right\|_p^p = \mu\left(\cup_{n=K}^{\infty} E_n\right) \to 0.$$

Since φ is continuous we then have

$$\nu(E) = \sum_{n=1}^{\infty} \varphi(1_{E_n}) = \sum_{n=1}^{\infty} \nu(E_n).$$

Therefore ν describes a complex measure. Also, if $\mu(E) = 0$ then $1_E = 0$ in $L^p(X, \mu)$ and hence $\nu(E) = \varphi(0) = 0$, so that μ is absolutely continuous with respect to μ. By the Radon-Nikodym theorem there then exists $g \in L^1(X, \mu)$ so that $\nu(E) = \int_E g\, d\mu$ for any measurable set E, and hence for any simple function $f \in L^p(X, \mu)$ we have

$$\varphi(f) = \int fg\, d\mu. \tag{5.1}$$

If $f \in L^\infty(X, \mu)$ then taking simple functions $f_n \in L^\infty(X, \mu)$ with $\|f - f_n\|_\infty$ we have $\int f_n g\, d\mu \to \int fg\, d\mu$. Moreover, since $\|f - f_n\|_p \leq \mu(X)\|f - f_n\|_\infty$ we also have $\varphi(f_n) \to \varphi(f)$. Thus, for any $f \in L^\infty(X, \mu)$ we have $\varphi(f) = \int fg\, d\mu$ and hence

$$\sup_{f \in L^\infty(X,\mu), \|f\|_p \leq 1} \left| \int fg\, d\mu \right| = \sup_{f \in L^\infty(X,\mu), \|f\|_p \leq 1} |\varphi(f)| \leq \|\varphi\|.$$

By Lemma 5.2.1 we then have $g \in L^p(X, \mu)$ with $\|g\|_p \leq \|\varphi\|$. Since $L^\infty(X, \mu)$ is dense in $L^p(X, \mu)$, applying Hölder's inequality to approximate f in $L^p(X, \mu)$ we see that equation (5.1) holds for all $f \in L^p(X, \mu)$.

We now consider the case when μ is σ-finite, so that we may write $X = \cup_{n=1}^{n} F_n$ where F_n are increasing measurable sets of finite measure. Considering $L^p(F_n, \mu)$ as a subset of $L^p(X, \mu)$ we may then restrict φ and from the argument above it then follows that there exists $g_n \in L^q(F_n, \mu)$ so that $\varphi(f) = \int fg\, d\mu$ for all $f \in L^p(F_n, \mu)$. It's easy to check that for $m \geq n$ we have $g_n = g_{m|F_n}$, μ-almost everywhere. Thus, we may essentially define a function $g : X \to \mathbb{C}$ by letting $g_{|F_n} = g_n$. Since $\|g_n\|_q \leq \|\varphi\|$ it follows from the monotone convergence theorem that $g \in L^q(X,\mu)$. If $f \in L^p(F_n,\mu)$ then we have $\varphi(f) = \int fg\,d\mu$ and since $\cup_{n=1}^\infty L^p(F_n,\mu)$ is dense in $L^p(X,\mu)$ it then follows that $\varphi(f) = \int fg\,d\mu$ for all $f \in L^p(X,\mu)$.

Finally, we consider the general case. From above, for each σ-finite set $E \subset X$ there exists an essentially unique function $g_E \in L^q(E,\mu)$ so that $\|g_E\|_q \leq \|\varphi\|$ and $\varphi(f) = \int fg_E\,d\mu$ for any $f \in L^p(E,\mu)$. We let $M \leq \|\varphi\|$ be the supremum of $\|g_E\|_q$ as E varies over all σ-finite subsets of X, and we take a sequence E_n so that $\|g_{E_n}\|_q \to M$. Set $F = \cup_{n=1}^\infty E_n$. Then F is σ-finite, and we have $\|g_F\|_q = M$. If $E \subset F^c$ is any σ-finite set then $M \geq \|g_F + g_E\|_q \geq \|g_F\|_q = M$, and hence it follows that g_E is essentially 0. In particular, if $f \in L^p(X,\mu)$ vanishes on F it then follows that $\varphi(f) = 0 = \int fg_F\,d\mu$. Thus, we then see that for general $f \in L^p(X,\mu)$ we have $\varphi(f) = \int fg_F\,d\mu$. ■

5.2.1 Exercises

Exercise 5.2.3. Define $\varphi_n \in \ell^\infty(\mathbb{N})^*$ by $\varphi_n(f) = \frac{1}{n}\sum_{k=1}^n f(n)$, show that if φ is a weak$*$ cluster point of $\{\varphi_n\}_n$ then $\varphi \notin \ell^1(\mathbb{N})$.

If $(X,\mathcal{M})$ is a measurable space then a (complex) **finitely additive measure** on $(X,\mathcal{M})$ is a function $m : \mathcal{M} \to \mathbb{C}$, such that there exists $K > 0$ so that whenever $E_1, \dots, E_n \in \mathcal{M}$ are disjoint we have

1. $m(\cup_{k=1}^n E_k) = \sum_{k=1}^n m(E_k)$;
2. $\sum_{k=1}^n |m(E_k)| \leq K$.

If μ is a measure on $(X,\mathcal{M})$ then we say that m is **absolutely continuous** with respect to μ if $m(E) = 0$ whenever $\mu(E) = 0$.

Exercise 5.2.4. Let $(X,\mathcal{M},\mu)$ be a measure space and suppose that m is a finitely additive measure on $(X,\mathcal{M})$ which is absolutely continuous with respect to μ. There exists a unique continuous linear functional $\varphi \in L^\infty(X,\mu)^*$ so that $\varphi(1_E) = m(E)$ for all $E \in \mathcal{M}$. Moreover, every continuous linear functional on $L^\infty(X,\mu)$ arrises in this way.

Chapter 6

Functional analysis

6.1 Topological vector spaces

Suppose $\mathbb{K} = \mathbb{C}$, or $\mathbb{K} = \mathbb{R}$. A **topological $\mathbb{K}$-vector space** consists of a $\mathbb{K}$-vector space X, which is also a Hausdorff topological space such that vector addition and scalar mulitiplication give continuous maps $X \times X \to X$ and $\mathbb{K} \times X \to X$ respectively.

Examples of topological vector spaces that we have already encountered include normed spaces, as well as the duals of normed spaces enowed with the weak$*$-topology. As was the case for normed spaces, if X is a topological vector space then we may consider the space X^* consisting of all continuous linear functions into $\mathbb{K}$. The **weak$*$-topology on** X^* is the coarsest topology such that the evaluation maps $X^* \ni \varphi \mapsto \varphi(x) \in \mathbb{K}$ are continuous for each $x \in X$. The **weak-topology on** X is the coarsest topology such that the evaluation maps $X \ni x \mapsto \varphi(x) \in \mathbb{K}$ are still continuous for each $\varphi \in X^*$.

If X is a topological vector space then a net $\{x_i\}_{i \in I}$ is **Cauchy** if for any neighborhood U of 0 there exists $\alpha \in I$ so that $x_i - x_j \in U$ whenever $i, j \geq \alpha$. We say that X is **complete** if every Cauchy net converges to a point in X. A metric d is **translation invariant** if $d(x+z, y+z) = d(x, y)$ for all $x, y, z \in X$. If d is a translation invariant metric on X which is compatible with the topology then the notions of Cauchy and completeness are the same as those notions for d. In particular, if X has two translation invariant metrics, both of which give the topology, then X is complete with respect to one if and only if X is complete with respect to the other.

6.1.1 Locally convex spaces

Suppose X is a vector space over $\mathbb{K}$ and $\mathcal{F}$ is a family of seminorms on X. We say that $\mathcal{F}$ **separates points** if $\rho(x) = 0$ for all $\rho \in \mathcal{F}$ only when $x = 0$. If $\mathcal{F}$ separates points then the coarsest topology for which every $\rho \in \mathcal{F}$ is continuous gives a Hausdorff topological vector space structure to X. In this case we see that a net $\{x_i\}_i$ is Cauchy in X if and only if for each $\rho \in \mathcal{F}$ the net is Cauchy with respect to (X, ρ). We also have that $x_i \to x$ in X if and only if $\rho(x - x_i) \to 0$ for each $\rho \in \mathcal{F}$.

We now wish to find a topological characterization of those spaces whose topology arises from seminorms. If X is a topological vector space and $C \subset X$, then C is **convex** if for all $x, y \in C$, and $0 \leq t \leq 1$ we have $tx + (1-t)y \in C$. The set C is **balanced** if for all $x \in C$ and $\lambda \in \mathbb{K}$ with $|\lambda| \leq 1$ we have $\lambda x \in C$. The set C is **absorbing** if $\cup_{t \geq 0} tC = X$.

If ρ is a semi-norm on X then the unit ball $\{x \in X \mid \rho(x) < 1\}$ is a convex, balanced, absorbing, open set. In general, if C is a an absorbing set we define the **Minkowski functional** $\rho_C : X \to [0, \infty)$ by

$$\rho_C(x) = \inf\{t \geq 0 \mid x \in tC\}.$$

Proposition 6.1.1. *If X is a topological vector space and C is a convex, balanced, and open set, then the Minkowski functional is a continuous semi-norm on X, such that $C = \{x \in X \mid \rho_C(x) < 1\}$.*

Proof. First note that since C is open if $x \in X$ then $\lim n \to \infty \frac{1}{n} x = 0$ and hence $\frac{1}{n} x \in C$ for some n. Therefore C is also absorbing and so we have $\rho_C(x) < \infty$ for all $x \in X$. If $t > 0$ then it is clear that $\rho_C(tx) = t\rho_C(x)$. Also if $\alpha \in \mathbb{C}$ with $|\alpha| = 1$ then as C is balanced we have that $\rho_C(t\alpha x) = \rho_C(tx) = t\rho_C(x)$.

If $0 < a, b < \infty$ and $x \in aC$, $y \in bC$, then as C is convex we have $\frac{a}{a+b} x + \frac{b}{a+b} y \in C$, and hence $x + y \in (a+b)C$.

If $x, y \in X$, $t = \rho_C(x)$, $s = \rho_C(y)$, then for every $\varepsilon > 0$ we have $x \in (t+\varepsilon)C$ and $y \in (s+\varepsilon)C$. Therefore we see that $x + y \in (s + t + 2\varepsilon)C$. Since $\varepsilon > 0$ is arbitrary we then have $\rho_C(x+y) \leq s + t = \rho_C(x) + \rho_C(y)$.

Thus, it remains to show $C = \{x \in X \mid \rho_C(x) < 1\}$. If $\rho_C(x) < 1$ then since C is absorbing it following that $x \in C$. Conversely, if $x \in C$, then as C is open we have $tx \in C$ for t sufficiently close to 1, therefore $\rho_C(x) < 1$. ∎

Theorem 6.1.2. *Let X be a topological vector space. Then there is a family of seminorms on X which generates the topology if and only if there exists a neighborhood base at 0 consisting of convex, balanced, absorbing sets.*

Proof. If $\mathcal{F}$ is a family of seminorms which generate the topology on X then we have a neighborhood base at 0 consisting of the convex balanced and absorbing sets $\{x \in X \mid \rho(x) < a\}$ for $\rho \in \mathcal{F}$ and $a > 0$. Conversely, if $\{C_\alpha\}_\alpha$ is a family of open convex balanced and absorbing sets which give a neighborhood base at 0 then by the previous proposition the seminorms $\{\rho_{C_\alpha}\}$ are continuous and satisfy $\{x \in X \mid \rho_{C_\alpha}(x) < 1\} = C_\alpha$, hence the family generates the topology on X. ∎

A topological vector space X is **locally convex** if it satisfies one of the hypotheses of the previous theorem.

Proposition 6.1.3. *Let X be a locally convex topological vector space, then the following conditions are equivalent:*

1. *The topology on X is given by a translation invariant metric.*

2. *X is metrizable.*

3. *X is first countable.*

4. *There exists a countable family of semi-norms on X which generate the topology on X.*

5. *There exists a countable neighborhood base at 0 consisting of convex, balanced, and absorbing sets.*

Proof. The equivalence between the last two conditions above follow as in Theorem 6.1.2. Also, every metric space is first countable.

If $\{\rho_n\}_n$ is a countable family of seminorms which generate the topology then consider the translation invariant metric

$$d(x, y) = \sum_n 2^{-n} \frac{\rho_n(x-y)}{1+\rho_n(x-y)}.$$

A net converges with respect to this metric if and only if it converges with respect to each seminorm ρ_n, thus the metric d describes the topology on X.

If X is first countable then there exists a countable neighborhood base $\{U_n\}_n$ at 0. Since X is locally convex, for each n there exists a convex, balanced, and absorbing set C_n such that $C_n \subset U_n$. We then have that $\{C_n\}_n$ gives a neighborhood base at 0 which consists of convex, balanced, and absorbing sets. ■

A **Fréchet space** is a locally convex topological vector space which is has a complete translation invariant metric.

6.1.2 The open mapping and closed graph theorems

Lemma 6.1.4. *Let X be a locally convex topological vector space, Y a Fréchet space, and $T : X \to Y$ a continuous surjective linear map. Then for any neighborhood G of 0 in X we have that $\overline{T(G)}$ is a neighborhood of 0 in Y.*

Proof. Let G be a convex, balanced, absorbing neighborhood of 0 in X. Then $X = \cup_{n \geq 1} nG$ and as T is surjective we have $\cup_{n\geq 1} nT(G) = \cup_{n \geq 1} T(nG) = T(X) = Y$.

Since Y is a Fréchet space it satisfies the Baire property, and so cannot be a countable union of nowhere dense sets. Hence for some n we must have that $\overline{nT(G)}$ contains a non-empty open set O. We then have that $U = O - O$ is an open neighborhood of 0 in Y and $U \subset \overline{nT(G)} - \overline{nT(G)} \subset \overline{2nT(G)}$. Thus $V = \frac{1}{2n} U$ is an open neighborhood of 0 and $V \subset \overline{T(G)}$. ■

Theorem 6.1.5 (The open mapping theorem)**.** *Let X and Y be Fréchet spaces and $T : X \to Y$ a continuous surjective linear map, then T is an open map. In particular, if T is a bijection then T is a homeomorphism.*

Proof. Fix compatible translation invariant metrics d_X and d_Y on X and Y respectively. The map T is an open map if neighborhoods of a point x are mapped to neighborhoods of Tx, and by translation it suffices to consider the case $x = 0$. Fix $0 < r \leq 1$, then by Lemma 6.1.4 we have that $\overline{T(B_X(r,0))}$ is a neighborhood of 0, and thus it suffices to show that $\overline{T(B_X(r,0))} \subset T(B_X(2r,0))$.

Fix $y \in \overline{T(B_X(r,0))}$. For $n \geq 0$ we inductively define a sequence x_n, so that

$$x_n \in B_X(r2^{-n}, 0),$$

and

$$y - T(x_0 + x_1 + \cdots + x_n) \in B_Y(2^{-n}, 0) \cap \overline{T(B_X(r2^{-n}, 0))}.$$

Indeed, $y \in \overline{T(B_X(r,0))}$ and by Lemma 6.1.4 $\overline{T(B_X(r2^{-1},0))}$ is a neighborhood of 0, therefore there exists $x_0 \in B_X(r,0)$ so that

$$y - Tx_0 \in B_Y(1,0) \cap \overline{T(B_X(r2^{-1},0))}.$$

Now suppose that $x_0, \ldots, x_{n-1}$ have been chosen. Since

$$y - T(x_0 + x_1 + \cdots + x_{n-1}) \in \overline{T(B_X(r2^{-n},0))}$$

and since $\overline{T(B_X(r2^{-n-1},0))}$ is a neighborhood of 0 there exists $x_n \in B_X(r2^{-n},0)$ such that

$$y - T(x_0 + x_1 + \cdots + x_n) \in B_Y(2^{-n},0) \cap \overline{T(B_X(r2^{-n-1},0))}.$$

Then $\{\sum_{k=0}^n x_k\}_{n\geq 1}$ gives a Cauchy sequence and we have $y = T(\sum_{n=0}^\infty x_n)$. Since $\sum_{n=0}^\infty x_n \in B(2r,0)$ we then have that $y \in T(B(2r,0))$. ■

Corollary 6.1.6. *Let X and Y be Fréchet spaces and suppose $T : X \to Y$ is a continuous linear bijective map, then T gives an isomorphism of topological vector spaces.*

Theorem 6.1.7 (The closed graph theorem)**.** *Let X and Y be Fréchet spaces and $T : X \to Y$ a linear map such that the graph of T*

$$\mathcal{G}(T) = \{(x, Tx) \mid x \in X\} \subset X \times Y$$

is closed in the product topology, then T is continuous.

Proof. $X \times Y$ is also a Fréchet space and hence so is $\mathcal{G}(T)$, being a closed subspace. The projection map $p_X : \mathcal{G}(T) \to X$ is continuous bijective and hence by the open mapping theorem has continuous inverse. Since p_Y is also continuous it then follows that $T = p_Y \circ p_X^{-1}$ is continuous. ■

Theorem 6.1.8 (The Banach-Steinhaus uniform boundedness principle)**.** *Let X be a Banach space and Y a normed vector space. Suppose $\mathcal{F} \subset \mathcal{B}(X,Y)$ is such that for all $x \in X$ one has*

$$\sup_{T \in \mathcal{F}} \|T(x)\|_Y < \infty,$$

then

$$\sup_{T \in \mathcal{F}} \|T\|_{\mathcal{B}(X,Y)} < \infty.$$

Proof. For each $n \in \mathbb{N}$ let

$$X_n = \{x \in X \mid \sup_{T \in \mathcal{F}} \|T(x)\|_Y \leq n\}.$$

Then X_n is closed and $\cup_{n=1}^{\infty} X_n = X$. By the Baire category theorem there exists n so that X_n has non-empty interior, i.e., there exists $x_0 \in X_n$ and $\varepsilon > 0$ so that $\overline{B(\varepsilon, x_0)} \subset X_n$.

Let $y \in X$ with $\|y\| \leq 1$, and suppose $T \in \mathcal{F}$. Then

$$\begin{aligned}\|T(y)\|_Y &= \varepsilon^{-1}\|T(x_0 + \varepsilon y) - T(x_0)\|_Y \\ &\leq \varepsilon^{-1}(\|T(x_0 + \varepsilon y)\|_Y + \|T(x_0)\|_Y) \leq 2\varepsilon^{-1} n.\end{aligned}$$

Therefore $\sup_{T \in \mathcal{F}} \|T\|_{\mathcal{B}(X,Y)} \leq 2\varepsilon^{-1} n < \infty$. ■

6.1.3 Exercises

Exercise 6.1.9. Let X be a σ-compact, locally compact Hausdorff space. Consider $C(X)$ endowed with the topology of uniform convergence on compact sets. Then $C(X)$ is a Fréchet space.

Exercise 6.1.10. Let X be a vector space over $\mathbb{K}$ and suppose $\|\cdot\|_1$ and $\|\cdot\|_2$ are two complete norms on X such that there exists $C > 0$ with $\|x\|_1 \leq C\|x\|_2$ for all $x \in X$. Show that there exists $C' > 0$ so that $\|x\|_2 \leq C'\|x\|_1$ for all $x \in X$.

Exercise 6.1.11. The vector space $C^m(\mathbb{R})$ of all m-times continuously differentiable functions is a Fréchet space with the semi-norms $\|f\|_{k,n} = \sup\{|f^{(k)}(x)| \mid x \in [-n, n]\}$, for $0 \leq k \leq m$.

Exercise 6.1.12. The vector space $C^\infty(\mathbb{R})$ of all infinitely differentiable functions $f : \mathbb{R} \to \mathbb{C}$ is a Fréchet space with the semi-norms $\|f\|_{k,n} = \sup\{|f^{(k)}(x)| \mid x \in [-n, n]\}$, for $0 \leq k < \infty$.

Exercise 6.1.13. If $\{X_i\}_{i \in I}$ is a family of locally convex topological vector spaces, then $\prod_{i \in I} X_i$ with the product topology and coordinatewise operations is again a locally convex topological vector space. Moreover, if I is countable and each X_i is a Fréchet space then so is $\prod_{i \in I} X_i$.

Exercise 6.1.14. If X is a topological vector space, then X^* with the weak$*$-topology, and X with the weak-topology are also topological vector spaces.

Exercise 6.1.15. Let X be a topological vector space and suppose $\varphi \in X^*$ is not the zero functional, then φ is an open map, i.e., if $G \subset X$ is open then $\varphi(G)$ is also open.

Exercise 6.1.16. If X is a topological vector space, then $(X, \mathrm{wk})^* = X^*$.

Exercise 6.1.17. Let X be a set. The pairing between $c_0 X$ and $\ell^1 X$ given by $\langle f, g \rangle = \sum_{x \in X} f(x)\overline{g(x)}$ gives an isomorphism $\ell^1 X \cong c_0 X^*$.

Exercise 6.1.18. Let (X, μ) be a σ-finite measure space, and let $\{f_n\}_{n\in\mathbb{N}} \subset L^1(X, \mu)$ be a uniformly bounded sequence of non-negative functions. Then $f_n \to 0$ weakly if and only if $\int f_n \, d\mu \to 0$.

Exercise 6.1.19. Consider $L^\infty([0,1]) = L^1([0,1])^*$. Then $\text{span}\{e^{i2\pi nt}\}_{n\in\mathbb{Z}}$ is weak∗-dense in $L^\infty([0,1])$. Hint: Use Lusin's theorem and the Stone-Weierstrass theorem.

Exercise 6.1.20. The sequence $\{e^{i2\pi nt}\}_{n\in\mathbb{N}} \subset L^\infty([0,1])$ converges to 0 in the weak∗-topology.

Exercise 6.1.21. A topological vector space X is locally convex if and only if it is isomorphic (as topological vector spaces) to a subspace of a product of Banach spaces.

Exercise 6.1.22. If X is a Fréchet space then a subspace $Y \subset X$ is a Fréchet space if and only if Y is closed.

Exercise 6.1.23. If X is a Fréchet space then there exists a complete metric d on X which is compatible with the topology and is translation invariant, i.e., $d(x+z, y+z) = d(x, y)$ for all $x, y, z \in X$. Hint: First find a translation invariant metric which is compatible with the topology and then use the previous exercise to show that this metric is complete.

Exercise 6.1.24. Let X be a locally convex topological vector space whose topology is defined by a family of seminorms $\mathcal{F}$. If $\varphi \in X^*$ then there exist $\rho_1, \ldots, \rho_n \in \mathcal{F}$ and $K > 0$ so that $|\varphi(x)| \leq K \sum_{i=1}^n \rho_i(x)$, for all $x \in X$.

6.2 The Hahn-Banach theorem

Let X be a real vector space. A function $f : X \to \mathbb{R}$ is a **sublinear functional** if

1. $f(tx) = tf(x)$, for all $t > 0$ and $x \in X$;
2. $f(x + y) \leq f(x) + f(y)$ for all $x, y \in X$.

Note that any seminorm is a sublinear functional. Further examples are given by the following lemma whose proof we leave to the reader.

Lemma 6.2.1. *Let X be a real vector space and let $C \subset X$ be a convex and absorbing set. Then the Minkowski functional*

$$\rho_C(x) = \inf\{t > 0 \mid x \in tC\}$$

is a sublinear funcitonal, and $C = \{x \in X \mid \rho_C(x) < 1\}$.

Theorem 6.2.2 (The Hahn-Banach theorem I). *Let X be a real vector space with a sublinear functional $f : X \to \mathbb{R}$. Suppose $Y \subset X$ is a subspace, and $\varphi : Y \to \mathbb{R}$ is a linear functional such that*

$$\varphi(y) \leq f(y), \qquad y \in Y.$$

Then there exists a linear functional $\psi : X \to \mathbb{R}$ such that $\psi(y) = \varphi(y)$ for $y \in Y$, and such that

$$\psi(x) \leq f(x), \qquad x \in X.$$

Proof. We let $\mathcal{F}$ denote the set of all linear functionals $\psi : Z \to \mathbb{K}$ such that $Y \subset Z \subset X$, $\psi_{|Y} = \varphi$ and $\psi(z) \leq f(z)$ for all $z \in Z$. If $\psi_1, \psi_2 \in \mathcal{F}$ with $\psi_1 : Z_1 \to \mathbb{K}$ and $\psi_2 : Z_2 \to \mathbb{K}$ then we write $\psi_1 \prec \psi_2$ if $Z_1 \subset Z_2$ and $\psi_{2|Z_1} = \psi_1$. This then gives a partial ordering on $\mathcal{F}$. If $\{\psi_\alpha\}_{\alpha \in I}$ is an increasing chain in $\mathcal{F}$ with $\psi_\alpha : Z_\alpha \to \mathbb{K}$ then by setting $Z = \cup_\alpha Z_\alpha$ and defining $\psi : Z \to \mathbb{K}$ by $\psi(x) = \psi_\alpha(x)$ for $x \in Z_\alpha$ we have that $\psi \in \mathcal{F}$ is well defined and $\psi_\alpha \prec \psi$ for each α. Thus every chain has an upper bound and hence by Zorn's lemma there exists a maximal element $\psi \in \mathcal{F}$.

To finish the theorem it then suffices to show that any maximal element in $\mathcal{F}$ must have the entirety of X in its domain. Suppose that $\psi \in \mathcal{F}$ with $\psi : Z \to \mathbb{R}$ such that $X \neq Z$. Take $x_0 \in X \setminus Z$ and set $\tilde{Z} = \{z + \alpha x_0 \mid \alpha \in \mathbb{R}\}$. If $t \in \mathbb{R}$ then we may define a linear function $\tilde{\psi} : \tilde{Z} \to \mathbb{R}$ by $\tilde{\psi}(z + \alpha x_0) = \psi(z) + \alpha t$. In order for $\tilde{\psi}$ to belong to $\mathcal{F}$ we need to be able to choose t so that for all $z \in Z$ we have

$$\psi(z) + \alpha t \leq f(z + \alpha x_0).$$

Equivalently, for $\alpha > 0$ we need

$$-f\left(\frac{z}{\alpha} - x_0\right) + \psi\left(\frac{z}{\alpha}\right) \leq t \leq f\left(\frac{z}{\alpha} + x_0\right) - \psi\left(\frac{z}{\alpha}\right),$$

for all $z \in Z$.

Note that for $z_1, z_2 \in Z$ we have

$$\psi(z_2) + \psi(z_1) = \psi(z_2 + z_1) \leq f(z_2 + z_1) \leq f(z_2 + x_0) + f(z_1 - x_0).$$

Therefore,

$$-f(z_1 - x_0) + \psi(z_1) \leq f(z_2 + x_0) - \psi(z_2).$$

If we set $c_1 = \sup_{z_1 \in Z}\{-f(z_1 - x_0) + \psi(z_1)\}$ and we set $c_2 = \inf_{z_2 \in Z}\{f(z_2 + x_0) - \psi(z_2)\}$ then we have shown that $c_1 \leq c_2$. Taking t so that $c_1 \leq t \leq c_2$ then gives the extension $\tilde{\psi} \in \mathcal{F}$, showing that ψ was not maximal. ∎

Theorem 6.2.3 (The Hahn-Banach theorem II). *Let X be a vector space over $\mathbb{K}$ and ρ a semi-norm on X. Suppose $Y \subset X$ is a subspace, and $\varphi : Y \to \mathbb{K}$ is a linear functional such that*

$$|\varphi(y)| \leq \rho(y), \qquad y \in Y.$$

Then there exists a linear functional $\psi : X \to \mathbb{K}$ such that $\psi(y) = \varphi(y)$ for $y \in Y$, and such that

$$|\psi(x)| \leq \rho(x), \qquad x \in X.$$

Proof. We first consider the case $\mathbb{K} = \mathbb{R}$. Since ρ is a sublinear functional and $\varphi(y) \leq |\varphi(y)| \leq \rho(y)$ for $y \in Y$ it follows from the previous theorem that there exists a linear functional $\psi : X \to \mathbb{R}$ so that $\psi_{|Y} = \varphi$ and $\psi(x) \leq \rho(x)$ for $x \in X$. Considering $-x$ then shows that $-\psi(x) = \psi(-x) \leq \rho(-x) = \rho(x)$ and hence $|\psi(x)| \leq \rho(x)$ for all $x \in X$.

We now consider the case $\mathbb{K} = \mathbb{C}$. We have $|\mathrm{Re}\,(\varphi(x))| \leq |\varphi(y)| \leq \rho(y)$ and hence viewing X as a real vector space it follows from above that there exists a $\mathbb{R}$-linear functional $\psi_0 : X \to \mathbb{R}$ so that $\psi_0(y) = \mathrm{Re}\,(\varphi(y))$ for $y \in Y$, and $|\psi_0(x)| \leq \rho(x)$ for $x \in X$. Set $\psi : X \to \mathbb{C}$ by $\psi(x) = \psi_0(x) - i\psi_0(ix)$. Then ψ is $\mathbb{R}$-linear and we also have $\psi(ix) = i\psi(x)$, hence ψ is $\mathbb{C}$-linear.

Moreover, for $y \in Y$ we have $\psi(y) = \mathrm{Re}\,(\varphi(y)) - i\mathrm{Re}\,(\varphi(iy)) = \varphi(y)$. Finally, if $x \in X$ choose θ so that $e^{i\theta}\psi(x) = |\psi(x)|$. Then

$$|\psi(x)| = \psi(e^{i\theta}x) = \mathrm{Re}\,(\psi(e^{i\theta}x)) = \psi_0(e^{i\theta}x) \leq \rho(e^{i\theta}x) = \rho(x).$$

■

Corollary 6.2.4. *Let X be a normed space, then the map $\iota : X \to X^{**}$ given by $\iota(x)(\varphi) = \varphi(x)$ is isometric.*

Proof. Fix $x \in X$ and define $\varphi : \mathbb{K}x \to \mathbb{K}$ by $\varphi(\alpha x) = \alpha\|x\|$. Then φ is linear and we have $|\varphi(\alpha x)| = \|\alpha x\|$. By the Hahn-Banach theorem there then exists a linear functional $\psi : X \to \mathbb{K}$ so that $\psi(x) = \|x\|$, and $|\psi(z)| \leq \|z\|$ for all $z \in X$, i.e., $\|\psi\| \leq 1$. Therefore we see $\|\iota(x)\| \geq |\iota(x)(\psi)| = |\psi(x)| = \|x\|$. The reverse inequality is trivial. ■

Lemma 6.2.5. *Let X be a vector space over $\mathbb{K}$ and suppose $\varphi, \varphi_1, \ldots, \varphi_n$ are linear functionals on X such that $\cap_{k=1}^n \ker(\varphi_k) \subset \ker(\varphi)$. Then we have $\varphi \in \mathrm{span}\{\varphi_1, \ldots, \varphi_n\}$.*

Proof. We may assume that $\{\varphi_1, \ldots, \varphi_n\}$ is linearly independent. Set $L = \cap_{k=1}^n \ker(\varphi_k)$. Then $\varphi, \varphi_1, \ldots, \varphi_n$ all give well defined linear functionals on X/L which has dimension at most n. Since $\{\varphi_1, \ldots, \varphi_n\} \subset (X/L)^*$ is a set of n linearly independent vectors it follows that $\{\varphi_1, \ldots, \varphi_n\}$ also spans $(X/L)^*$. Therefore, $\varphi \in \mathrm{span}\{\varphi_1, \ldots, \varphi_n\}$. ■

Proposition 6.2.6. *If X is a locally convex topological vector space, then the map $\Xi : X \to (X^*, \mathrm{wk}^*)^*$ given by $\Xi_x(\varphi) = \varphi(x)$ is bijective.*

Proof. If $x \in X$, $x \neq 0$ then there exists a continuous semi-norm ρ so that $\rho(x) > 0$. If we consider $\varphi : \mathbb{K}x \to \mathbb{K}$ given by $\varphi(\alpha x) = \alpha\rho(x)$, then as in Corollary 6.2.4 we may apply the Hahn-Banach theorem to produce a linear functional $\psi : X \to \mathbb{K}$ so that $\psi(x) = \rho(x)$ and $|\psi(z)| \leq \rho(z)$ for all $z \in X$. Hence ψ is continuous and $\Xi(x)(\psi) = \psi(x) = \rho(z) \neq 0$. Therefore Ξ is injective.

If $\zeta \in (X^*, \mathrm{wk}^*)^*$ then there exit $K > 0$ and $x_1, \ldots, x_n \in X$ so that $|\zeta(\varphi)| \leq K\sum_{i=1}^n |\varphi(x_i)|$, for all $\varphi \in X^*$. In particular we have $\ker(\zeta) \subset \cap_{i=1}^n \ker(\Xi(x_i))$, and it follows from the previous lemma that $\zeta = \Xi(x)$ for some $x \in \mathrm{span}\{x_1, \ldots, x_n\}$. ■

6.2.1 Separating convex sets

If X is a topological vector space over $\mathbb{K}$ and $A, B \subset X$, then A and B are **separated** if there exists a continuous linear functional $\varphi \in X^*$, and $\alpha \in \mathbb{R}$ so that

$$\mathrm{Re}\,(\varphi(a)) \leq \alpha \leq \mathrm{Re}\,(\varphi(b)), \qquad a \in A, b \in B \tag{6.1}$$

If the inequalities in Equation (6.1) may be taken to be strict then we say that A and B are strictly separated. Note, that if A and B are (strictly) separated then the convex sets they generate are also (strictly) separated.

Lemma 6.2.7. *Let X be a topological vector space over $\mathbb{K}$. If $G \subset X$ is convex open, and $x_0 \notin G$, then G and x are separated.*

Proof. We first consider the case $\mathbb{K} = \mathbb{R}$ and we assume that G is nonempty. Taking $g \in G$ and replacing G with $G - g$ and x_0 with $x_0 - g$ it is enough to consider the case when $0 \in G$. Since G is open and contains 0 we have that G is absorbing and hence by Lemma 6.2.1 the Minkowski functional $\rho_G(y) = \inf\{t > 0 \mid y \in tG\}$ is sublinear and satisfies $G = \{y \in X \mid \rho_G(y) < 1\}$. It is also easy to see that ρ_G is continuous.

Define $\varphi : \mathbb{R}x_0 \to \mathbb{R}$ by $\varphi(\alpha x_0) = \alpha\rho_G(x_0) \leq \rho_G(\alpha x_0)$. By the Hahn-Banach theorem there then exists a linear functional $\psi : X \to \mathbb{R}$ so that $\psi(\alpha x_0) = \alpha\rho_G(x_0)$ and $\psi(z) \leq \rho_G(z)$ for all $z \in X$. If $z_\alpha \to z$ then we have $\rho_G(z_\alpha - z) \to 0$ and hence $\limsup_{\alpha\to\infty} \psi(z_\alpha - z) \leq 0$. Since we also have $\rho_G(z - z_\alpha) \to 0$ we also obtain $\liminf_{\alpha\to\infty} \psi(z_\alpha - z) \geq 0$ and hence $\lim_{\alpha\to\infty} \psi(z_\alpha - z) = 0$. Therefore ψ is continuous. Finally, for $x \in G$ we have

$$\psi(x) \leq \rho_G(x) < 1 \leq \rho_G(x_0) = \psi(x_0).$$

We now consider the case $\mathbb{K} = \mathbb{C}$. Treating X as a $\mathbb{R}$-vector space it follows from above that there exists a continuous $\mathbb{R}$-linear functional ψ and $\alpha \in \mathbb{R}$ so that for all $x \in G$ we have $\psi(x) \leq \alpha \leq \psi(x_0)$. Setting $\tilde{\psi}(x) = \psi(x) - i\psi(ix)$ then gives the result. ■

Proposition 6.2.8. *Let X be a topological vector space over $\mathbb{K}$. If $G, H \subset X$ are disjoint convex sets such that G is open, then G and H are separated. Moreover, if H is also open then G and H are strictly separated.*

Proof. Consider the set $G - H = \{x - y \mid x \in G, y \in H\}$. Since G and H are convex it follows that $G - H$ is also convex. Moreover, as G and H are disjoint we have that $0 \notin G - H$. Writing $G - H = \cup_{y\in H}(G - y)$ we see that $G - H$ is open. Thus, by Lemma 6.2.7 we can separate $G - H$ and 0 by some linear functional φ, i.e., replacing φ with $-\varphi$ if needed, we have $0 \leq \mathrm{Re}\,(\varphi(x - y))$ for all $x \in G$ and $y \in H$. This then shows that $\mathrm{Re}\,(\varphi(y)) \leq \mathrm{Re}\,(\varphi(x))$ for all $x \in G$ and $y \in H$, and hence G and H are separated.

If H is also open then by Exercise 6.1.15 we have that $\mathrm{Re}\,(\varphi(G))$ and $\mathrm{Re}\,(\varphi(H))$ are disjoint open sets in $\mathbb{R}$ showing that G and H are strictly separated. ■

Theorem 6.2.9 (The Hahn-Banach separation theorem)**.** *Let X be a locally convex topological vector space over $\mathbb{K}$. If $K, F \subset X$ are disjoint closed convex sets and K is compact, then K and F are strictly separated.*

Proof. Since F is closed and X is locally convex for each $k \in K$ there exists a balanced convex open neighborhood U_k of 0 so that $(k + U_k) \cap F = \emptyset$. The family $\{k + \frac{1}{2}U_k\}_{k \in K}$ gives an open over of K and so by compactness there exists a finite subcover $\{k_i + \frac{1}{2}U_{k_i}\}_{i=1}^n$. Set $U = \cap_{i=1}^n \frac{1}{4}U_{k_i}$. Then $K + U$ and $F + U$ are convex open sets.

We claim that $K + U$ and $F + U$ are disjoint. Indeed if $k + u = f + v$ with $k \in K$ and $u, v \in U$ then we have $k \in (k_i + \frac{1}{2}U_{k_i})$ for some i and so

$$f = k + u - v \in \left(k_i + \frac{1}{2}U_{k_i}\right) + \frac{1}{4}U + \frac{1}{4}U \subset (k_i + U_{k_i}) \subset F^c.$$

Hence $(K + U) \cap (F + U) = \emptyset$ and thus these sets are strictly separated by Proposition 6.2.8. Since $K \subset K + U$ and $F \subset F + U$ it then follows that K and F are strictly separated. ■

Note that the hypothesis that K is compact is necessary. Indeed, even in $\mathbb{R}^2$ the closed sets $\{(t, 0) \mid t > 0\}$ and $\{(t, t^{-1}) \mid t > 0\}$ cannot be strictly separated.

Corollary 6.2.10. *Let X be a $\mathbb{K}$-vector space and suppose that $\mathcal{T}_1$ and $\mathcal{T}_2$ are topologies on X giving the structure of a locally convex topological vector space and such that $(X, \mathcal{T}_1)$ and $(X, \mathcal{T}_2)$ have the same continuous linear functionals. Then a convex set $C \subset X$ is closed in the $\mathcal{T}_1$-topology if and only if C is closed in the $\mathcal{T}_2$-topology.*

Proof. Suppose $C \subset X$ is convex and closed in the $\mathcal{T}_1$-topology. For each point $x \notin C$ it follows from the Hahn-Banach separation theorem that x and C are strictly separated. Thus, there exists a $\mathcal{T}_1$-continuous linear functional φ_x and $\alpha \in \mathbb{R}$ so that $\mathrm{Re}\,(\varphi_x(y)) \leq \alpha$ for all $y \in C$ and $\mathrm{Re}\,(\varphi_x(x)) > \alpha$. We then have $C = \cap_{x \in X}\{y \in X \mid \mathrm{Re}\,(\varphi_x(y)) \leq \alpha\}$. Since the two topologies have the same continuous linear functionals it then follows that C is also closed in the $\mathcal{T}_2$-topology. The converse also holds by symmetry. ■

Corollary 6.2.11. *Let X be a locally convex topological vector space and suppose $C \subset X$ is a convex set, then C is closed if and only if C is weakly closed. In particular, a subspace $Y \subset X$ is closed if and only if it is weakly closed.*

Proof. Since X with its given topology and X with the weak topology have the same continuous linear functionals this follows from the previous corollary. ■

Recall from Corollary 6.2.4 that for a normed space X we have a natural isometric embedding $\iota : X \to X^{**}$ given by $\iota(x)(\varphi) = \varphi(x)$. We will therefore identify X as a subspace of X^{**}.

Proposition 6.2.12. *Let X be a normed space, then the unit ball of X is weak*-dense in the unit ball of X^{**}. In particular, X is weak*-dense in X^{**}.*

Proof. Let B be the unit ball of X and let C be the weak*-closure of B in X^{**}. By Proposition 6.2.6 any weak*-continuous linear functional on X^{**} is of the form $\eta \mapsto \eta(\varphi)$ for some $\varphi \in X^*$. Therefore by the Hahn-Banach separation theorem if $\zeta \in X^{**}$ is not in C then there exists $\varphi \in X^*$ and $\alpha \in \mathbb{R}$ so that for any $x \in B$ we have $\mathrm{Re}\,(\varphi(x)) < \alpha < \mathrm{Re}\,(\zeta(\varphi))$. Since $0 \in B$ we have $0 < \alpha$.

Since $e^{i\theta}B = B$ it then follows that $|\varphi(x)| < \alpha$ for each $x \in B$ and hence $\|\varphi\| \leq \alpha$. Therefore $\alpha < \mathrm{Re}\,(\zeta(\varphi)) \leq |\zeta(\varphi)| \leq \|\zeta\|\|\varphi\| \leq \alpha\|\zeta\|$, and we conclude that $\|\zeta\| > 1$. By contraposition it then follows that C agrees with the unit ball in X^{**}. ■

A Banach space X is **reflexive** if $X^{**} = X$. For example, Theorem 5.2.2 shows that if (X, μ) is a measure space and $1 < p < \infty$, then $L^p(X, \mu)$ is reflexive.

Theorem 6.2.13. *Let X be a Banach space, then the following conditions are equivalent:*

1. *X is reflexive.*
2. *The weak and weak* topologies on X^* agree.*
3. *X^* is reflexive.*
4. *The closed unit ball in X is weakly compact.*

Proof. By Proposition 6.2.12 the unit ball of X is weak*-dense in the unit ball of X^{**} in the weak*-topology, which agrees with the weak topology on X. Therefore if the unit ball of X is weakly compact we must have that the unit ball of X is equal to the unit ball of X^{**} and hence it follows that $X = X^{**}$. This then shows that (4) $\implies$ (1). While the Banach-Alaoglu theorem gives (1) $\implies$ (4).

We also clearly have (3) $\implies$ (2). And (2) together with the Banach-Alaoglu theorem shows that the closed unit ball of X^* is weakly compact which then shows that X^* is reflexive from the implication (4) $\implies$ (1) applied to X^*. Thus we see that (2) and (3) are also equivalent.

To complete the theorem it then suffices to show (3) $\implies$ (1) as (1) $\implies$ (3) would then follow by considering X^{**}. Suppose therefore that (3) (and hence also (2)) holds. If $\zeta \in X^{**}$ then ζ is continuous with respect to the weak topology on X^*. Since the weak and weak* topologies agree we then have that ζ is continuous with respect to the weak* topology and hence $\zeta \in X$ by Proposition 6.2.6. This then shows that X is reflexive. ■

6.2.2 The Krein-Milman theorem

If X is a $\mathbb{K}$-vector space and $C \subset X$ is a nonempty convex set, then a point $k \in C$ is an **extreme point** if it cannot be written as a non-trivial convex combination of elements in C, i.e., if $x, y \in C$ $x, y \neq k$ and $t \in [0, 1]$ then $tx + (1-t)y \neq k$. More generally, we say that a nonempty subset $K \subset C$ is an extreme subset if whenever $x, y \in C \setminus K$ and $t \in [0,1]$ then we have $tx + (1-t)y \notin K$. We denote by $\text{ext}(C)$ the set of extreme points of C. If $F \subset X$ we let $\overline{\text{co}}(F)$ denote the smallest closed convex set which contains F.

Lemma 6.2.14. *Let X be a vector space over $\mathbb{K}$ and suppose $C \subset X$ is a nonempty convex set, φ is a linear functional and $\alpha \in \mathbb{R}$ so that $K = C \cap \{x \in X \mid \text{Re}\,(\varphi(x)) \leq \alpha\}$ is nonempty. Then K is an extreme subset of C.*

Proof. Suppose $x, y \in C \setminus K$ and $t \in [0,1]$ then $\varphi(tx + (1-t)y) = t\varphi(x) + (1-t)\varphi(y) > t\alpha + (1-t)\alpha = \alpha$. ■

Theorem 6.2.15 (The Krein-Milman theorem)**.** *Suppose K is a compact convex subset of a locally convex topological vector space over $\mathbb{K}$, then $K = \overline{\text{co}}(\text{ext}K)$.*

Proof. We consider the family $\mathcal{F}$ consisting of compact convex extreme subsets of C which we consider ordered by decreasing inclusion. If $\{K_\alpha\}_\alpha$ is a chain of elements in $\mathcal{F}$ then set $K = \cap_\alpha K_\alpha$. Then K is convex and by compactness we have that K is nonempty. It is also easy to see that K is an extreme subset. Thus any chain has an upper bound and by Zorn's lemma there then exists a nonempty convex extreme set K which has no proper subset with this property. If $x, y \in K$ with $x \neq y$ then by the Hahn-Banach theorem there exists a continuous linear functional φ so that $\text{Re}\,(\varphi(x)) < \text{Re}\,(\varphi(y))$. We would then have that $K \cap \{z \in X \mid \text{Re}\,(\varphi(z)) \leq \text{Re}\,(\varphi(x))\}$ is a non-empty convex compact extreme subset which does not contain y contradicting minimality of K. Thus we conclude that K consists of a single point. and so $\text{ext}(C) \neq \emptyset$.

Let $B = \overline{\text{co}}(\text{ext}(C))$ and suppose $B \neq C$ so that there exists $x \in C \setminus B$. As B is a closed convex set from the Hahn-Banach separation theorem there exists a continuous linear functional φ and $\alpha \in \mathbb{R}$ so that $\text{Re}\,(\varphi(x)) \leq \alpha$ and $\text{Re}\,(\varphi(y)) > \alpha$ for all $y \in B$. We then have that $C \cap \{x \in X \mid \text{Re}\,(\varphi(x)) \leq \alpha\}$ is a nonempty convex compact extreme subset which is disjoint from B. However the argument above shows that $C \cap \{x \in X \mid \text{Re}\,(\varphi(x)) \leq \alpha\}$ must contain an extreme point and hence cannot be disjoint from B. Therefore we conclude that $B = C$. ■

6.2.3 Exercises

Exercise 6.2.16. Let X be a Banach space and suppose $\{x_n\}_{n=1}^\infty \subset X$ is a sequence which converges weakly, then $\{x_n\}_{n=1}^\infty$ is bounded.

Exercise 6.2.17. Let (X, μ) be a σ-finite measure space. Then $L^1(X, \mu)$ is reflexive if and only if $L^1(X, \mu)$ is finite dimensional.

Exercise 6.2.18. Let X be a Fréchet space, then X^* is a Fréchet space in the weak*-topology if and only if X is isomorphic to a separable Banach space. Hint: If $\{\rho_n\}_{n \in \mathbb{N}}$ is a family of seminorms which give the topology on X, first show that $A_n = \{\varphi \in X^* \mid |\varphi(x)| \leq n \sum_{k=1}^n \rho_k(x) \text{ for all } x \in X\}$ is weak*-compact by the Banach-Alaoglu theorem, then show that $X = \cup_n A_n$ and use the Baire property.

Exercise 6.2.19 (The Markov-Kakutani fixed point theorem)**.** Let $K \subset X$ be a non-empty compact convex subset of a locally convex space X, and suppose $\mathcal{S}$ is a non-empty family of pairwise commuting continuous maps which are affine, i.e., $T(tk_1 + (1-t)k_2) = tT(k_1) + (1-t)T(k_2)$ whenever $T \in \mathcal{S}$, $k_1, k_2 \in K$ and $t \in [0,1]$. Then there is a point in K which is a common fixed points for all maps in $\mathcal{S}$. Hint: First consider the case when $\mathcal{S}$ consists of a single map T, take $k_0 \in K$ and consider the sequence $\frac{1}{n}\sum_{i=0}^{n-1} T^i k_0$.

6.3 Hilbert space

6.3.1 Inner product spaces

Let $\mathcal{H}$ be a vector space over $\mathbb{K}$, where $\mathbb{K} = \mathbb{R}$, or $\mathbb{K} = \mathbb{C}$. An **inner product** on $\mathcal{H}$ is a map $(\xi, \eta) \mapsto \langle \xi, \eta \rangle$ from $\mathcal{H} \times \mathcal{H} \to \mathbb{K}$ so that

1. $\langle \xi, \xi \rangle \in (0, \infty)$ for all nonzero $\xi \in \mathcal{H}$.
2. $\langle \alpha\xi + \eta, \zeta \rangle = \alpha\langle \xi, \zeta \rangle + \langle \eta, \zeta \rangle$, for $\alpha \in \mathbb{K}$, $\xi, \eta, \zeta \in \mathcal{H}$.
3. $\langle \xi, \eta \rangle = \overline{\langle \eta, \xi \rangle}$, for $\xi, \eta \in \mathcal{H}$.

Observe that from the last two conditions above we also have $\langle \xi, \alpha\eta + \zeta \rangle = \overline{\alpha}\langle \xi, \eta \rangle + \langle \xi, \zeta \rangle$ for $\alpha \in \mathbb{K}$, and $\xi, \eta, \zeta \in \mathcal{H}$. Given $\xi \in \mathcal{H}$ we define

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

An **inner product space** is a vector space together with an inner product on that space. As an example, suppose (X, μ) is a measure space, and consider $L^2(X, \mu) = \{f \in M(X, \mu) \mid |f|^2 \in L^1(X, \mu)\}$ where we identity two functions which agree almost everywhere. From the inequalities

$$|ab| \leq |a|^2 + |b|^2, \qquad |a+b|^2 \leq 2(|a|^2 + |b|^2), \qquad a, b \in \mathbb{C},$$

we deduce that for $f, g \in L^2(X, \mu)$ we have $\overline{g}f \in L^1(X, \mu)$, and $f + g \in L^2(X, \mu)$. Therefore, $L^2(X, \mu)$ is a vector space and we obtain an inner product by setting

$$\langle f, g \rangle = \int \overline{g} f \, d\mu.$$

Proposition 6.3.1 (The Cauchy-Schwarz inequality)**.** *Let $\mathcal{H}$ be an inner product space, then for all $\xi, \eta \in \mathcal{H}$ we have*

$$|\langle \xi, \eta \rangle| \leq \|\xi\| \|\eta\|.$$

Proof. For any $\lambda \in \mathbb{K}$, $\xi_0, \eta_0 \in \mathcal{H}$ we have

$$\|\xi_0\|^2 + 2\mathrm{Re}\,(\lambda\langle \xi_0, \eta_0 \rangle) + |\lambda|^2 \|\eta_0\|^2 = \|\xi_0 + \lambda\eta_0\|^2 \geq 0.$$

Taking λ so that $|\lambda| = 1$ and $\lambda\langle \xi_0, \eta_0 \rangle \leq 0$ gives

$$\|\xi_0\|^2 + \|\eta_0\|^2 \geq 2|\langle \xi_0, \eta_0 \rangle|.$$

Setting $\xi_0 = \|\eta\|\xi$ and $\eta_0 = \|\xi\|\eta$ gives

$$2\|\xi\|^2\|\eta\|^2 \geq 2\|\xi\|\|\eta\||\langle \xi, \eta\rangle|,$$

and the inequality follows. ■

Proposition 6.3.2. *Let $\mathcal{H}$ be an inner product space, then the map $\mathcal{H} \ni \xi \mapsto \|\xi\|$ defines a norm on $\mathcal{H}$.*

Proof. The only nontrivial thing to check is the triangle inequality. Suppose $\xi, \eta \in \mathcal{H}$, then from the Cauchy-Schwarz inequality we have

$$\|\xi + \eta\|^2 = \|\xi\|^2 + 2\mathrm{Re}\,\langle \xi, \eta\rangle + \|\eta\|^2 \leq \|\xi\|^2 + 2\|\xi\|\|\eta\| + \|\eta\|^2 = (\|\xi\| + \|\eta\|)^2.$$

■

Lemma 6.3.3. *Let $\mathcal{H}$ be an inner product space, then the inner product is jointly continuous with respect to the topology induced by the norm.*

Proof. Suppose $\xi_n \to \xi$ and $\eta_n \to \eta$, then $\{\eta_n\}_{n\in\mathbb{N}}$ and the Cauchy-Schwarz inequality gives

$$\begin{aligned} |\langle \xi_n, \eta_n\rangle - \langle \xi, \eta\rangle| &\leq |\langle \xi_n - \xi, \eta_n\rangle| + |\langle \xi, \eta_n - \eta\rangle| \\ &\leq \|\xi_n - \xi\|\|\eta_n\| + \|\xi\|\|\eta_n - \eta\| \to 0. \end{aligned}$$

■

A **Hilbert space** is an inner product space which is complete with respect to the given norm. For example, if (X, μ) is any measure space then by Theorem 5.1.5 we have that $L^2(X, \mu)$ is a Hilbert space with inner product $\langle f, g\rangle = \int \overline{g} f \, d\mu$.

Proposition 6.3.4 (The parallelogram identity)**.** *Let $\mathcal{H}$ be an inner product space, then for $\xi, \eta \in \mathcal{H}$ we have*

$$\|\xi + \eta\|^2 + \|\xi - \eta\|^2 = 2(\|\xi\|^2 + \|\eta\|^2).$$

Proof. Just add the formulas $\|\xi \pm \eta\|^2 = \|\xi\|^2 \pm 2\mathrm{Re}\,\langle \xi, \eta\rangle + \|\eta\|^2$. ■

The next proposition shows that the norm completely determines the inner product.

Proposition 6.3.5 (The polarization identity)**.** *Let $\mathcal{H}$ be a complex inner-product space and suppose μ is a measure on the circle $\mathbb{T}$ so that $\mu(\mathbb{T}) = 1$, and $\int \lambda \, d\mu(\lambda) = \int \lambda^2 \, d\mu(\lambda) = 0$. Then for $\xi, \eta \in \mathcal{H}$ we have*

$$\int \lambda \|\xi + \lambda\eta\|^2 \, d\mu(\lambda) = \langle \xi, \eta\rangle$$

Proof. We may compute directly

$$\begin{aligned}\int \lambda\|\xi + \lambda\eta\|^2 \, d\mu(\lambda) &= \int \lambda\|\xi\|^2 \, d\mu(\lambda) + \int \langle \xi, \eta\rangle \, d\mu(\lambda) \\ &\quad + \int \lambda^2 \langle \eta, \xi\rangle \, d\mu(\lambda) + \int \lambda\|\eta\|^2 \, d\mu(\lambda) \\ &= \langle \xi, \eta\rangle.\end{aligned}$$

■

The most commonly used case of the polarization identity is when $\mu = \frac{1}{4}\left(\delta_{\{1\}} + \delta_{\{i\}} + \delta_{\{-1\}} + \delta_{\{-i\}}\right)$, in which case we obtain the formula

$$\langle \xi, \eta\rangle = \frac{1}{4}\sum_{k=0}^{3} i^k \|\xi + i^k\eta\|^2.$$

Another case is if we take Lebesgue measure on the interval $[0, 1]$ and we consider the corresponding Lebesgue measure on the circle, which is the push-forward under the map $t \mapsto e^{2\pi i t}$.

Corollary 6.3.6. *Let $\mathcal{H}$ and $\mathcal{K}$ be two complex inner product spaces. A linear map $U : \mathcal{H} \to \mathcal{K}$ is isometric if and only if $\langle U\xi, U\eta\rangle = \langle \xi, \eta\rangle$ for all $\xi, \eta \in \mathcal{H}$.*

Proof. If $U : \mathcal{H} \to \mathcal{K}$ is isometric, and $\xi, \eta \in \mathcal{H}$, then by the polarization identity we have

$$\langle U\xi, U\eta\rangle = \frac{1}{4}\sum_{k=0}^{3} i^k \|U(\xi + i^k\eta)\|^2 = \frac{1}{4}\sum_{k=0}^{3} i^k \|(\xi + i^k\eta)\|^2 = \langle \xi, \eta\rangle.$$

■

The previous result is also valid for real inner product spaces, and we leave it as an exercise.

6.3.2 Orthogonal subspaces and the Riesz representation theorem

Given an inner product space $\mathcal{H}$, two vectors $\xi, \eta \in \mathcal{H}$ are **orthogonal** if $\langle \xi, \eta\rangle = 0$. A set $\{\xi_i\}_{i \in I}$ is **orthogonal** if $\langle \xi_i, \xi_j\rangle = 0$ for $i \neq j$, and $\{\xi_i\}_{i \in I}$ is **orthonormal** if it is orthogonal and we also have $\|\xi_i\| = 1$ for all $i \in I$.

If $A \subset \mathcal{H}$ we set

$$A^\perp = \{\xi \in \mathcal{H} \mid \langle \xi, \eta\rangle = 0 \text{ for all } \eta \in A\}.$$

Theorem 6.3.7. *Let $\mathcal{H}$ be a Hilbert space, $K \subset \mathcal{H}$ a nonempty closed convex subset, and $\eta_0 \in \mathcal{H}$. Then there exists a unique element $\xi_0 \in K$ with minimal distance to η_0.*

Proof. By considering $\tilde{K} = K - \eta_0$ it suffices to consider the case when $\eta_0 = 0$. Set $d = \inf\{\|\xi\| \mid \xi \in K\}$, and choose a sequence $\xi_n \in \mathcal{K}$ such that $\|\xi_n\| \to d$. Then for $n, m \in \mathbb{N}$ we have

$$d^2 \leq \left\| \frac{1}{2}\xi_n + \frac{1}{2}\xi_m \right\|^2 = \frac{1}{4}\|\xi_n\|^2 + \frac{1}{2}\mathrm{Re}\,\langle \xi_n, \xi_m \rangle + \frac{1}{4}\|\xi_m\|^2.$$

Hence, $\lim_{n,m\to\infty} \mathrm{Re}\,\langle \xi_n, \xi_m \rangle = d^2$.

Therefore,

$$\begin{aligned}\lim_{n,m\to\infty} \|\xi_n - \xi_m\|^2 &= \lim_{n,m\to\infty} \|\xi_n\|^2 - 2\mathrm{Re}\,\langle \xi_n, \xi_m \rangle + \|\xi_m\|^2 \\ &= d^2 - 2d^2 + d^2 = 0.\end{aligned}$$

Hence $\{\xi_n\}_{n\in\mathbb{N}}$ is Cauchy and converges to vector $\xi_0 \in K$, which satisfies $\|\xi_0\| = d$. Since the sequence $\{\xi_n\}_{n\in\mathbb{N}}$ was chosen arbitrary it follows that ξ_0 must be unique. ■

Let $\mathcal{H}$ be a Hilbert space and suppose $\mathcal{K} \subset \mathcal{H}$ is a closed subspace. If $\xi \in \mathcal{H}$ we let $P_{\mathcal{K}}(\xi)$ denote the unique vector in $\mathcal{K}$ with minimal distance to ξ and we call the map $P_{\mathcal{K}}$ the **orthogonal projection** from $\mathcal{H}$ onto $\mathcal{K}$.

Proposition 6.3.8. *Let $\mathcal{H}$ be a Hilbert space, $\mathcal{K} \subset \mathcal{H}$ a closed subspace and fix $\xi \in \mathcal{H}$. Then $P_{\mathcal{K}}(\xi)$ is the unique vector in $\mathcal{K}$ such that $\xi - P_{\mathcal{K}}(\xi) \in \mathcal{K}^\perp$.*

Proof. As $P_{\mathcal{K}}(\xi)$ minimizes the distance to ξ it follows that for all $a > 0$ and $\eta \in \mathcal{K}$ we have

$$\|\xi - P_{\mathcal{K}}(\xi)\|^2 \leq \|\xi - P_{\mathcal{K}}(\xi) - a\eta\|^2 = \|\xi - P_{\mathcal{K}}(\xi)\|^2 - 2a\mathrm{Re}\,\langle \xi - P_{\mathcal{K}}(\xi), \eta \rangle + a^2\|\eta\|^2.$$

Rearranging and dividing by a then gives

$$\mathrm{Re}\,\langle \xi - P_{\mathcal{K}}(\xi), \eta \rangle \leq a\|\eta\|^2.$$

As $a > 0$ was arbitrary this then shows that $\mathrm{Re}\,\langle \xi - P_{\mathcal{K}}(\xi), \eta \rangle \leq 0$, and replacing η with $i^k\eta$ for $k = 0, 1, 2, 3$ then shows that $\langle \xi - P_{\mathcal{K}}(\xi), \eta \rangle = 0$. Therefore $\xi - P_{\mathcal{K}}(\xi) \in \mathcal{K}^\perp$.

Conversely, suppose that $\eta \in \mathcal{K}$ is such that $\xi - \eta \in \mathcal{K}^\perp$. Then for $\zeta \in \mathcal{K}$ we have $\langle \xi - \eta, \zeta \rangle = 0$, hence

$$\|\xi - \eta - \zeta\|^2 = \|\xi - \eta\|^2 + \|\zeta\|^2 \geq \|\xi - \eta\|^2.$$

It then follows that $\eta - P_{\mathcal{K}}(\xi)$. ■

Corollary 6.3.9. *Let $\mathcal{H}$ be a Hilbert space, $\mathcal{K} \subset \mathcal{H}$ a closed subspace, then $P_{\mathcal{K}}$ is a linear map and $\|P_{\mathcal{K}}\| \leq 1$.*

Proof. If $\xi, \eta \in \mathcal{H}$, $\alpha \in \mathbb{C}$ and $\zeta \in \mathcal{K}$ then by the previous proposition we have

$$\langle \alpha\xi + \eta - \alpha P_{\mathcal{K}}(\xi) - P_{\mathcal{K}}(\eta), \zeta \rangle = \alpha\langle \xi - P_{\mathcal{K}}(\xi), \zeta\rangle + \langle \eta - P_{\mathcal{K}}(\eta), \zeta\rangle = 0.$$

Hence, again by the previous proposition we have $\alpha P_{\mathcal{K}}(\xi) + P_{\mathcal{K}}(\eta) = P_{\mathcal{K}}(\alpha\xi + \eta)$, which shows that $P_{\mathcal{K}}$ is linear.

Also, since $P_{\mathcal{K}}(\xi)$ and $\xi - P_{\mathcal{K}}(\xi)$ are orthogonal we have $\|P_{\mathcal{K}}(\xi)\|^2 \leq \|P_{\mathcal{K}}(\xi)\|^2 + \|\xi - P_{\mathcal{K}}(\xi)\|^2 = \|\xi\|^2$ so that $P_{\mathcal{K}}$ is a contraction. ■

Theorem 6.3.10 (Riesz representation theorem)**.** *Let $\mathcal{H}$ be a Hilbert space, and for each $\eta \in \mathcal{H}$ consider the map $\Xi_\eta \in \mathcal{H}^*$ given by $\Xi_\eta(\xi) = \langle \xi, \eta\rangle$. Then $\Xi : \mathcal{H} \to \mathcal{H}^*$ gives an isometric anti-linear surjection.*

Proof. Clearly Ξ is anti-linear. By the Cauchy-Schwarz inequality we have $|\Xi_\eta(\xi)| = |\langle \xi, \eta\rangle| \leq \|\xi\|\|\eta\|$ which shows that Ξ is a contraction. Moreover $\Xi_\eta(\eta) = \|\eta\|^2$ which then shows that Ξ is isometric.

Suppose now that we have $\varphi \in \mathcal{H}^*$, and assume that $\varphi \neq 0$. Then $\ker(\varphi)$ is a proper closed subspace. Take $\eta \in \mathcal{H}$ so that $\varphi(\eta) = 1$ and by replacing η with $\eta - P_{\ker(\varphi)}(\eta)$ we assume that $\eta \in \ker(\varphi)^\perp$.

If $\xi \in \mathcal{H}$, then $\xi - \varphi(\xi)\eta \in \ker(\varphi)$ and hence is orthogonal to η. Thus,

$$0 = \langle \xi - \varphi(\xi)\eta, \eta\rangle = \langle \xi, \eta\rangle - \varphi(\xi)\|\eta\|^2.$$

Thus, $\varphi(\xi) = \langle \xi, \eta\|\eta\|^{-2}\rangle$ for all $\xi \in \mathcal{H}$ which shows that Ξ is surjective. ■

6.3.3 Orthonormal bases and dimension

Note that if $\{\xi_1, \ldots, \xi_n\}$ is a finite orthonormal set then expanding the inner product gives

$$\left\| \sum_{i=1}^n \alpha_i \xi_i \right\|^2 = \sum_{i\in I} |\alpha_i|^2.$$

We will use this equality throughout this section.

Proposition 6.3.11 (Bessel's inequality)**.** *Let $\mathcal{H}$ be a Hilbert space and suppose $\{\xi_i\}_{i\in I}$ is an orthonormal set, then for any $\eta \in \mathcal{H}$ we have*

$$\sum_{i\in I} |\langle \eta, \xi_i\rangle|^2 \leq \|\eta\|^2.$$

In particular, $\{i \in I \mid \langle \eta, \xi_i\rangle \neq 0\}$ is countable.

Proof. If $\{\xi_1, \ldots, \xi_n\}$ is an orthonormal set and $\mathcal{K}$ denotes it's span, then for $\eta \in \mathcal{H}$ set $\eta_0 = \sum_{i=1}^n \langle \eta_0, \xi_i\rangle \xi_i$. For $1 \leq j \leq n$ we then have

$$\langle \eta - \eta_0, \xi_j\rangle = \langle \eta, \xi_j\rangle - \sum_{i=1}^n \langle \eta_0, \xi_i\rangle\langle \xi_i, \xi_j\rangle = \langle \eta, \xi_j\rangle - \langle \eta, \xi_j\rangle = 0.$$

Thus, by Proposition 6.3.8 we have $\mathcal{P}_{\mathcal{K}}(\eta) = \eta_0 = \sum_{i=1}^n \langle \eta, \xi_i\rangle \xi_i$.

In particular, we have $\sum_{i=1}^n |\langle \eta, \xi_i\rangle|^2 = \|\mathcal{P}_{\mathcal{K}}(\eta)\|^2 \leq \|\eta\|^2$. Thus, Bessel's inequality holds for finite sets and the general case then follows easily. ■

If $\mathcal{H}$ is a Hilbert space then an orthonormal set $\{\xi_i\}_{i\in I}$ is an **orthonormal basis** if 0 is the only vector which is orthogonal to every ξ_i. For example, if X is a set then the family of Dirac functions $\{\delta_x\}_{x\in X}$ forms an orthonormal basis.

Proposition 6.3.12 (Parseval's identity)**.** *Suppose $\mathcal{H}$ is a Hilbert space with orthonormal basis $\{\xi_i\}_{i\in I}$, then for $\eta \in \mathcal{H}$ we have*

$$\|\eta\|^2 = \sum_{i\in I} |\langle \eta, \xi_i\rangle|^2,$$

and $\eta = \sum_{i\in I}\langle \eta, \xi_i\rangle \xi_i$, where the sum converges absolutely in $\mathcal{H}$.

Proof. Given $\eta \in \mathcal{H}$ set $\eta_0 = \sum_{i\in I}\langle \eta, \xi_i\rangle\xi_i$ and note that this sum converges absolutely in $\mathcal{H}$ by Bessel's inequality.

For $j \in I$ we have $\langle \eta - \eta_0, \xi_j\rangle = \langle \eta, \xi_j\rangle - \langle \eta, \xi_j\rangle = 0$. Since $\{\xi_i\}_{i\in I}$ is an orthonormal basis we then have $\eta = \eta_0 = \sum_{i\in I}\langle \eta, \xi_i\rangle\xi_i$. By approximating η by finite sums it then follows that

$$\|\eta\|^2 = \sum_{i\in I} |\langle \eta, \xi_i\rangle|^2.$$

■

Theorem 6.3.13. *Every Hilbert space has an orthonormal basis. Moreover, any two orthonormal bases have the same cardinality.*

Proof. Let $\mathcal{H}$ be a Hilbert space. By Zorn's lemma it follows easily that $\mathcal{H}$ has a maximal (with respect to inclusion) orthonormal set $\{\xi_i\}_{i\in I}$. Suppose $\eta \in \mathcal{H}$ such that $\langle \eta, \xi_i\rangle = 0$ for all $i \in I$. If $\eta \neq 0$ then the set $\{\eta\|\eta\|^{-1}\} \cup \{\xi_i\}_{i\in I}$ would be an orthonormal set which is strictly larger and hence would contradict Zorn's lemma. Thus, we must have $\eta = 0$ and hence $\{\xi_i\}_{i\in I}$ is an orthonormal basis.

To see that any two bases have the same cardinality we consider separately the finite and infinite cases. In the finite case an orthonormal basis is also an algebraic basis and this is then a standard fact from abstract linear algebra which we will not present here.

Suppose therefore that $\{\xi_i\}_{i\in I}$ and $\{\eta_j\}_{j\in J}$ are two infinite orthonormal bases. By Bessel's inequality to each $i \in I$ the set of j's such that $\langle \xi_i, \eta_j\rangle \neq 0$ is a non-empty countable set. Since every η_j is not orthogonal to some ξ_i it follows that there exists a surjective map from $I \times \mathbb{N}$ onto J. Since I is infinite we have $|I| = |I \times \mathbb{N}| \geq |J|$. By symmetry we also have that $|J| \geq |I|$ and hence $|I| = |J|$. ■

If $\mathcal{H}$ is a Hilbert space then the **dimension** of $\mathcal{H}$ is the cardinality of any orthonormal basis, and denoted by $\dim \mathcal{H}$. If $\mathcal{H}$ and $\mathcal{K}$ are Hilbert spaces then a **unitary operator** is a surjective linear isometry $U : \mathcal{H} \to \mathcal{K}$.

Theorem 6.3.14. *If $\mathcal{H}$ is a Hilbert space, and X is a set then* $\dim \mathcal{H} = |X|$ *if and only if there exists a unitary operator $U : \ell^2(X) \to \mathcal{H}$.*

Proof. If $U : \ell^2(X) \to \mathcal{H}$ is a unitary operator then $\{U\delta_x\}_{x\in X}$ gives an orthonormal basis with cardinality $|X|$.

Conversely, if $\mathcal{H}$ has an orthonormal basis $\{\xi_i\}_{i\in I}$ with cardinality $|I| = |X|$, then there exists a bijection $\theta : X \to I$. We define an operator $U : \ell^2 X \to \mathcal{H}$, by setting $U(f) = \sum_{x\in X} f(x)\xi_{\theta(x)}$. It's then an easy calculation to see that U is a unitary operator. ■

6.3.4 Exercises

Exercise 6.3.15. If we consider $\mathbb{N}$ with the counting measure, then $\ell^2(\mathbb{N})$ is a Hilbert space.

Exercise 6.3.16. Let X be a real normed space with norm $\|\cdot\|^2$. Then $\|\cdot\|^2$ comes from an inner product if and only if the parallelogram identity $\|\xi+\eta\|^2 + \|\xi-\eta\|^2 = 2(\|\xi\|^2 + \|\eta\|^2)$ holds.

Exercise 6.3.17. Let $\mathcal{H}$ and $\mathcal{K}$ be two real inner product spaces. A linear map $U : \mathcal{H} \to \mathcal{K}$ is isometric if and only if $\langle U\xi, U\eta\rangle = \langle \xi, \eta\rangle$ for all $\xi, \eta \in \mathcal{H}$.

Exercise 6.3.18 (The Banach-Saks theorem)**.** Let $\mathcal{H}$ be a Hilbert space and $\{\xi_n\}_{n\in\mathbb{N}} \subset \mathcal{H}$ a uniformly bounded sequence, then there exists a subsequence $\{\xi_{n_k}\}_k$ so that the Cesàro means $\frac{1}{K}\sum_{k=1}^K \xi_{n_k}$ converges in $\mathcal{H}$. Hint: Using the Banach-Alaoglu theorem you may assume that ξ_n has a weak limit.

Exercise 6.3.19. Let $\mathcal{H}$ be an inner product space, and $A \subset \mathcal{H}$, then $A^\perp$ is a closed subspace.

Exercise 6.3.20. If $\mathcal{H}$ is a Hilbert space and $A \subset \mathcal{H}$ is a subspace then $(A^\perp)^\perp = \overline{A}$. This does not hold for general inner product spaces.

Exercise 6.3.21. Let $\mathcal{H}$, and $\mathcal{K}$ be Hilbert spaces and suppose $T : \mathcal{H} \to \mathcal{K}$ is a bounded linear operator, then there exists a unique bounded linear operator $T^* : \mathcal{K} \to \mathcal{H}$ so that for all $\xi \in \mathcal{H}$ and $\eta \in \mathcal{K}$ we have

$$\langle T\xi, \eta\rangle = \langle \xi, T^*\eta\rangle.$$

The operator $T^* : \mathcal{K} \to \mathcal{H}$ is called the **adjoint** of the operator T.

Exercise 6.3.22. Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces.

1. A bounded linear operator $P \in \mathcal{B}(\mathcal{H})$ is an orthogonal projection operator if and only if $P = P^*$ and $P^2 = P$.

2. A bouned operator $U \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ is isometric if and only if $U^*U = \text{id}$.

Exercise 6.3.23. Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces, and suppose $T \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. Then $\ker(T) = \text{Range}(T^*)^\perp$.

A linear operator $V \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ is a **partial isometry** if V^*V is an orthogonal projection.

Exercise 6.3.24. Suppose $V \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ is a partial isometry.

1. V^*V is the orthogonal projection onto $\ker(V)^\perp$.
2. Range(V) is closed and VV^* is the orthogonal projection onto Range(V). In particular, V^* is also a partial isometry.

Exercise 6.3.25. Let M be a closed subspace of $L^2([0,1], \lambda)$ such that M is contained in $C([0,1])$.

1. There exists $C > 0$ such that $\|f\|_\infty \leq C\|f\|_{L^2}$ for all $f \in M$.
2. For each $x \in [0,1]$ there exists $g_x \in M$ so that $f(x) = \langle f, g_x \rangle$ for all $f \in M$. Moreover, $\|g_x\|_{L^2} \leq C$.
3. $\dim M \leq C^2$. Hint: If $\{f_i\}_i$ is an orthonormal sequence in M then $\sum_i |f_i(x)|^2 \leq C^2$ for all $x \in [0,1]$.

Exercise 6.3.26 (Von Neumann's mean ergodic theorem)**.** Let U be a unitary operator on a Hilbert space $\mathcal{H}$, set $\mathcal{K} = \{\xi \in \mathcal{H} \mid U\xi = \xi\}$, and let P denote the orthogonal projection onto $\mathcal{K}$. If $S_n = \frac{1}{n}\sum_{k=0}^{n-1} U^k$ then for all $\xi \in \mathcal{U}$ we have $S_n\xi \to P\xi$. Hint: Use Exercise 6.3.23 applied to the operator $1 - U$.

Exercise 6.3.27. Let (X, μ) be a finite measure space and fix $f \in L^\infty(X, \mu)$. Set $a_n = \int |f|^n \, d\mu$.

1. The sequence $\frac{a_{n+1}}{a_n}$ is non-decreasing as n increases. Hint: Apply the Cauchy-Schwarz inequality for $\xi = |f|^{n/2}$ and $\eta = |f|^{(n+2)/2}$.
2. The sequence $\frac{a_{n+1}}{a_n}$ converges to $\|f\|_\infty$. Hint: Show that the series $\sum_{n=1}^\infty b^n|f|^n$ converges absolutely in $L^1(X, \mu)$ if $b < \|f\|_\infty^{-1}$, and diverges in $L^1(X, \mu)$ if $b > \|f\|_\infty^{-1}$.

Index

Bibliography

[hg] Andrey Gogolev (http://mathoverflow.net/users/2029/andrey gogolev), *If f is infinitely differentiable then f coincides with a polynomial*, MathOverflow, URL:http://mathoverflow.net/q/34067 (version: 2011-11-01).