Notes Following Folland's *Real Analysis*

Greyson C. Wesley's Notes on Folland

April 30, 2024

Abstract

These notes follow a course in real analysis from 2023–2024 at The Ohio State University. We follow [Fol99] very closely, and at times even exactly.

Contents

0 Preliminaries

The purpose of this introductory chapter is to establish the notation and terminology that will be used throughout the book and to present a few diverse results from set theory and analysis that will be needed later. The style here is deliberately terse, since this chapter is intended as a reference rather than a systematic exposition.

0.1 The Language of Set Theory

It is assumed that the reader is familiar with the basic concepts of set theory; the following discussion is meant mainly to fix our terminology.

Number Systems. Our notation for the fundamental number systems is as follows:

$$\mathbb{Z}_{\geqslant 1} = \text{ the set of positive integers (not including zero)}$$
$$\mathbb{Z} = \text{ the set of integers}$$
$$\mathbb{Q} = \text{ the set of rational numbers}$$
$$\mathbb{R} = \text{ the set of real numbers}$$
$$\mathbb{C} = \text{ the set of complex numbers}$$

Logic. We shall avoid the use of special symbols from mathematical logic, preferring to remain reasonably close to standard English. We shall, however, sometimes use the abbreviation iff for "if and only if."

One point of elementary logic that is often insufficiently appreciated by students is the following: If A and B are mathematical assertions and $-A, -B$ are their negations, the statement " A implies B" is logically equivalent to the contrapositive statement " $-B$ implies $-A$." Thus one may prove that A implies B by assuming $-B$ and deducing $-A$, and we shall frequently do so. This is not the same as *reductio ad absurdum*, which consists of assuming both A and $-B$ and deriving a contradiction.

Sets. The words "family" and "collection" will be used synonymously with "set," usually to avoid phrases like "set of sets." The empty set is denoted by $\varnothing$, and the family of all subsets of a set X is denoted by $\mathcal{P}(X)$:

$$\mathcal{P}(X) = \{E \mid E \subset X\}$$

Here and elsewhere, the inclusion sign $\subset$ is interpreted in the weak sense; that is, the assertion " $E \subset X$" includes the possibility that $E = X$.

Remark 1 (Human language conversion)**.** *What follows in this remark is an excerpt from Professor Nicolaescu's course notes from the past few years. Suppose that we are given a family of subsets $(S_i)_{i\in I}$ of a set Ω. Let us observe that the statement*

$$\omega \in \bigcap_{i\in I} S_i$$

translates into the formula $\forall i \in I, \omega \in S_i$ or, in human language, ω belongs to all the sets in the family. The statement

$$\omega \in \bigcup_{i\in I} S_i$$

translates into the formula $\exists i \in I, \omega \in S_i$ or, in human language, ω belongs to at least one of the sets S_i. For example a statement of the form

$$\omega \in \bigcup_{n\in\mathbb{Z}_{\geqslant 1}} \bigcap_{k\geqslant n} S_k$$

translates into

$$\exists n \in \mathbb{Z}_{\geqslant 1}, \quad \forall k \geqslant n, \quad \omega \in S_k.$$

Equivalently, this means that ω belongs to all but finitely many of the sets S_k. Conversely, statements involving the quantifiers $\exists, \forall$ can be translated into set theoretic statements using the conversion rules

$$\exists \to \cup, \quad \forall \to \cap.$$

If $\mathcal{E}$ is a family of sets, we can form the union and intersection of its members:

$$\bigcup_{E\in\mathcal{E}} E = \{x : x \in E \text{ for some } E \in \mathcal{E}\}$$
$$\bigcap_{E\in\mathcal{E}} E = \{x : x \in E \text{ for all } E \in \mathcal{E}\}$$

Usually it is more convenient to consider indexed families of sets:

$$\mathcal{E} = \{E_\alpha \mid \alpha \in A\} = \{E_\alpha\}_{\alpha \in A},$$

in which case the union and intersection are denoted by

$$\bigcup_{\alpha \in A} E_\alpha, \quad \bigcap_{\alpha \in A} E_\alpha$$

If $E_\alpha \cap E_\beta = \varnothing$ whenever $\alpha \neq \beta$, the sets E_α are called disjoint. The terms "disjoint collection of sets" and "collection of disjoint sets" are used interchangeably, as are "disjoint union of sets" and "union of disjoint sets."

When considering families of sets indexed by $\mathbb{Z}_{\geqslant 1}$, our usual notation will be

$$\{E_n\}_{n=1}^\infty \text{ or } \{E_n\}_1^\infty$$

and likewise for unions and intersections. In this situation, the notions of limit superior and limit inferior are sometimes useful:

$$\limsup E_n = \bigcap_{k=1}^\infty \bigcup_{n=k}^\infty E_n, \quad \liminf E_n = \bigcup_{k=1}^\infty \bigcap_{n=k}^\infty E_n$$

The reader may verify that

$\limsup E_n = \{x \mid x \in E_n \text{ for infinitely many } n\}$,

$\liminf E_n = \{x \mid x \in E_n \text{ for all but finitely many } n\}$.

If E and F are sets, we denote their difference by $E \setminus F$:

$$E \setminus F = \{x \mid x \in E \text{ and } x \notin F\}$$

and their symmetric difference by $E \triangle F$:

$$E \triangle F = (E \setminus F) \cup (F \setminus E).$$

When it is clearly understood that all sets in question are subsets of a fixed set X, we define the complement E^c of a set E (in X):

$$E^c = X \setminus E.$$

In this situation we have deMorgan's laws:

$$\left(\bigcup_{\alpha \in A} E_\alpha\right)^c = \bigcap_{\alpha \in A} E_\alpha^c, \quad \left(\bigcap_{\alpha \in A} E_\alpha\right)^c = \bigcup_{\alpha \in A} E_\alpha^c$$

If X and Y are sets, their Cartesian product $X \times Y$ is the set of all ordered pairs (x, y) such that $x \in X$ and $y \in Y$. A relation from X to Y is a subset of $X \times Y$. (If $Y = X$, we speak of a relation on X.) If R is a relation from X to Y, we shall sometimes write xRy to mean that $(x, y) \in R$. The most important types of relations are the following: - Equivalence relations. An equivalence relation on X is a relation R on X such that

$$xRx \text{ for all } x \in X$$
$$xRy \text{ iff } yRx$$
$$xRz \text{ whenever } xRy \text{ and } yRz \text{ for some } y.$$

The equivalence class of an element x is $\{y \in X \mid xRy\}$.X is the disjoint union of these equivalence classes. - Orderings. See Folland Section 0.2. - Mappings. A mapping $f \colon X \to Y$ is a relation R from X to Y with the property that for every $x \in X$ there is a unique $y \in Y$ such that xRy, in which case we write $y = f(x)$. Mappings are sometimes called maps or functions; we shall generally reserve the latter name for the case when Y is $\mathbb{C}$ or some subset thereof.

If $f\colon X \to Y$ and $g\colon Y \to Z$ are mappings, we denote by $g \circ f$ their composition:

$$g \circ f\colon X \to Z, \quad g \circ f(x) = g(f(x))$$

If $D \subset X$ and $E \subset Y$, we define the image of D and the inverse image of E under a mapping $f\colon X \to Y$ by

$$f(D) = \{f(x) \mid x \in D\}, \quad f^{-1}(E) = \{x \mid f(x) \in E\}$$

It is easily verified that the map $f^{-1}\colon \mathcal{P}(Y) \to \mathcal{P}(X)$ defined by the second formula commutes with union, intersections, and complements:

$$f^{-1}\Big(\bigcup_{\alpha\in A} E_\alpha\Big) = \bigcup_{\alpha\in A} f^{-1}(E_\alpha), \quad f^{-1}\Big(\bigcap_{\alpha\in A} E_\alpha\Big) = \bigcap_{\alpha\in A} f^{-1}(E_\alpha),$$
$$f^{-1}(E^c) = (f^{-1}(E))^c.$$

(The direct image mapping $f\colon \mathcal{P}(X) \to \mathcal{P}(Y)$ commutes with unions, but in general not with intersections or complements.)

If $f\colon X \to Y$ is a mapping, X is called the domain of f and $f(X)$ is called the range of $f. f$ is said to be injective if $f(x_1) = f(x_2)$ only when $x_1 = x_2$, surjective if $f(X) = Y$, and bijective if it is both injective and surjective. If f is bijective, it has an inverse $f^{-1}\colon Y \to X$ such that $f^{-1} \circ f$ and $f \circ f^{-1}$ are the identity mappings on X and Y, respectively. If $A \subset X$, we denote by $f|A$ the restriction of f to A:

$$(f|A) : A \to Y, \quad (f|A)(x) = f(x) \text{ for } x \in A$$

A sequence in a set X is a mapping from $\mathbb{Z}_{\geqslant 1}$ into X. (We also use the term finite sequence to mean a map from $\{1, \ldots, n\}$ into X where $n \in \mathbb{Z}_{\geqslant 1}$.) If $f\colon \mathbb{Z}_{\geqslant 1} \to X$ is a sequence and $g\colon \mathbb{Z}_{\geqslant 1} \to \mathbb{Z}_{\geqslant 1}$ satisfies $g(n) < g(m)$ whenever $n < m$, the composition $f \circ g$ is called a subsequence of f. It is common, and often convenient, to be careless about distinguishing between sequences and their ranges, which are subsets of X indexed by $\mathbb{Z}_{\geqslant 1}$. Thus, if $f(n) = x_n$, we speak of the sequence $\{x_n\}_1^\infty$; whether we mean a mapping from $\mathbb{Z}_{\geqslant 1}$ to X or a subset of X will be clear from the context.

Earlier we defined the Cartesian product of two sets. Similarly one can define the Cartesian product of n sets in terms of ordered n-tuples. However, this definition becomes awkward for infinite families of sets, so the following approach is used instead. If $\{X_\alpha\}_{\alpha\in A}$ is an indexed family of sets, their Cartesian product $\prod_{\alpha\in A} X_\alpha$ is the set of all maps $f\colon A \to \bigcup_{\alpha\in A} X_\alpha$ such that $f(\alpha) \in X_\alpha$ for every $\alpha \in A$. (It should be noted, and then promptly forgotten, that when $A = \{1, 2\}$, the previous definition of $X_1 \times X_2$ is set-theoretically different from the present definition of $\prod_1^2 X_j$. Indeed, the latter concept depends on mappings, which are defined in terms of the former one.) If $X = \prod_{\alpha\in A} X_\alpha$ and $\alpha \in A$, we define the αth projection or coordinate map $\pi_\alpha\colon X \to X_\alpha$ by $\pi_\alpha(f) = f(\alpha)$. We also frequently write x and x_α instead of f and $f(\alpha)$ and call x_α the αth coordinate of x.

If the sets X_α are all equal to some fixed set Y, we denote $\prod_{\alpha\in A} X_\alpha$ by Y^A:

$$Y^A = \text{ the set of all mappings from } A \text{ to } Y.$$

If $A = \{1, \ldots, n\}, Y^A$ is denoted by Y^n and may be identified with the set of ordered n-tuples of elements of Y.

0.2 Orderings

A partial ordering on a nonempty set X is a relation R on X with the following properties:

- if xRy and yRz, then xRz;
- if xRy and yRx, then $x = y$;
- xRx for all x.

If R also satisfies

- if $x, y \in X$, then either xRy or yRx,

then R is called a linear (or total) ordering. For example, if E is any set, then $\mathcal{P}(E)$ is partially ordered by inclusion, and $\mathbb{R}$ is linearly ordered by its usual ordering. Taking this last example as a model, we shall usually denote partial orderings by $\leqslant$, and we write $x < y$ to mean that $x \leqslant y$ but $x \neq y$. We observe that a partial ordering on X naturally induces a partial ordering on every nonempty subset of X. Two partially ordered sets X and Y are said to be order isomorphic if there is a bijection $f\colon X \to Y$ such that $x_1 \leqslant x_2$ iff $f(x_1) \leqslant f(x_2)$.

If X is partially ordered by $\leqslant$, a maximal (resp. minimal) element of X is an element $x \in X$ such that the only $y \in X$ satisfying $x \leqslant y$ (resp. $x \geqslant y$) is x itself. Maximal and minimal elements may or may not exist, and they need not be unique unless the ordering is linear. If $E \subset X$, an upper (resp. lower) bound for E is an element $x \in X$ such that $y \leqslant x$ (resp. $x \leqslant y$) for all $y \in E$. An upper bound for E need not be an element of E, and unless E is linearly ordered, a maximal element of E need not be an upper bound for E. (The reader should think up some examples.)

If X is linearly ordered by $\leqslant$ and every nonempty subset of X has a (necessarily unique) minimal element, X is said to be well ordered by $\leqslant$, and (in defiance of the laws of grammar) $\leqslant$ is called a well ordering on X. For example, $\mathbb{Z}_{\geqslant 1}$ is well ordered by its natural ordering.

We now state a fundamental principle of set theory and derive some consequences of it.

Theorem 0.2: 0.1: The Hausdorff Maximal Principle.

Every partially ordered set has a maximal linearly ordered subset.

In more detail, this means that if X is partially ordered by $\leqslant$, there is a set $E \subset X$

that is linearly ordered by $\leqslant$, such that no subset of X that properly includes E is linearly ordered by $\leqslant$. Another version of this principle is the following:

Theorem 0.3: 0.2: Zorn's Lemma.

If X is a partially ordered set and every linearly ordered subset of X has an upper bound, then X has a maximal element.

Clearly the Hausdorff maximal principle implies Zorn's lemma: An upper bound for a maximal linearly ordered subset of X is a maximal element of X. It is also not difficult to see that Zorn's lemma implies the Hausdorff maximal principle. (Apply Zorn's lemma to the collection of linearly ordered subsets of X, which is partially ordered by inclusion.)

Theorem 0.4: 0.3: The Well Ordering Principle.

Every nonempty set X can be well ordered.

Proof. Let $\mathcal{W}$ be the collection of well orderings of subsets of X, and define a partial ordering on $\mathcal{W}$ as follows. If $\leqslant_1$ and $\leqslant_2$ are well orderings on the subsets E_1 and E_2, then $\leqslant_1$ precedes $\leqslant_2$ in the partial ordering if (i) $\leqslant_2$ extends $\leqslant_1$, i.e., $E_1 \subset E_2$ and $\leqslant_1$ and $\leqslant_2$ agree on E_1, and (ii) if $x \in E_2 \setminus E_1$ then $y \leqslant_2 x$ for all $y \in E_1$. The reader may verify that the hypotheses of Zorn's lemma are satisfied, so that $\mathcal{W}$ has a maximal element. This must be a well ordering on X itself, for if $\leqslant$ is a well ordering on a proper subset E of X and $x_0 \in X \setminus E$, then $\leqslant$ can be extended to a well ordering on $E \cup \{x_0\}$ by declaring that $x \leqslant x_0$ for all $x \in E$. □

Theorem 0.5: 0.4: The Axiom of Choice.

If $\{X_\alpha\}_{\alpha \in A}$ is a nonempty collection of nonempty sets, then $\prod_{\alpha \in A} X_\alpha$ is nonempty.

Proof. Let $X = \bigcup_{\alpha \in A} X_\alpha$. Pick a well ordering on X and, for $\alpha \in A$, let $f(\alpha)$ be the minimal element of X_α. Then $f \in \prod_{\alpha \in A} X_\alpha$. □

Corollary 0.6: 0.5.

If $\{X_\alpha\}_{\alpha \in A}$ is a disjoint collection of nonempty sets, there is a set $Y \subset \bigcup_{\alpha \in A} X_\alpha$ such that $Y \cap X_\alpha$ contains precisely one element for each $\alpha \in A$.

Proof. Take $Y = f(A)$ where $f \in \prod_{\alpha \in A} X_\alpha$. □

We have deduced the axiom of choice from the Hausdorff maximal principle; in fact, it can be shown that the two are logically equivalent.

0.3 Cardinality

If X and Y are nonempty sets, we define the expressions

$$\operatorname{card}(X) \leqslant \operatorname{card}(Y), \quad \operatorname{card}(X) = \operatorname{card}(Y), \quad \operatorname{card}(X) \geqslant \operatorname{card}(Y)$$

to mean that there exists $f \colon X \to Y$ which is injective, bijective, or surjective, respectively. We also define

$$\operatorname{card}(X) < \operatorname{card}(Y), \quad \operatorname{card}(X) > \operatorname{card}(Y)$$

to mean that there is an injection but no bijection, or a surjection but no bijection, from X to Y. Observe that we attach no meaning to the expression " $\operatorname{card}(X)$" when it stands alone; there are various ways of doing so, but they are irrelevant for our purposes (except when X is finite— see below). These relationships can be extended to the empty set by declaring that

$$\operatorname{card}(\varnothing) < \operatorname{card}(X) \text{ and } \operatorname{card}(X) > \operatorname{card}(\varnothing) \text{ for all } X \neq \varnothing$$

For the remainder of this section we assume implicitly that all sets in question are nonempty in order to avoid special arguments for $\varnothing$. Our first task is to prove that the relationships defined above enjoy the properties that the notation suggests.

Proposition 0.7: 0.6.

$\operatorname{card}(X) \leqslant \operatorname{card}(Y)$ iff $\operatorname{card}(Y) \geqslant \operatorname{card}(X)$.

Proof. If $f \colon X \to Y$ is injective, pick $x_0 \in X$ and define $g \colon Y \to X$ by $g(y) = f^{-1}(y)$ if $y \in f(X), g(y) = x_0$ otherwise. Then g is surjective. Conversely, if $g \colon Y \to X$ is surjective, the sets $g^{-1}(\{x\})(x \in X)$ are nonempty and disjoint, so any $f \in \prod_{x\in X} g^{-1}(\{x\})$ is an injection from X to Y. □

Proposition 0.8: 0.7.

For any sets X and Y, either $\operatorname{card}(X) \leqslant \operatorname{card}(Y)$ or $\operatorname{card}(Y) \leqslant \operatorname{card}(X)$.

Proof. Consider the set $\mathcal{J}$ of all injections from subsets of X to Y. The members of $\mathcal{J}$ can be regarded as subsets of $X \times Y$, so $\mathcal{J}$ is partially ordered by inclusion. It is easily verified that Zorn's lemma applies, so $\mathcal{J}$ has a maximal element f, with (say) domain A and range B. If $x_0 \in X \setminus A$ and $y_0 \in Y \setminus B$, then f can be extended to an injection from $A \cup \{x_0\}$ to $Y \cup \{y_0\}$ by setting $f(x_0) = y_0$, contradicting maximality. Hence either $A = X$, in which case $\operatorname{card}(X) \leqslant \operatorname{card}(Y)$, or $B = Y$, in which case f^{-1} is an injection from Y to X and $\operatorname{card}(Y) \leqslant \operatorname{card}(X)$. □

Theorem 0.9: 0.8: The Schröder-Bernstein Theorem.

If $\operatorname{card}(X) \leqslant \operatorname{card}(Y)$ and $\operatorname{card}(Y) \leqslant \operatorname{card}(X)$ then $\operatorname{card}(X) = \operatorname{card}(Y)$.

Proof. Let $f\colon X \to Y$ and $g\colon Y \to X$ be injections. Consider a point $x \in X$: If $x \in g(Y)$, we form $g^{-1}(x) \in Y$; if $g^{-1}(x) \in f(X)$, we form $f^{-1}(g^{-1}(x))$; and so forth. Either this process can be continued indefinitely, or it terminates with an element of $X \setminus g(Y)$ (perhaps x itself), or it terminates with an element of $Y \setminus f(X)$. In these three cases we say that x is in X_∞, X_X, or X_Y; thus X is the disjoint union of X_∞, X_X, and X_Y. In the same way, Y is the disjoint union of three sets Y_∞, Y_X, and Y_Y. Clearly f maps X_∞ onto Y_∞ and X_X onto Y_X, whereas g maps Y_Y onto X_Y. Therefore, if we define $h\colon X \to Y$ by $h(x) = f(x)$ if $X \in X_\infty \cup X_X$ and $h(x) = g^{-1}(x)$ if $x \in X_Y$, then h is bijective. □

Proposition 0.10: 0.9: Proposition.

For any set X, $\text{card}(X) < \text{card}(\mathcal{P}(X))$.

Proof. On the one hand, the map $f(x) = \{x\}$ is an injection from X to $\mathcal{P}(X)$. On the other, if $g\colon X \to \mathcal{P}(X)$, let $Y = \{x \in X \mid x \notin g(x)\}$. Then $Y \notin g(X)$, for if $Y = g(x_0)$ for some $x_0 \in X$, any attempt to answer the question "Is $x_0 \in Y$?" quickly leads to an absurdity. Hence g cannot be surjective. □

A set X is called countable (or denumerable) if $\text{card}(X) \leqslant \text{card}(\mathbb{Z}_{\geqslant 1})$. In particular, all finite sets are countable, and for these it is convenient to interpret "card (X)" as the number of elements in X:

$$\text{card}(X) = n \text{ iff } \text{card}(X) = \text{card}(\{1, \ldots, n\})$$

If X is countable but not finite, we say that X is countably infinite.

Proposition 0.11: 0.10.

(a) If X and Y are countable, so is $X \times Y$.
(b) If A is countable and X_α is countable for every $\alpha \in A$, then $\bigcup_{\alpha \in A} X_\alpha$ is countable.
(c) If X is countably infinite, then $\text{card}(X) = \text{card}(\mathbb{Z}_{\geqslant 1})$.

Proof. . To prove (a) it suffices to prove that $\mathbb{Z}^2_{\geqslant 1}$ is countable. But we can define a bijection from $\mathbb{Z}_{\geqslant 1}$ to $\mathbb{Z}^2_{\geqslant 1}$ by listing, for n successively equal to $2, 3, 4, \ldots$, those elements $(j, k) \in \mathbb{Z}^2_{\geqslant 1}$ such that $j + k = n$ in order of increasing j, thus:

$$(1,1), (1,2), (2,1), (1,3), (2,2), (3,1), (1,4), (2,3), (3,2), (4,1), \ldots$$

As for (b), for each $\alpha \in A$ there is a surjective $f_\alpha\colon \mathbb{Z}_{\geqslant 1} \to X_\alpha$, and then the map $f\colon \mathbb{Z}_{\geqslant 1} \times A \to \bigcup_{\alpha \in A} X_\alpha$ defined by $f(n, \alpha) = f_\alpha(n)$ is surjective; the result therefore follows from (a). Finally, for (c) it suffices to assume that X is an infinite subset of $\mathbb{Z}_{\geqslant 1}$. Let $f(1)$ be the smallest element of X, and define $f(n)$ inductively to be the smallest element of $E \setminus \{f(1), \ldots, f(n-1)\}$. Then f is easily seen to be a bijection from $\mathbb{Z}_{\geqslant 1}$ to X. □

Corollary 0.12: 0.11.

$\mathbb{Z}$ and $\mathbb{Z}^2$ are countable.

Proof. $\mathbb{Z}$ is the union of the countable sets $\mathbb{Z}, \{-n : n \in \mathbb{Z}\}$, and $\{0\}$, and one can define a surjection $f\colon \mathbb{Z}^2 \to \mathbb{Z}$ by $f(m,n) = m/n$ if $n \neq 0$ and $f(m,0) = 0$. □

A set X is said to have the cardinality of the continuum if $\text{card}(X) = \text{card}(\mathbb{R})$. We shall use the letter $\mathfrak{c}$ as an abbreviation for $\text{card}(\mathbb{R})$:

$$\text{card}(X) = \mathfrak{c} \text{ iff } \text{card}(X) = \text{card}(\mathbb{R})$$

Proposition 0.13: 0.12.

$\text{card}(\mathcal{P}(\mathbb{Z}_{\geqslant 1})) = \mathfrak{c}$.

Proof. If $A \subset \mathbb{Z}_{\geqslant 1}$, define $f(A) \in \mathbb{Z}_{\geqslant 1}$ to be $\sum_{n\in A} 2^{-n}$ if $\mathbb{Z}_{\geqslant 1} \setminus A$ is infinite and $1+\sum_{n\in A} 2^{-n}$ if $\mathbb{Z}_{\geqslant 1} \setminus A$ is finite. (In the two cases, $f(A)$ is the number whose base-2 decimal expansion is $0.a_1a_2\cdots$ or $1.a_1a_2\cdots$, where $a_n = 1$ if $n \in A$ and $a_n = 0$ otherwise.) Then $f\colon \mathcal{P}(\mathbb{Z}_{\geqslant 1}) \to \mathbb{Z}_{\geqslant 1}$ is injective. On the other hand, define $g\colon \mathcal{P}(\mathbb{Z}_{\geqslant 1}) \to \mathbb{Z}_{\geqslant 1}$ by $g(A) = \log(\sum_{n\in A} 2^{-n})$ if A is bounded below and $g(A) = 0$ otherwise. Then g is surjective since every positive real number has a base-2 decimal expansion. Since $\text{card}(\mathcal{P}(\mathbb{Z}_{\geqslant 1})) = \text{card}(\mathcal{P}(\mathbb{Z}_{\geqslant 1}))$, the result follows from the Schröder-Bernstein theorem. □

Corollary 0.14: 0.13.

If $\text{card}(X) \geqslant \mathfrak{c}$, then X is uncountable.

Proof. Apply Proposition 10. □

The converse of this corollary is the so-called continuum hypothesis, whose validity is one of the famous undecidable problems of set theory; see Folland Section 0.7.

Proposition 0.15: 0.14.

(a) If $\text{card}(X) \leqslant \mathfrak{c}$ and $\text{card}(Y) \leqslant \mathfrak{c}$, then $\text{card}(X \times Y) \leqslant \mathfrak{c}$.
(b) If $\text{card}(A) \leqslant \mathfrak{c}$ and $\text{card}(X_\alpha) \leqslant \mathfrak{c}$ for all $\alpha \in A$, then $\text{card}(\bigcup_{\alpha\in A} X_\alpha) \leqslant \mathfrak{c}$.

Proof. For (a) it suffices to take $X = Y = \mathcal{P}(\mathbb{Z}_{\geqslant 1})$. Define $\phi, \psi\colon \mathbb{Z}_{\geqslant 1} \to \mathbb{Z}_{\geqslant 1}$ by $\phi(n) = 2n$ and $\psi(n) = 2n - 1$. It is then easy to check that the map $f\colon \mathcal{P}(\mathbb{Z}_{\geqslant 1})^2 \to \mathcal{P}(\mathbb{Z}_{\geqslant 1})$ defined by $f(A,B) = \phi(A) \cup \psi(B)$ is bijective. (b) follows from (a) as in the proof of Proposition 11. □

0.4 More About Well-Ordered Sets

The material in this section is optional; it is used only in a few exercises and in some notes at the ends of chapters.

Let X be a well ordered set. If $A \subset X$ is nonempty, A has a minimal element, which is its maximal lower bound or infimum; we shall denote it by $\inf A$. If A is bounded above, it also has a minimal upper bound or supremum, denoted by $\sup A$. If $x \in X$, 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.

The principle of mathematical induction is equivalent to the fact that $\mathbb{Z}_{\geqslant 1}$ is well ordered. It can be extended to arbitrary well ordered sets as follows:

Theorem 0.16: 0.15: The Principle of Transfinite Induction.

Let X be a well ordered set. If A is a subset of X such that $x \in A$ whenever $I_x \subset A$, then $A = X$.

Proof. If $X \neq A$, let $x = \inf(X \setminus A)$. Then $I_x \subset A$ but $x \notin A$. □

Proposition 0.17: 0.16.

If X is well ordered and $A \subset X$, then $\bigcup_{x\in A} I_x$ is either an initial segment or X itself.

Proof. Let $J = \bigcup_{x\in A} I_x$. If $J \neq X$, let $b = \inf(X \setminus J)$. If there existed $y \in J$ with $y > b$, we would have $y \in I_x$ for some $x \in A$ and hence $b \in I_x$, contrary to construction. Hence $J \subset I_b$, and it is obvious that $I_b \subset J$. □

Proposition 0.18: 0.17.

If X and Y are well ordered, then either X is order isomorphic to Y, or X is order isomorphic to an initial segment in Y, or Y is order isomorphic to an initial segment in X.

Proof. Consider the set $\mathcal{F}$ of order isomorphisms whose domains are initial segments in X or X itself and whose ranges are initial segments in Y or Y itself. $\mathcal{F}$ is nonempty since the unique $f\colon \{\inf X\} \to \{\inf Y\}$ belongs to $\mathcal{F}$, and $\mathcal{F}$ is partially ordered by inclusion (its members being regarded as subsets of $X \times Y$).

An application of Zorn's lemma shows that $\mathcal{F}$ has a maximal element f, with (say) domain A and range B. If $A = I_x$ and $B = I_y$, then $A \cup \{x\}$ and $B \cup \{y\}$ are again initial segments of X and Y, and f could be extended by setting $f(x) = y$, contradicting maximality. Hence either $A = X$ or $B = Y$ (or both), and the result follows. □

Proposition 0.19: 0.18.

There is an uncountable well ordered set Ω such that I_x is countable for each $x \in \Omega$. If Ω' is another set with the same properties, then Ω and Ω' are order isomorphic.

Proof. Uncountable well ordered sets exist by the well ordering principle; let X be one. Either X has the desired property or there is a minimal element x_0 such that I_{x_0} is uncountable, in which case we can take $\Omega = I_{x_0}$. If Ω' is another such set, Ω' cannot be order isomorphic to an initial segment of Ω or vice versa, because Ω and Ω' are uncountable while their initial segments are countable, so Ω and Ω' are order isomorphic by Proposition 18. □

The set Ω in Proposition 19, which is essentially unique qua well ordered set, is called the set of countable ordinals. It has the following remarkable property:

Proposition 0.20: 0.19.

Every countable subset of Ω has an upper bound.

Proof. If $A \subset \Omega$ is countable, $\bigcup_{x\in A} I_x$ is countable and hence is not all of Ω. By Proposition 17, there exists $y \in \Omega$ such that $\bigcup_{x\in A} I_x = I_y$, and y is thus an upper bound for A.

The set $\mathbb{Z}_{\geqslant 1}$ of positive integers may be identified with a subset of Ω as follows. Set $f(1) = \inf \Omega$, and proceeding inductively, set $f(n) = \inf(\Omega \setminus \{f(1), \ldots, f(n-1)\})$. The reader may verify that f is an order isomorphism from $\mathbb{Z}_{\geqslant 1}$ to I_ω, where ω is the minimal element of Ω such that I_ω is infinite. □

It is sometimes convenient to add an extra element ω_1 to Ω to form a set $\Omega^* = \Omega \cup \{\omega_1\}$ and to extend the ordering on Ω to Ω^* by declaring that $x < \omega_1$ for all $x \in \Omega$. ω_1 is called the first uncountable ordinal. (The usual notation for ω_1 is Ω, since ω_1 is generally taken to be the set of countable ordinals itself.)

0.5 The Extended Real Number System

It is frequently useful to adjoin two extra points $\infty(= +\infty)$ and $-\infty$ to $\mathbb{R}$ to form the extended real number system $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$, and to extend the usual ordering on $\mathbb{R}$ by declaring that $-\infty < x < \infty$ for all $x \in \mathbb{R}$. The completeness of $\mathbb{R}$ can then be stated as follows: Every subset A of $\overline{\mathbb{R}}$ has a least upper bound, or supremum, and a greatest lower bound, or infimum, which are denoted by $\sup A$ and $\inf A$. If $A = \{a_1, \ldots a_n\}$, we also write

$$\max(a_1, \ldots, a_n) = \sup A, \quad \min(a_1, \ldots, a_n) = \inf A$$

From completeness it follows that every sequence $\{x_n\}$ in $\overline{\mathbb{R}}$ has a limit superior and a limit inferior:

$$\limsup x_n = \inf_{k\geqslant 1}(\sup_{n\geqslant k} x_n), \quad \liminf x_n = \sup_{k\geqslant 1}\left(\inf_{n\geqslant k} x_n\right)$$

The sequence $\{x_n\}$ converges (in $\mathbb{R}$) iff these two numbers are equal (and finite), in which case its limit is their common value. One can also define limsup and lim inf for functions $f\colon \mathbb{R} \to \overline{\mathbb{R}}$, for instance:

$$\limsup_{x\to a} f(x) = \inf_{\delta>0}(\sup_{0<|x-a|<\delta} f(x))$$

The arithmetical operations on $\mathbb{R}$ can be partially extended to $\overline{\mathbb{R}}$:

$$x \pm \infty = \pm\infty (x \in \mathbb{R}), \quad \infty + \infty = \infty, \quad -\infty - \infty = -\infty$$

$$x \cdot (\pm\infty) = \pm\infty (x > 0), \quad x \cdot (\pm\infty) = \mp\infty (x < 0)$$

We make no attempt to define $\infty - \infty$, but we abide by the convention that, unless otherwise stated,

$$0 \cdot (\pm\infty) = 0$$

(The expression $0 \cdot \infty$ turns up now and then in measure theory, and for various reasons its proper interpretation is almost always 0.)

We employ the following notation for intervals in $\overline{\mathbb{R}}$: if $-\infty \leqslant a < b \leqslant \infty$,

$$(a, b) = \{x \mid a < x < b\}, \qquad [a, b] = \{x \mid a \leqslant x \leqslant b\}$$

$$(a, b] = \{x \mid a < x \leqslant b\}, \qquad [a, b) = \{x \mid a \leqslant x < b\}$$

We shall occasionally encounter uncountable sums of nonnegative numbers. If X is an arbitrary set and $f\colon X \to [0, \infty]$, we define $\sum_{x\in X} f(x)$ to be the supremum of its finite partial sums:

$$\sum_{x\in X} f(x) = \sup\left\{\sum_{x\in F} f(x) \,\middle|\, F \subset X, F \text{ finite}\right\}$$

(Later we shall recognize this as the integral of f with respect to counting measure on X.)

Proposition 0.21: 0.20.

Given $f\colon X \to [0, \infty]$, let $A = \{x \mid f(x) > 0\}$. If A is uncountable, then $\sum_{x\in X} f(x) = \infty$. If A is countably infinite, then $\sum_{x\in X} f(x) = \sum_1^\infty f(g(n))$ where $g\colon \mathbb{Z}_{\geqslant 1} \to A$ is any bijection and the sum on the right is an ordinary infinite series.

Proof. We have $A = \bigcup_1^\infty A_n$ where $A_n = \{x \mid f(x) > 1/n\}$. If A is uncountable, then some A_n must be uncountable, and $\sum_{x\in F} f(x) > \operatorname{card}(F)/n$ for F a finite subset of A_n; it follows that $\sum_{x\in X} f(x) = \infty$. If A is countably infinite, $g\colon \mathbb{Z}_{\geqslant 1} \to A$ is a bijection, and $B_N = g(\{1, \ldots, N\})$, then every finite subset F of A is contained in some B_N. Hence

$$\sum_{x\in F} f(x) \leqslant \sum_1^N f(g(n)) \leqslant \sum_{x\in X} f(x)$$

Taking the supremum over N, we find

$$\sum_{x\in F} f(x) \leqslant \sum_1^\infty f(g(n)) \leqslant \sum_{x\in X} f(x)$$

and then taking the supremum over F, we obtain the desired result. □

Some terminology concerning (extended) real-valued functions: A relation between numbers that is applied to functions is understood to hold pointwise. Thus $f \leqslant g$ means that $f(x) \leqslant g(x)$ for every x, and $\max(f, g)$ is the function whose value at x is $\max(f(x), g(x))$. If $X \subset \overline{\mathbb{R}}$ and $f\colon X \to \overline{\mathbb{R}}$, f is called increasing if $f(x) \leqslant f(y)$ whenever $x \leqslant y$ and strictly increasing if $f(x) < f(y)$ whenever $x < y$; similarly for decreasing. A function that is either increasing or decreasing is called monotone.

If $f\colon \mathbb{R} \to \mathbb{R}$ is an increasing function, then f has right- and left-hand limits at each point:

$$f(a+) = \lim_{x\searrow a} f(x) = \inf_{x>a} f(x), \quad f(a-) = \lim_{x\nearrow a} f(x) = \sup\nolimits_{x<a} f(x).$$

Moreover, the limiting values $f(\infty) = \sup_{a\in\mathbb{R}} f(x)$ and $f(-\infty) = \inf_{a\in\mathbb{R}} f(x)$ exist (possibly equal to $\pm\infty$). f is called right continuous if $f(a) = f(a+)$ for all $a \in \mathbb{R}$ and left continuous if $f(a) = f(a-)$ for all $a \in \mathbb{R}$.

For points x in $\mathbb{R}$ or $\mathbb{R}$, $|x|$ denotes the ordinary absolute value or modulus of x, $|a + ib| = \sqrt{a^2 + b^2}$. For points x in $\mathbb{R}^n$ or $\mathbb{R}^n$, $|x|$ denotes the Euclidean norm:

$$|x| = \left[\sum_1^n |x_j|^2\right]^{1/2}$$

We recall that a set $U \subset \mathbb{R}$ is open if, for every $x \in U, U$ includes an interval centered at x.

Proposition 0.22: 0.21.

Every open set in $\mathbb{R}$ is a countable disjoint union of open intervals.

Proof. If U is open, for each $x \in U$ consider the collection $\mathcal{J}_x$ of all open intervals I such that $x \in I \subset U$. It is easy to check that the union of any family of open intervals containing a point in common is again an open interval, and hence $J_x = \bigcup_{I\in\mathcal{J}_x} I$ is an open interval; it is the largest element of $\mathcal{J}_x$. If $x, y \in U$ then either $J_x = J_y$ or $J_x \cap J_y = \varnothing$, for otherwise $J_x \cup J_y$ would be a larger open interval than J_x in $\mathcal{J}_x$. Thus if $\mathcal{J} = \{J_x \mid x \in U\}$, the (distinct) members of $\mathcal{J}$ are disjoint, and $U = \bigcup_{J\in\mathcal{J}} J$. For each $J \in \mathcal{J}$, pick a rational number $f(J) \in J$. The map $f\colon \mathcal{J} \to \mathbb{Q}$ thus defined is injective, for if $J \neq J'$ then $J \cap J' = \varnothing$; therefore $\mathcal{J}$ is countable. □

0.6 Metric Spaces

A metric on a set X is a function $\rho\colon X \times X \to [0, \infty)$ such that

- $\rho(x, y) = 0$ iff $x = y$;
- $\rho(x, y) = \rho(y, x)$ for all $x, y \in X$;

- $\rho(x, z) \leqslant \rho(x, y) + \rho(y, z)$ for all $x, y, z \in X$.

(Intuitively, $\rho(x, y)$ is to be interpreted as the distance from x to y.) A set equipped with a metric is called a metric space.

Example 23. *The following are some example of metric spaces.*

(i) The Euclidean distance $\rho(x, y) = |x - y|$ is a metric on $\mathbb{R}^n$.

(ii) $\rho_1(f, g) = \int_0^1 |f(x) - g(x)|dx$ and $\rho_\infty(f, g) = \sup_{0 \leqslant x \leqslant 1} |f(x) - g(x)|$ are metrics on the space of continuous functions on $[0, 1]$.

(iii) If ρ is a metric on X and $A \subset X$, then $\rho|(A \times A)$ is a metric on A.

(iv) If (X_1, ρ_1) and (X_2, ρ_2) are metric spaces, the product metric ρ on $X_1 \times X_2$ is given by

$$\rho((x_1, x_2), (y_1, y_2)) = \max(\rho_1(x_1, y_1), \rho_2(x_2, y_2))$$

Other metrics are sometimes used on $X_1 \times X_2$, for instance,

$$\rho_1(x_1, y_1) + \rho_2(x_2, y_2) \qquad \text{or } \left[\rho_1(x_1, y_1)^2 + \rho_2(x_2, y_2)^2\right]^{1/2}$$

These, however, are equivalent to the product metric in the sense that we shall define at the end of this section.

Let (X, ρ) be a metric space. If $x \in X$ and $r > 0$, the (open) ball of radius r about x is

$$B(r, x) = \{y \in X \mid \rho(x, y) < r\}$$

A set $E \subset X$ is open if for every $x \in E$ there exists $r > 0$ such that $B(r, x) \subset E$, and closed if its complement is open. For example, every ball $B(r, x)$ is open, for if $y \in B(r, x)$ and $\rho(x, y) = s$ then $B(r - s, y) \subset B(r, x)$. Also, X and $\varnothing$ are both open and closed. Clearly the union of any family of open sets is open, and hence the intersection of any family of closed sets is closed. Also, the intersection (resp. union) of any finite family of open (resp. closed) sets is open (resp. closed). Indeed, if $U_1, \ldots U_n$ are open and $x \in \bigcap_1^n U_j$, for each j there exists $r_j > 0$ such that $B(r_j, x) \subset U_j$, and then $B(r, x) \subset \bigcap_1^n U_j$ where $r = \min(r_1, \ldots, r_n)$, so $\bigcap_1^n U_j$ is open.

If $E \subset X$, the union of all open sets $U \subset E$ is the largest open set contained in E; it is called the interior of E and is denoted by E^o. Likewise, the intersection of all closed sets $F \supset E$ is the smallest closed set containing E; it is called the closure of E and is denoted by $\overline{E}$. E is said to be dense in X if $\overline{E} = X$, and nowhere dense if $\overline{E}$ has empty interior. X is called separable if it has a countable dense subset. (For example, $\mathbb{Q}^n$ is a countable dense subset of $\mathbb{Q}^n$.) A sequence $\{x_n\}$ in X converges to $x \in X$ (symbolically: $x_n \to x$ or $\lim x_n = x$) if $\lim_{n \to \infty} \rho(x_n, x) = 0$.

Proposition 0.24: 0.22.

If X is a metric space, $E \subset X$, and $x \in X$, the following are equivalent:

(a) $x \in \overline{E}$.

(b) $B(r, x) \cap E \neq \varnothing$ for all $r > 0$.

(c) There is a sequence $\{x_n\}$ in E that converges to x.

Proof. If $B(r,x) \cap E = \varnothing$, then $B(r,x)^c$ is a closed set containing E but not x, so $x \notin \overline{E}$. Conversely, if $x \notin \overline{E}$, since $(\overline{E})^c$ is open there exists $r > 0$ such that $B(r,x) \subset (\overline{E})^c \subset E^c$. Thus (a) is equivalent to (b). If (b) holds, for each $n \in \mathbb{Z}_{\geqslant 1}$ there exists $x_n \in B(n^{-1},x) \cap E$, so that $x_n \to x$. On the other hand, if $B(r,x) \cap E = \varnothing$, then $\rho(y,x) \geqslant r$ for all $y \in E$, so no sequence of E can converge to x. Thus (b) is equivalent to (c). □

If (X_1,ρ_1) and (X_2,ρ_2) are metric spaces, a map $f\colon X_1 \to X_2$ is called continuous at $x \in X$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\rho_2(f(y),f(x)) < \varepsilon$ whenver $\rho_1(x,y) < \delta$—in other words, such that $f^{-1}(B(\varepsilon,f(x))) \supset B(\delta,x)$. The map f is called continuous if it is continuous at each $x \in X_1$ and uniformly continuous if, in addition, the δ in the definition of continuity can be chosen independent of x.

Proposition 0.25: 0.23.

$f\colon X_1 \to X_2$ is continuous iff $f^{-1}(U)$ is open in X_1 for every open $U \subset X_2$.

Proof. If the latter condition holds, then for every $x \in X_1$ and $\varepsilon > 0$, the set $f^{-1}(B(\varepsilon,f(x)))$ is open and contains x, so it contains some ball about x; this means that f is continuous at x. Conversely, suppose that f is continuous and U is open in X_2. For each $y \in U$ there exists $\varepsilon_y > 0$ such that $B(\varepsilon_y,y) \subset U$, and for each $x \in f^{-1}(\{y\})$ there exists $\delta_x > 0$ such that $B(\delta_x,x) \subset f^{-1}(B(\varepsilon_y,y)) \subset f^{-1}(U)$. Thus $f^{-1}(U) = \bigcup_{x \in f^{-1}(U)} B(\delta_x,x)$ is open. □

A sequence $\{x_n\}$ in a metric space (X,ρ) is called Cauchy if $\rho(x_n,x_m) \to 0$ as $n,m \to \infty$. A subset E of X is called complete if every Cauchy sequence in E converges and its limit is in E. For example, $\mathbb{R}^n$ (with the Euclidean metric) is complete, whereas $\mathbb{R}^n$ is not.

Proposition 0.26: 0.24.

A closed subset of a complete metric space is complete, and a complete subset of an arbitrary metric space is closed.

Proof. If X is complete, $E \subset X$ is closed, and $\{x_n\}$ is a Cauchy sequence in E, $\{x_n\}$ has a limit in X. By Proposition 0.22, $x \in \overline{E} = E$. If $E \subset X$ is complete and $x \in \overline{E}$, by Proposition (0.22) there is a sequence $\{x_n\}$ in E converging to $x.\{x_n\}$ is Cauchy, so its limit lies in E; thus $E = \overline{E}$. □

In a metric space (X,ρ) we can define the distance from a point to a set and the distance between two sets. Namely, if $x \in X$ and $E,F \subset X$,

$$\rho(x,E) = \inf\{\rho(x,y) \mid y \in E\}$$

$$\rho(E,F) = \inf\{\rho(x,y) \mid x \in E, y \in F\} = \inf\{\rho(x,F) \mid x \in E\}$$

Observe that, by Proposition 24, $\rho(x,E) = 0$ iff $x \in \overline{E}$. We also define the diameter of $E \subset X$ to be

$$\operatorname{diam} E = \sup\{\rho(x,y) \mid x, y \in E\}$$

E is called bounded if $\operatorname{diam} E < \infty$.

If $E \subset X$ and $\{V_\alpha\}_{\alpha \in A}$ is a family of sets such that $E \subset \bigcup_{\alpha \in A} V_\alpha$, $\{V_\alpha\}_{\alpha \in A}$ is called a cover of E, and E is said to be covered by the V_αs. E is called totally bounded if, for every $\varepsilon > 0$, E can be covered by finitely many balls of radius ε. Every totally bounded set is bounded, for if $x, y \in \bigcup_1^n B(\varepsilon, z_j)$, say $x \in B(\varepsilon, z_1)$ and $y \in B(\varepsilon, z_2)$, then

$$\rho(x,y) \leqslant \rho(x,z_1) + \rho(z_1,z_2) + \rho(z_2,y) \leqslant 2\varepsilon + \max\{\rho(z_j,z_k) \mid 1 \leqslant j,k \leqslant n\}$$

(The converse is false in general.) If E is totally bounded, so is $\overline{E}$, for it is easily seen that if $E \subset \bigcup_1^n B(\varepsilon, z_j)$, then $\overline{E} \subset \bigcup_1^n B(2\varepsilon, z_j)$.

Theorem 0.27: 0.25.

If E is a subset of the metric space (X,ρ), the following are equivalent:

(a) E is complete and totally bounded.

(b) (The Bolzano-Weierstrass Property) Every sequence in E has a subsequence that converges to a point of E.

(c) (The Heine-Borel Property) If $\{V_\alpha\}_{\alpha \in A}$ is a cover of E by open sets, there is a finite set $F \subset A$ such that $\{V_\alpha\}_{\alpha \in F}$ covers E.

Proof. We shall show that (a) and (b) are equivalent, that (a) and (b) together imply (c), and finally that (c) implies (b).

(a) implies (b): Suppose that (a) holds and $\{x_n\}$ is a sequence in E. E can be covered by finitely many balls of radius 2^{-1}, and at least one of them must contain x_n for infinitely many n: say, $x_n \in B_1$ for $n \in N_1$.$E \cap B_1$ can be covered by finitely many balls of radius 2^{-2}, and at least one of them must contain x_n for infinitely many $n \in N_1$: say, $x_n \in B_2$ for $n \in N_2$. Continuing inductively, we obtain a sequence of balls B_j of radius 2^{-j} and a decreasing sequence of subsets N_j of $\mathbb{Z}_{\geqslant 1}$ such that $x_n \in B_j$ for $n \in N_j$. Pick $n_1 \in N_1, n_2 \in N_2, \ldots$ such that $n_1 < n_2 < \cdots$. Then $\{x_{n_j}\}$ is a Cauchy sequence, for $\rho(x_{n_j}, x_{n_k}) < 2^{1-j}$ if $k > j$, and since E is complete, it has a limit in E.

(b) implies (a): We show that if either condition in (a) fails, then so does (b). If E is not complete, there is a Cauchy sequence $\{x_n\}$ in E with no limit in E. No subsequence of $\{x_n\}$ can converge in E, for otherwise the whole squence would converge to the same limit. On the other hand, if E is not totally bounded, let $\varepsilon > 0$ be such that E cannot be covered by finitely many balls of radius ε. Choose $x_n \in E$ inductively as follows. Begin with any $x_1 \in E$, and having chosen $x_1, \ldots, x_n$, pick $x_{n+1} \in E \setminus \bigcup_1^n B(\varepsilon, x_j)$. Then $\rho(x_n, x_m) > \varepsilon$ for all n, m, so $\{x_n\}$ has no convergent subsequence.

(a) and (b) imply (c): It suffices to show that if (b) holds and $\{V_\alpha\}_{\alpha\in A}$ is a cover of E by open sets, there exists $\varepsilon > 0$ such that every ball of radius ε that intersects E is contained in some V_α, for E can be covered by finitely many such balls by (a). Suppose to the contrary that for each $n \in \mathbb{Z}_{\geqslant 1}$ there is a ball B_n of radius 2^{-n} such that $B_n \cap E \neq \varnothing$ and B_n is contained in no V_α. Pick $x_n \in B_n \cap E$; by passing to a subsequence we may assume that $\{x_n\}$ converges to some $x \in E$. We have $x \in V_\alpha$ for some α, and since V_α is open, there exists $\varepsilon > 0$ such that $B(\varepsilon, x) \subset V_\alpha$. But if n is large enough so that $\rho(x_n, x) < \varepsilon/3$ and $2^{-n} < \varepsilon/3$, then $B_n \subset B(\varepsilon, x) \subset V_\alpha$, contradicting the assumption on B_n.

(c) implies (b): If $\{x_n\}$ is a sequence in E with no convergent subsequence, for each $x \in E$ there is a ball B_x centered at x that contains x_n for only finitely many n (otherwise some subsequence would converge to x). Then $\{B_x\}_{x\in E}$ is a cover of E by open sets with no finite subcover. □

A set E that possesses the properties (a)-(c) of Theorem 27 is called compact. Every compact set is closed (by Proposition 26) and bounded; the converse is false in general but true in $\mathbb{R}^n$.

Proposition 0.28: 0.26.

Every closed and bounded subset of $\mathbb{R}^n$ is compact.

Proof. Since closed subsets of $\mathbb{R}^n$ are complete, it suffices to show that bounded subsets of $\mathbb{R}^n$ are totally bounded. Since every bounded set is contained in some cube

$$Q = [-R, R]^n = \{x \in \mathbb{R}^n \mid \max(|x_1|, \ldots, |x_n|) \leqslant R\}$$

it is enough to show that Q is totally bounded. Given $\varepsilon > 0$, pick an integer $k > R\sqrt{n}/\varepsilon$, and express Q as the union of k^n congruent subcubes by dividing the interval $[-R, R]$ into k equal pieces. The side length of these subcubes is $2R/k$ and hence their diameter is $\sqrt{n}(2R/k) < 2\varepsilon$, so they are contained in the balls of radius ε about their centers. □

Two metrics ρ_1 and ρ_2 on a set X are called equivalent if

$$C\rho_1 \leqslant \rho_2 \leqslant C'\rho_1 \text{ for some } C, C' > 0$$

It is easily verified that equivalent metrics define the same open, closed, and compact sets, the same convergent and Cauchy sequences, and the same continuous and uniformly continuous mappings. Consequently, most results concerning metric spaces depend not on the particular metric chosen but only on its equivalence class.

1 Measures

In this chapter we set forth the basic concepts of measure theory, develop a general procedure for constructing nontrivial examples of measures, and apply this procedure to construct measures on the real line.

1.1 Introduction

One of the most venerable problems in geometry is to determine the area or volume of a region in the plane or in 3-space. The techniques of integral calculus provide a satisfactory solution to this problem for regions that are bounded by "nice" curves or surfaces but are inadequate to handle more complicated sets, even in dimension one. Ideally, for $n \in \mathbb{Z}_{\geqslant 1}$ we would like to have a function μ that assigns to each $E \subset \mathbb{Z}^n_{\geqslant 1}$ a number $\mu(E) \in [0, \infty]$, the n-dimensional measure of E, such that $\mu(E)$ is given by the usual integral formulas when the latter apply. Such a function μ should surely possess the following properties:

(i) If $E_1, E_2, \ldots$ is a finite or infinite sequence of disjoint sets, then

$$\mu(E_1 \cup E_2 \cup \cdots) = \mu(E_1) + \mu(E_2) + \cdots$$

(ii) If E is congruent to F (that is, if E can be transformed into F by translations, rotations, and reflections), then $\mu(E) = \mu(F)$.

(iii) $\mu(Q) = 1$, where Q is the unit cube

$$Q = \{x \in \mathbb{R}^n \mid 0 \leqslant x_j < 1 \text{ for } j = 1, \ldots, n\}$$

Unfortunately, these conditions are mutually inconsistent. Let us see why this is true for $n = 1$. (The argument can easily be adapted to higher dimensions.) To begin with, we define an equivalence relation on $[0, 1)$ by declaring that $x \sim y$ if and only if $x - y$ is rational. Let N be a subset of $[0, 1)$ that contains precisely one member of each equivalence class. (To find such an N, one must invoke the axiom of choice.) Next, let $R = \mathbb{Q} \cap [0, 1)$, and for each $r \in R$ let

$$N_r = \{x + r \mid x \in N \cap [0, 1 - r)\} \cup \{x + r - 1 \mid x \in N \cap [1 - r, 1)\}$$

That is, to obtain N_r, shift N to the right by r units and then shift the part that sticks out beyond $[0, 1)$ one unit to the left. Then $N_r \subset [0, 1)$, and every $x \in [0, 1)$ belongs to precisely one N_r. Indeed, if y is the element of N that belongs to the equivalence class of x, then $x \in N_r$ where $r = x - y$ if $x \geqslant y$ or $r = x - y + 1$ if $x < y$; on the other hand, if $x \in N_r \cap N_s$, then $x - r$ (or $x - r + 1$) and $x - s$ (or $x - s + 1$) would be distinct elements of N belonging to the same equivalence class, which is impossible.

Suppose now that $\mu\colon \mathcal{P}(\mathbb{R}) \to [0, \infty]$ satisfies (i), (ii), and (iii). By (i) and (ii),

$$\mu(N) = \mu(N \cap [0, 1 - r)) + \mu(N \cap [1 - r, 1)) = \mu(N_r)$$

for any $r \in R$. Also, since R is countable and $[0, 1)$ is the disjoint union of the N_rs,

$$\mu([0, 1)) = \sum_{r \in R} \mu(N_r)$$

by (i) again. But $\mu([0, 1)) = 1$ by (iii), and since $\mu(N_r) = \mu(N)$, the sum on the right is either 0 (if $\mu(N) = 0$) or ∞ (if $\mu(N) > 0$). Hence no such μ can exist.

Faced with this discouraging situation, one might consider weakening (i) so that additivity is required to hold only for finite sequences. This is not a very good idea, as we shall see: The additivity for countable sequences is what makes all the limit and continuity results of the theory work smoothly. Moreover, in dimensions $n \in \mathbb{Z}_{\geqslant 3}$, even this weak form of (i) is inconsistent with (ii) and (iii). Indeed, in 1924 Banach and Tarski proved the following amazing result:

Let U and V be arbitrary bounded open sets in $\mathbb{R}^n, n \in \mathbb{Z}_{\geqslant 3}$. There exist $k \in \mathbb{R}$ and subsets $E_1, \ldots, E_k, F_1, \ldots, F_k$ of $\mathbb{R}^n$ such that

- the E_js are disjoint and their union is U;
- the F_js are disjoint and their union is V;
- E_j is congruent to F_j for $j = 1, \ldots, k$.

Thus one can cut up a ball the size of a pea into a finite number of pieces and rearrange them to form a ball the size of the earth! Needless to say, the sets E_j and F_j are very bizarre. They cannot be visualized accurately, and their construction depends on the axiom of choice. But their existence clearly precludes the construction of any $\mu\colon \mathcal{P}(\mathbb{R}^n) \to [0, \infty]$ that assigns positive, finite values to bounded open sets and satisfies (i) for finite sequences as well as (ii).

The moral of these examples is that $\mathbb{R}^n$ contains subsets which are so strangely put together that it is impossible to define a geometrically reasonable notion of measure for them, and the remedy for the situation is to discard the requirement that μ should be defined on all subsets of $\mathbb{R}^n$. Rather, we shall content ourselves with constructing μ on a class of subsets of $\mathbb{R}^n$ that includes all the sets one is likely to meet in practice unless one is deliberately searching for pathological examples. This construction will be carried out for $n = 1$ in Folland Section 1.5 and for $n > 1$ in Folland Section 2.6.

It is worthwhile, and not much extra work, to develop the theory in much greater generality. The conditions (ii) and (iii) are directly related to Euclidean geometry, but set functions satisfying (i), called measures, arise also in a great many other situations. For example, in a physics problem involving mass distributions, $\mu(E)$ could represent the total mass in the region E. For another example, in probability theory one considers a set X that represents the possible outcomes of an experiment, and for $E \subset X, \mu(E)$ is the probability that the outcome lies in E. We therefore begin by studying the theory of measures on abstract sets.

1.2 Sigma Algebras

In this section we discuss the families of sets that serve as the domains of measures.

Let X be a nonempty set. An algebra of sets on X is a nonempty collection $\mathcal{A}$ of subsets of X that is closed under finite unions and complements; in other words, if $E_1, \ldots, E_n \in \mathcal{A}$, then $\bigcup_1^n E_j \in \mathcal{A}$; and if $E \in \mathcal{A}$, then $E^c \in \mathcal{A}$. A σ-algebra is an algebra that is closed under countable unions. (Some authors use the terms field and $\boldsymbol{\sigma}$-field instead of algebra and σ-algebra.)

We observe that since $\bigcap_j E_j = (\bigcup_j E_j^c)^c$, algebras (resp. σ-algebras) are also closed under finite (resp. countable) intersections. Moreover, if $\mathcal{A}$ is an algebra, then $\varnothing \in \mathcal{A}$ and $X \in \mathcal{A}$, for if $E \in \mathcal{A}$ we have $\varnothing = E \cap E^c$ and $X = E \cup E^c$.

It is worth noting that an algebra $\mathcal{A}$ is a σ-algebra provided that it is closed under countable disjoint unions. Indeed, suppose $\{E_j\}_1^\infty \subset \mathcal{A}$. Set

$$F_k = E_k \setminus \left[\bigcup_1^{k-1} E_j\right] = E_k \cap \left[\bigcup_1^{k-1} E_j\right]^c$$

Then the F_ks belong to $\mathcal{A}$ and are disjoint, and $\bigcup_1^\infty E_j = \bigcup_1^\infty F_k$. This device of replacing a sequence of sets by a disjoint sequence is worth remembering; it will be used a number of times below.

Some examples: If X is any set, $\mathcal{P}(X)$ and $\{\varnothing, X\}$ are σ-algebras. If X is uncountable, then

$$\mathcal{A} = \{E \subset X \mid E \text{ is countable or } E^c \text{ is countable }\}$$

is a σ-algebra, called the σ-algebra of countable or co-countable sets. (The point here is that if $\{E_j\}_1^\infty \subset \mathcal{A}$, then $\bigcup_1^\infty E_j$ is countable if all E_j are countable and is co-countable otherwise.)

It is trivial to verify that the intersection of any family of σ-algebras on X is again a σ-algebra. It follows that if $\mathcal{E}$ is any susbset of $\mathcal{E}(X)$, there is a unique smallest σ-algebra $\mathcal{E}(\mathcal{E})$ containing $\mathcal{E}$, namely, the intersection of all σ-algebras containing $\mathcal{E}$. (There is always at least one such, namely, $\mathcal{E}(X)$.) $\mathcal{E}(\mathcal{E})$ is called the σ-algebra generated by $\mathcal{E}$. The following observation is often useful:

Lemma 1.1: 1.1.

If $\mathcal{E} \subset \mathcal{M}(\mathcal{F})$ then $\mathcal{M}(\mathcal{E}) \subset \mathcal{M}(\mathcal{F})$.

The proof is an exercise.

If X is any metric space, or more generally any topological space (see Folland Chapter 4), the σ-algebra generated by the family of open sets in X (or, equivalently, by the family of closed sets in X) is called the Borel σ-algebra on X and is denoted by $\mathcal{B}_X$. Its members are called Borel sets. $\mathcal{B}_X$ thus includes open sets, closed sets, countable intersections of open sets, countable unions of closed sets, and so forth.

There is a standard terminology for the levels in this hierarchy. A countable intersection of open sets is called a $\boldsymbol{G_\delta}$ set; a countable union of closed sets is called an $\boldsymbol{F_\sigma}$ set; a countable union of G_δ sets is called a $\boldsymbol{G_{\delta\sigma}}$ set; a countable intersection of F_σ sets is called an $\boldsymbol{F_{\sigma\delta}}$ set; and so forth. (δ and σ stand for the German Durchschnitt and Summe, that is, intersection and union.)

The Borel σ-algebra on $\mathbb{R}$ will play a fundamental role in what follows. For future reference we note that it can be generated in a number of different ways:

Proposition 1.2: 1.2.

$\mathcal{B}_{\mathbb{R}}$ is generated by each of the following:

(a) the open intervals: $\varepsilon_1 = \{(a,b) \mid a < b\}$,
(b) the closed intervals: $\mathcal{E}_2 = \{[a,b] \mid a < b\}$,
(c) the half-open intervals: $\mathcal{E}_3 = \{(a,b] \mid a < b\}$ or $\mathcal{E}_4 = \{[a,b) \mid a < b\}$,
(d) the open rays: $\mathcal{E}_5 = \{(a,\infty) \mid a \in \mathbb{R}\}$ or $\mathcal{E}_6 = \{(-\infty,a) \mid a \in \mathbb{R}\}$,
(e) the closed rays: $\mathcal{E}_7 = \{[a,\infty) \mid a \in \mathbb{R}\}$ or $\mathcal{E}_8 = \{(-\infty,a] \mid a \in \mathbb{R}\}$.

Proof. The elements of $\mathcal{E}_j$ for $j \neq 3,4$ are open or closed, and the elements of ε_3 and $\mathcal{E}_4$ are G_δ sets—for example, $(a,b] = \bigcap_1^\infty (a, b+n^{-1})$. All of these are Borel sets, so by Lemma 1, $\mathcal{E}(\mathcal{E}_j) \subset \mathcal{E}_{\mathbb{R}}$ for all j. On the other hand, every open set in $\mathbb{R}$ is a countable union of open intervals, so by Lemma 1 again, $\mathcal{E}_{\mathbb{R}} \subset \mathcal{E}(\mathcal{E}_1)$. That $\mathcal{E}_{\mathbb{R}} \subset \mathcal{E}(\varepsilon_j)$ for $j \geqslant 2$ can now be established by showing that all open intervals lie in $\mathcal{E}(\mathcal{E}_j)$ and applying Lemma 1. For example, $(a,b) = \bigcup_1^\infty [a+n^{-1}, b-n^{-1}] \in \mathcal{E}(\mathcal{E}_2)$. Verification of the other cases is left to the reader (Folland Exercise 1.2). □

Let $\{X_\alpha\}_{\alpha \in A}$ be an indexed collection of nonempty sets, $X = \prod_{\alpha \in A} X_\alpha$, and $\pi_\alpha \colon X \to X_\alpha$ the coordinate maps. If $\mathcal{M}_\alpha$ is a σ-algebra on X_α for each α, the product σ-algebra on X is the σ-algebra generated by

$$\{\pi_\alpha^{-1}(E_\alpha) \mid E_\alpha \in \mathcal{M}_\alpha, \alpha \in A\}$$

We denote this σ-algebra by $\bigotimes_{\alpha \in A} \mathcal{M}_\alpha$. (If $A = \{1, \ldots, n\}$ we also write $\bigotimes_1^n \mathcal{M}_j$ or $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$.) The significance of this definition will become clearer in Folland Section 2.1; for the moment we give an alternative, and perhaps more intuitive, characterization of product σ-algebras in the case of countably many factors.

Proposition 1.3: 1.3.

If A is countable, then $\bigotimes_{\alpha \in A} \mathcal{M}_\alpha$ is the σ-algebra generated by $\{\prod_{\alpha \in A} E_\alpha \mid E_\alpha \in \mathcal{M}_\alpha\}$.

Proof. If $E_\alpha \in \mathcal{M}_\alpha$, then $\pi_\alpha^{-1}(E_\alpha) = \prod_{\beta \in A} E_\beta$ where $E_\beta = X$ for $\beta \neq \alpha$; on the other hand, $\prod_{\alpha \in A} E_\alpha = \bigcap_{\alpha \in A} \pi_\alpha^{-1}(E_\alpha)$. The result therefore follows from Lemma 1. □

Proposition 1.4: 1.4.

Suppose that $\mathcal{M}_\alpha$ is generated by $\varepsilon_\alpha, \alpha \in A$. Then $\bigotimes_{\alpha \in A} \mathcal{M}_\alpha$ is generated by $\mathcal{M}_1 = \{\pi_\alpha^{-1}(E_\alpha) \mid E_\alpha \in \mathcal{M}_\alpha, \alpha \in A\}$. If A is countable and $X_\alpha \in \mathcal{M}_\alpha$ for all α, $\bigotimes_{\alpha \in A} \mathcal{M}_\alpha$ is generated by $\mathcal{M}_2 = \{\prod_{\alpha \in A} E_\alpha \mid E_\alpha \in \mathcal{M}_\alpha\}$.

Proof. Obviously $\mathcal{M}(\mathcal{M}_1) \subset \bigotimes_{\alpha \in A} \mathcal{M}_\alpha$. On the other hand, for each α, the collection $\{E \subset X_\alpha \mid \pi_\alpha^{-1}(E) \in \mathcal{M}(\mathcal{M}_1)\}$ is easily seen to be a σ-algebra on X_α that contains $\mathcal{M}_\alpha$

and hence $\mathcal{M}_\alpha$. In other words, $\pi_\alpha^{-1}(E) \in \mathcal{M}(\mathcal{M}_1)$ for all $E \in \mathcal{M}_\alpha$, $\alpha \in A$, and hence $\bigotimes_{\alpha\in A} \mathcal{M}_\alpha \subset \mathcal{M}(\mathcal{M}_1)$. The second assertion follows from the first as in the proof of Proposition 3. □

Proposition 1.5: 1.5.

Let $X_1, \ldots, X_n$ be metric spaces and let $X = \prod_1^n X_j$, equipped with the product metric. Then $\bigotimes_1^n \mathcal{B}_{X_j} \subset \mathcal{B}_X$. If the X_js are separable, then $\bigotimes_1^n \mathcal{B}_{X_j} = \mathcal{B}_X$.

Proof. By Proposition 4, $\otimes_1^n \mathcal{B}_{X_j}$ is generated by the sets $\pi_j^{-1}(U_j), 1 \leqslant j \leqslant n$, where U_j is open in X_j. Since these sets are open in X, Lemma 1 implies that $\otimes_1^n \mathcal{B}_{X_j} \subset \mathcal{B}_X$. Suppose now that C_j is a countable dense set in X_j, and let $\mathcal{B}_j$ be the collection of balls in X_j with rational radius and center in C_j. Then every open set in X_j is a union of members of $\mathcal{B}_j$—in fact, a countable union since $\mathcal{B}_j$ itself is countable. Moreover, the set of points in X whose jth coordinate is in C_j for all j is a countable dense subset of X, and the balls of radius r in X are merely products of balls of radius r in the X_js. It follows that $\mathcal{B}_{X_j}$ is generated by $\mathcal{B}_j$ and $\mathcal{B}_X$ is generated by $\{\prod_1^n E_j \mid E_j \in \mathcal{B}_j\}$. Therefore $\mathcal{B}_X = \bigotimes_1^n \mathcal{B}_{X_j}$ by Proposition 4. □

Corollary 1.6: 1.6.

$\mathcal{B}_{\mathbb{R}^n} = \bigotimes_1^n \mathcal{B}_{\mathbb{R}}$.

We conclude this section with a technical result that will be needed later. We define an elementary family to be a collection $\mathcal{E}$ of subsets of X such that

- $\varnothing \in \mathcal{E}$,
- if $E, F \in \mathcal{E}$ then $E \cap F \in \mathcal{E}$,
- if $E \in \mathcal{E}$ then E^c is a finite disjoint union of members of $\mathcal{E}$.

Proposition 1.7: 1.7.

If $\mathcal{E}$ is an elementary family, the collection $\mathcal{E}$ of finite disjoint unions of members of $\mathcal{E}$ is an algebra.

Proof. . If $A, B \in \mathcal{E}$ and $B^c = \bigcup_1^J C_j (C_j \in \mathcal{E}$, disjoint), then $A \setminus B = \cup_1^J (A \cap C_j)$ and $A \cup B = (A \setminus B) \cup B$, where these unions are disjoint, so $A \setminus B \in \mathcal{E}$ and $A \cup B \in \mathcal{E}$. It now follows by induction that if $A_1 \ldots, A_n \in \mathcal{E}$, then $\bigcup_1^n A_j \in \mathcal{E}$; indeed, by inductive hypothesis we may assume that $A_1, \ldots, A_{n-1}$ are disjoint, and then $\bigcup_1^n A_j = A_n \cup \bigcup_1^{n-1} (A_j \setminus A_n)$, which is a disjoint union. To see that $\mathcal{E}$ is closed under complements, suppose $A_1, \ldots A_n \in \mathcal{E}$ and $A_m^c = \bigcup_{j=1}^{J_m} B_m^j$ with $B_m^1, \ldots, B_m^{J_m}$ disjoint members of $\mathcal{E}$. Then

$$\left(\bigcup_{m=1}^n A_m\right)^c = \bigcap_{m=1}^n \left(\bigcup_{j=1}^{J_m} B_m^j\right) = \bigcup \{B_1^{j_1} \cap \cdots \cap B_n^{j_n} \mid 1 \leqslant j_m \leqslant J_m, 1 \leqslant m \leqslant n\},$$

which is in $\mathcal{A}$. □

Exercise 1.8: Folland Exercise 1.1.

A family of sets $\mathcal{R} \subset \mathcal{R}(X)$ is called a ring if it is closed under finite unions and differences (i.e., if $E_1, \dots, E_n \in \mathcal{R}$, then $\bigcup_1^n E_j \in \mathcal{R}$, and if $E, F \in \mathcal{R}$, then $E \setminus F \in \mathcal{R}$). A ring that is closed under countable unions is called a σ-ring.

(a) Rings (resp. σ-rings) are closed under finite (resp. countable) intersections.
(b) If $\mathcal{R}$ is a ring (resp. σ-ring), then $\mathcal{R}$ is an algebra (resp. σ-algebra) if and only if $X \in \mathcal{R}$.
(c) If $\mathcal{R}$ is a σ-ring, then $\{E \subset X \mid E \in \mathcal{R} \text{ or } E^c \in \mathcal{R}\}$ is a σ-algebra.
(d) If $\mathcal{R}$ is a σ-ring, then $\{E \subset X \mid E \cap F \in \mathcal{R} \text{ for all } F \in \mathcal{R}\}$ is a σ-algebra.

Solution.

(a) First, we make the following claim: Let $\{E_i\}_{i\in I}$ be a family of sets indexed by I, and let $E = \bigcup_{i\in I} E_i$, then

$$\bigcap_{i\in I} E_i = E \setminus \left(\bigcup_{i\in I} E \setminus E_i\right).$$

We will prove this by showing each is a subset of the other.

($\subseteq$) Let $e \in \bigcap_{i\in I} E_i$. Then $e \in E_i$ for all $i \in I$, and clearly $e \in E$. Also, since $e \in E_i$ for all i, we see that $e \notin E \setminus E_i$ for any $i \in I$. Then $e \notin \bigcup_{i\in I} E \setminus E_i$. Since $e \in E$ and $e \notin \bigcup_{i\in I} E \setminus E_i$, then $e \in E \setminus (\bigcup_{i\in I} E \setminus E_i)$.

($\supseteq$) Now assume $e \in E \setminus (\bigcup_{i\in I} E \setminus E_i)$. Then we have $e \in E$ and $e \notin E \setminus E_i$ for any i. That is to say $e \notin E$ or $e \in E_i$ for all $i \in I$. We've already established $e \in E$, so we must have $e \in E_i$ for all $i \in I$, then we have $e \in \bigcap_{i\in I} E_i$.

Now we have established the two sets are equal. Let $\mathcal{R}$ be a ring (σ-ring), and let $\{E_i\}_{i\in I}$ be a family of sets indexed by the finite (countable) set I. Then we have $E = \bigcup_{i\in I} E_i$ is the finite (countable) union of sets in $\mathcal{R}$, and so $E \in \mathcal{R}$. Since $\mathcal{R}$ is closed under finite set differences, we have that $\bigcap_{i\in I} E_i = E \setminus (\bigcup_{i\in I} E \setminus E_i) \in \mathcal{R}$. Therefore, any ring ($\sigma$-ring) $\mathcal{R}$ is closed under finite (countable) intersection.

(b) We suppose $\mathcal{R}$ is a ring (σ-ring) and is therefore closed under finite (countable) unions and differences. We also make the assumption that $\mathcal{R}$ is nonempty, as otherwise, $\mathcal{R}$ is trivially closed under unions, differences, and complements.

($\implies$) Suppose $\mathcal{R}$ is an algebra (σ-algebra). Then $\mathcal{R}$ is closed under complements. As $\mathcal{R}$ is nonempty, there exists $E \in \mathcal{R}$, where $E \subset X$. As $\mathcal{R}$ is closed under complements, $E^c \in \mathcal{R}$. As $\mathcal{R}$ is closed under finite (countable) unions, $E \cup E^c \in \mathcal{R}$. But $X = E \cup E^c$, so $X \in \mathcal{R}$.

($\impliedby$) Suppose $X \in \mathcal{R}$. As $\mathcal{R}$ is closed under differences, then for all $E \in \mathcal{R}$, we know $X \setminus E \in \mathcal{R}$. But $E^c = X \setminus E$, so for all $E \in \mathcal{R}$, we have that $E^c \in \mathcal{R}$, so $\mathcal{R}$ is closed under complements. Hence, as $\mathcal{R}$ is closed under complements and finite (countable) unions, then $\mathcal{R}$ is an algebra (σ-algebra).

(c) We let $S = \{E \subset X \mid E \in \mathcal{R} \text{ or } E^c \in \mathcal{R}\}$. To prove S is a σ-algebra, we must show it is closed under countable unions and complements.

To prove S is closed under complements, we consider an arbitrary subset, $E \in S$. Because $E \in S$, it must be that either $E \in \mathcal{R}$ or $E^c \in \mathcal{R}$. As $(E^c)^c = E$, this means that $(E^c)^c \in \mathcal{R}$ or $E^c \in \mathcal{R}$, which means that $E^c \in S$ by the definition of S. Hence, S is closed under taking complements.

To prove that S is closed under countable unions, let $\{E_j\}_{j\in\mathbb{Z}_{\geqslant 1}}$ be a countable collection of sets in S. We need to show that either $\bigcup_{j\in\mathbb{Z}_{\geqslant 1}} E_j \in \mathcal{R}$ or $(\bigcup_{j\in\mathbb{Z}_{\geqslant 1}} E_j)^c = \bigcap_{j\in\mathbb{Z}_{\geqslant 1}} E^c \in \mathcal{R}$. We know that for each $E_j \in S$, we have that either $E_j \in \mathcal{R}$ or $E_j^c \in \mathcal{R}$. Let $A = \{n \in \mathbb{Z}_{\geqslant 1} \mid E_n \in \mathcal{R}\}$ and note that this implies if $n \in \mathbb{Z}_{\geqslant 1} \setminus A$ then $E_n \notin \mathcal{R}$ and so $E_n^c \in \mathcal{R}$. As $\mathbb{Z}_{\geqslant 1}$ is countable, then A and $\mathbb{Z}_{\geqslant 1} \setminus A$ are both countable as subsets of a countable set are countable. Therefore, $\bigcup_{j\in A} E_j \in \mathcal{R}$ as this is a countable union of elements in $\mathcal{R}$ and $\mathcal{R}$ is a σ-ring, and $\bigcap_{j\in\mathbb{Z}_{\geqslant 1}\setminus A} E_j^c \in \mathcal{R}$ as σ-rings are closed under countable intersections by part (a) We note that we can always split up a set into a disjoint union $E = (E \cap F) \cup (E \cap F^c)$, which we use to say that

$$\bigcap_{j\in\mathbb{Z}_{\geqslant 1}\setminus A} E_j^c = \Big(\bigcap_{j\in\mathbb{Z}_{\geqslant 1}\setminus A} E_j^c \cap \bigcup_{j\in A} E_j\Big) \cup \Big(\bigcap_{j\in\mathbb{Z}_{\geqslant 1}\setminus A} E_j^c \cap (\bigcup_{j\in A} E_j)^c\Big)$$

noting that $(\bigcup_{j\in A} E_j)^c = \bigcap_{j\in A} E_j^c$ we get that

$$\bigcap_{j\in\mathbb{Z}_{\geqslant 1}} E_j^c = \Big(\bigcap_{j\in\mathbb{Z}_{\geqslant 1}\setminus A} E_j^c \cap (\bigcup_{j\in A} E_j)^c\Big) = \bigcap_{j\in\mathbb{Z}_{\geqslant 1}\setminus A} E_j^c \setminus \Big(\bigcap_{j\in\mathbb{Z}_{\geqslant 1}\setminus A} E_j^c \cap \bigcup_{j\in A} E_j\Big)$$

As $\bigcap_{j\in\mathbb{Z}_{\geqslant 1}\setminus A} E_j^c \setminus (\bigcap_{j\in\mathbb{Z}_{\geqslant 1}\setminus A} E_j^c \cap \bigcup_{j\in A} E_j)$ is taking the difference and finite intersection of elements of $\mathcal{R}$, and $\mathcal{R}$ is a σ-ring, this implies $\bigcap_{j\in\mathbb{Z}_{\geqslant 1}} E_j^c \in \mathcal{R}$. Therefore, $\bigcup_{j\in\mathbb{Z}_{\geqslant 1}} E_j = (\bigcap_{j\in\mathbb{Z}_{\geqslant 1}} E_j^c)^c$ is an element of S, and so S is closed under countable unions.

Hence, as S is closed under countable unions and complements, S is a σ-algebra.

(d) We let $S = \{E \subset X \mid E \cap F \in \mathcal{R} \text{ for all } F \in \mathcal{R}\}$. We want to show that S is a σ-algebra given that $\mathcal{R}$ is a σ-ring, which means we must show it is closed under countable unions and complements.

To show S is closed under complements, suppose $E \in S$. Then for all $F \in \mathcal{R}$, we have that $E \cap F \in \mathcal{R}$, and as $\mathcal{R}$ is closed under taking differences, we have that

$$E^c \cap F = F \setminus (E \cap F) \in \mathcal{R}$$

as F and $E \cap F$ are both in $\mathcal{R}$. Hence, for all $F \in \mathcal{R}$, we have that $E^c \cap F \in \mathcal{R}$, and so $E^c \in S$. As $E \in S$ was arbitrary, S is closed under taking complements.

To show S is closed under countable unions, let $\{E_j\}_{j\in\mathbb{Z}_{\geqslant 1}}$ be a countable collection of sets in S. Fix $F \in \mathcal{R}$. For each $E_j \in S$, we have that $E_j \cap F \in \mathcal{R}$. As $\mathcal{R}$ is a σ-ring, it is closed under countable unions, and so

$$\bigcup_{j\in\mathbb{Z}_{\geqslant 1}} (E_j \cap F) \in \mathcal{R}$$

We can show that

$$\bigcup_{j\in\mathbb{Z}_{\geqslant 1}} (E_j \cap F) = \Big(\bigcup_{j\in\mathbb{Z}_{\geqslant 1}} E_j\Big) \cap F$$

by observing that if $x \in \bigcup_{j\in\mathbb{Z}_{\geqslant 1}}(E_j \cap F)$, then $x \in (E_j \cap F)$ for some $j \in \mathbb{Z}_{\geqslant 1}$, which implies that $x \in E_j$ and $x \in F$, so $x \in (\bigcup_{j\in\mathbb{Z}_{\geqslant 1}} E_j)$ and $x \in F$, and therefore $x \in (\bigcup_{j\in\mathbb{Z}_{\geqslant 1}} E_j) \cap F$. Similarly, if $x \in (\bigcup_{j\in\mathbb{Z}_{\geqslant 1}} E_j) \cap F$, then $x \in F$ and $x \in \bigcup_{j\in\mathbb{Z}_{\geqslant 1}} E_j$, so $x \in F$ and $x \in E_j$ for some $j \in \mathbb{Z}_{\geqslant 1}$, so $x \in E_j \cap F$, and hence $x \in \bigcup_{j\in\mathbb{Z}_{\geqslant 1}}(E_j \cap F)$. Therefore, we know that

$$\left(\bigcup_{j\in\mathbb{Z}_{\geqslant 1}} E_j\right) \cap F = \bigcup_{j\in\mathbb{Z}_{\geqslant 1}}(E_j \cap F) \in \mathcal{R}$$

As $F \in \mathcal{R}$ is arbitrary, we have that for all $F \in \mathcal{R}$, $(\bigcup_{j\in\mathbb{Z}_{\geqslant 1}} E_j) \cap F \in \mathcal{R}$, and so $\bigcup_{j\in\mathbb{Z}_{\geqslant 1}} E_j \in S$ by definition. Therefore, S is closed under countable unions and complements, so S is a σ-algebra. □

Exercise 1.9: Folland Exercise 1.2.

Complete the proof of Proposition 2.

Solution. To complete the proof of Proposition 2, it suffices to show $\mathcal{B}_{\mathbb{R}} \subset \mathcal{B}(\mathcal{B}_j)$ for $2 \leqslant j \leqslant 8$. Folland shows that $\mathcal{M}(\mathcal{M}_j)$ for $1 \leqslant j \leqslant 8$ and $\mathcal{M}_{\mathbb{R}} \subset \mathcal{M}(\mathcal{M}_1)$ is trivial. We have the following representations:

- $\mathcal{M}(\mathcal{M}_2) : (a, b) = \bigcup_1^\infty (a + 1/n, b - 1/n]$.
- $\mathcal{M}(\mathcal{M}_3) : (a, b) = \bigcup_1^\infty (a, b - 1/n]$.
- $\mathcal{M}(\mathcal{M}_4) : (a, b) = \bigcup_1^\infty [a + 1/n, b)$.
- $\mathcal{M}(\mathcal{M}_5) : (a, b) = (a, \infty) \cap (-\infty, b)$.
- $\mathcal{M}(\mathcal{M}_6) : (a, b) = (a, \infty) \cap (-\infty, b)$.
- $\mathcal{M}(\mathcal{M}_7) : (a, b) = (a, \infty) \cap (-\infty, b)$.
- $\mathcal{M}(\mathcal{M}_8) : (a, b) = (a, \infty) \cap (-\infty, b)$.

It follows that every open set is generated by taking countable unions, complements, and intersections of sets from $\mathcal{E}_j$ for all $1 \leqslant j \leqslant 8$. □

Exercise 1.10: Folland Exercise 1.3.

Let $\mathcal{M}$ be an infinite σ-algebra.

(a) $\mathcal{M}$ contains an infinite sequence of disjoint sets.
(b) $\operatorname{card}(\mathcal{M}) \geqslant \mathfrak{c}$.

Exercise 1.11: Folland Exercise 1.4.

An algebra $\mathcal{A}$ is a σ-algebra if and only if $\mathcal{A}$ is closed under countable increasing unions (i.e., if $\{E_j\}_1^\infty \subset \mathcal{A}$ and $E_1 \subset E_2 \subset \cdots$, then $\bigcup_1^\infty E_j \in \mathcal{A}$).

Solution. Let $\mathcal{A}$ be a σ-algebra. Then by definition, $\mathcal{A}$ is closed under countable unions, so in particular $\mathcal{A}$ is closed under countable increasing unions.

Conversely, suppose $\mathcal{A}$ is an algebra and $\mathcal{A}$ is closed under countable increasing unions. Let $\{E_i\}_{i\in\mathbb{Z}_{\geqslant 1}}$ be a countable family of sets where $E_i \in \mathcal{A}$ for all $i \in \mathbb{Z}_{\geqslant 1}$. Define $F_i = \bigcup_{k=1}^{i} E_k$ for all $i \in \mathbb{Z}_{\geqslant 1}$. Since $\mathcal{A}$ is an algebra, it is closed under finite unions, and so $F_i \in \mathcal{A}$ for all $i \in \mathbb{Z}_{\geqslant 1}$. Also note that $\{F_i\}_{i\in\mathbb{Z}_{\geqslant 1}}$ as defined is an increasing family of sets, that is $F_i \subseteq F_{i+1}$ for all $i \in \mathbb{Z}_{\geqslant 1}$. Therefore we have $\bigcup_{i\in\mathbb{Z}_{\geqslant 1}} F_i \in \mathcal{A}$. Finally note that $\bigcup_{i\in\mathbb{Z}_{\geqslant 1}} F_i = \bigcup_{i\in\mathbb{Z}_{\geqslant 1}} E_i$ and so $\bigcup_{i\in\mathbb{Z}_{\geqslant 1}} E_i \in \mathcal{A}$. □

Exercise 1.12: Folland Exercise 1.5.

If $\mathcal{M}$ is the σ-algebra generated by $\mathcal{M}$, then $\mathcal{M}$ is the union of the σ-algebras generated by $\mathcal{M}$ as $\mathcal{M}$ ranges over all countable subsets of $\mathcal{M}$. (Hint: Show that the latter object is a σ-algebra.)

1.3 Measures

Let X be a set equipped with a σ-algebra $\mathcal{M}$. A measure on $\mathcal{M}$ (or on $(X, \mathcal{M})$, or simply on X if $\mathcal{M}$ is understood) is a function $\mu\colon \mathcal{M} \to [0, \infty]$ satisfying the following two properties.

(i) $\mu(\varnothing) = 0$.

(ii) If $\{E_j\}_1^\infty$ is a sequence of disjoint sets in $\mathcal{M}$, then $\mu(\bigcup_1^\infty E_j) = \sum_1^\infty \mu(E_j)$.

Property (ii) is called countable additivity. It implies finite additivity:

(ii') If $E_1, \dots E_n$ are disjoint sets in $\mathcal{M}$, then $\mu(\bigcup_1^n E_j) = \sum_1^n \mu(E_j)$,

because one can take $E_j = \varnothing$ for $j > n$. A function μ that satisfies (i) and (ii') but not necessarily (ii) is called a finitely additive measure.

If X is a set and $\mathcal{M} \subset \mathcal{M}(X)$ is a σ-algebra, $(X, \mathcal{M})$ is called a measurable space and the sets in $\mathcal{M}$ are called measurable sets. If μ is a measure on $(X, \mathcal{M})$, then $(X, \mathcal{M}, \mu)$ is called a measure space.

Let $(X, \mathcal{M}, \mu)$ be a measure space. Here is some standard terminology concerning the "size" of μ. If $\mu(X) < \infty$ (which implies that $\mu(E) < \infty$ for all $E \in \mathcal{M}$ since $\mu(X) = \mu(E) + \mu(E^c)$), μ is called finite. If $X = \bigcup_1^\infty E_j$ where $E_j \in \mathcal{M}$ and $\mu(E_j) < \infty$ for all j, μ is called σ-finite. More generally, if $E = \bigcup_1^\infty E_j$ where $E_j \in \mathcal{M}$ and $\mu(E_j) < \infty$ for all j, the set E is said to be σ-finite for μ. (It would be correct but more cumbersome to say that E is of σ-finite measure.) If for each $E \in \mathcal{M}$ with $\mu(E) = \infty$ there exists $F \in \mathcal{M}$ with $F \subset E$ and $0 < \mu(F) < \infty, \mu$ is called semifinite.

Every σ-finite measure is semifinite (Folland Exercise 1.13), but not conversely. Most measures that arise in parctice are σ-finite, which is fortunate since non- σ-finite measures tend to exhibit pathological behavior. The properties of non- σ-finite measures will be explored from time to time in the exercises.

Example 13. *Let us examine a few examples of measures. These examples are of a rather trivial nature, although the first one is of practical importance. The construction of more* *interesting examples is a task to which we shall turn in the next two sections.*

- *Let X be any nonempty set, $\mathcal{M} = \mathcal{M}(X)$, and f any function from X to $[0, \infty]$. Then f determines a measure μ on $\mathcal{M}$ by the formula $\mu(E) = \sum_{x \in E} f(x)$. (For the definition of such possibly uncountable sums, see Folland Section 0.5.) The reader may verify that μ is semifinite if and only if $f(x) < \infty$ for every $x \in X$, and μ is σ-finite if and only if μ is semifinite and $\{x \mid f(x) > 0\}$ is countable. Two special cases are of particular significance: If $f(x) = 1$ for all x, μ is called counting measure; and if, for some $x_0 \in X$, f is defined by $f(x_0) = 1$ and $f(x) = 0$ for $x \neq x_0$, μ is called the point mass or Dirac measure at x_0. (The same names are also applied to the restrictions of these measures to smaller σ-algebras on X.)*
- *Let X be an uncountable set, and let $\mathcal{M}$ be the σ-algebra of countable or cocountable sets. The function μ on $\mathcal{M}$ defined by $\mu(E) = 0$ if E is countable and $\mu(E) = 1$ if E is co-countable is easily seen to be a measure.*
- *Let X be an infinite set and $\mathcal{M} = \mathcal{M}(X)$. Define $\mu(E) = 0$ if E is finite, $\mu(E) = \infty$ if E is infinite. Then μ is a finitely additive measure but not a measure.*

The basic properties of measures are summarized in the following theorem.

Theorem 1.14: 1.8.

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

(a) (Monotonicity) If $E, F \in \mathcal{M}$ and $E \subset F$, then $\mu(E) \leqslant \mu(F)$.
(b) (Subadditivity) If $\{E_j\}_1^\infty \subset \mathcal{M}$, then $\mu(\bigcup_1^\infty E_j) \leqslant \sum_1^\infty \mu(E_j)$.
(c) (Continuity from below) If $\{E_j\}_1^\infty \subset \mathcal{M}$ and $E_1 \subset E_2 \subset \cdots$, then $\mu(\bigcup_1^\infty E_j) = \lim_{j\to\infty} \mu(E_j)$.
(d) (Continuity from above) If $\{E_j\}_1^\infty \subset \mathcal{M}, E_1 \supset E_2 \supset \cdots$, and $\mu(E_1) < \infty$, then $\mu(\bigcap_1^\infty E_j) = \lim_{j\to\infty} \mu(E_j)$.

Proof. (a) If $E \subset F$, then $\mu(F) = \mu(E) + \mu(F \setminus E) \geqslant \mu(E)$.

(b) Let $F_1 = E_1$ and $F_k = E_k \setminus (\bigcup_1^{k-1} E_j)$ for $k > 1$. Then the F_ks are disjoint and $\bigcup_1^n F_j = \bigcup_1^n E_j$ for all n. Therefore, by (a),

$$\mu\Big(\bigcup_1^\infty E_j\Big) = \mu\Big(\bigcup_1^\infty F_j\Big) = \sum_1^\infty \mu(F_j) \leqslant \sum_1^\infty \mu(E_j)$$

(c) Setting $E_0 = \varnothing$, we have

$$\mu\Big(\bigcup_1^\infty E_j\Big) = \sum_1^\infty \mu(E_j \setminus E_{j-1}) = \lim_{n\to\infty} \sum_1^n \mu(E_j \setminus E_{j-1}) = \lim_{n\to\infty} \mu(E_n)$$

(d) Let $F_j = E_1 \setminus E_j$; then $F_1 \subset F_2 \subset \cdots, \mu(E_1) = \mu(F_j) + \mu(E_j)$, and $\bigcup_1^\infty F_j = E_1 \setminus (\bigcap_1^\infty E_j)$. By (c), then,

$$\mu(E_1) = \mu\Big(\bigcap_1^\infty E_j\Big) + \lim_{j\to\infty} \mu(F_j) = \mu\Big(\bigcap_1^\infty E_j\Big) + \lim_{j\to\infty} [\mu(E_1) - \mu(E_j)]$$

Since $\mu(E_1) < \infty$, we may subtract it from both sides to yield the desired result. □

We remark that the condition $\mu(E_1) < \infty$ in part (d) could be replaced by $\mu(E_n) < \infty$ for some $n > 1$, as the first $n-1 E_j$s can be discarded from the sequence without affecting the intersection. However, some finiteness assumption is necessary, as it can happen that $\mu(E_j) = \infty$ for all j but $\mu(\bigcap_1^\infty E_j) < \infty$. (For example, let μ be counting measure on $(\mathbb{Z}_{\geqslant 1}, \mathcal{P}(\mathbb{Z}_{\geqslant 1}))$ and let $E_j = \{n \mid n \geqslant j\}$; then $\bigcap_1^\infty E_j = \varnothing$.)

If $(X, \mathcal{M}, \mu)$ is a measure space, a set $E \in \mathcal{M}$ such that $\mu(E) = 0$ is called a null set. By subadditivity, any countable union of null sets is a null set, a fact which we shall use frequently. If a statement about points $x \in X$ is true except for x in some null set, we say that it is true almost everywhere (abbreviated a.e.), or for almost every x. (If more precision is needed, we shall speak of a $\boldsymbol{\mu}$-null set, or $\boldsymbol{\mu}$-almost everywhere).

If $\mu(E) = 0$ and $F \subset E$, then $\mu(F) = 0$ by monotonicity provided that $F \in \mathcal{M}$, but in general it need not be true that $F \in \mathcal{M}$. A measure whose domain includes all subsets of null sets is called complete. Completeness can sometimes obviate annoying technical points, and it can always be achieved by enlarging the domain of μ, as follows.

Theorem 1.15: 1.9.

Suppose that $(X, \mathcal{M}, \mu)$ is a measure space. Let $\mathcal{M} = \{N \in \mathcal{M} \mid \mu(N) = 0\}$ and $\overline{\mathcal{M}} = \{E \cup F \colon E \in \mathcal{M} \text{ and } F \subset N \text{ for some } N \in \mathcal{M}\}$. Then $\overline{\mathcal{M}}$ is a σ-algebra, and there is a unique extension $\overline{\mu}$ of μ to a complete measure on $\overline{\mathcal{M}}$.

Proof. Since $\mathcal{M}$ and $\mathcal{M}$ are closed under countable unions, so is $\overline{\mathcal{M}}$. If $E \cup F \in \overline{\mathcal{M}}$ where $E \in \mathcal{M}$ and $F \subset N \in \mathcal{M}$, we can assume that $E \cap N = \varnothing$ (otherwise, replace F and N by $F \setminus E$ and $N \setminus E$). Then $E \cup F = (E \cup N) \cap (N^c \cup F)$, so $(E \cup F)^c = (E \cup N)^c \cup (N \setminus F)$. But $(E \cup N)^c \in \mathcal{M}$ and $N \setminus F \subset N$, so that $(E \cup F)^c \in \overline{\mathcal{M}}$. Thus $\overline{\mathcal{M}}$ is a σ-algebra.

If $E \cup F \in \overline{\mathcal{M}}$ as above, we set $\overline{\mu}(E \cup F) = \mu(E)$. This is well defined, since if $E_1 \cup F_1 = E_2 \cup F_2$ where $F_j \subset N_j \in \mathcal{M}$, then $E_1 \subset E_2 \cup N_2$ and so $\mu(E_1) \leqslant \mu(E_2) + \mu(N_2) = \mu(E_2)$, and likewise $\mu(E_2) \leqslant \mu(E_1)$. It is easily verified that $\overline{\mu}$ is a complete measure on $\overline{\mathcal{M}}$, and that $\overline{\mu}$ is the only measure on $\overline{\mathcal{M}}$ that extends μ; details are left to the reader (Folland Exercise 1.6). □

The measure $\overline{\mu}$ in Theorem 15 is called the ***completion*** of μ, and $\overline{\mathcal{M}}$ is called the completion of $\mathcal{M}$ with respect to μ.

Exercise 1.16: Folland Exercise 1.6.

Complete the proof of Theorem 15.

Exercise 1.17: Folland Exercise 1.7.

If $\mu_1, \ldots, \mu_n$ are measures on $(X, \mathcal{M})$ and $a_1, \ldots, a_n \in [0, \infty)$, then $\sum_{j=1}^n a_j \mu_j$ is a measure on $(X, \mathcal{M})$.

Solution. We have
$$\mu(\varnothing) = \sum_{j=1}^{n} a_j\mu_j(\varnothing) = \sum_{j=1}^{n} a_j \cdot 0 = 0,$$
where the second equality is because μ_j is a measure for each $j \in \{1, \ldots, n\}$. Now let $\{E_j\}_{j=1}^{\infty}$ be a countable subset of $\mathcal{M}$ consisting of mutually pairwise disjoint subsets. Then
$$\begin{aligned}
\mu\Big(\bigcup_{j=1}^{\infty} E_j\Big) &= \sum_{j=1}^{n} a_j\mu_j\Big(\bigcup_{k=1}^{\infty} E_k\Big) \\
&= \sum_{j=1}^{n} a_j\Big(\sum_{k=1}^{\infty} \mu_j(E_k)\Big) && \text{(by countable additivity of each } \mu_j\text{)} \\
&= \sum_{j=1}^{n} a_j \lim_{N\to\infty}\Big(\sum_{k=1}^{N} \mu_j(E_k)\Big) && \text{(definition of infinite sum)} \\
&= \lim_{N\to\infty}\Big(\sum_{j=1}^{n} a_j \sum_{k=1}^{N} \mu_j(E_k)\Big) && \text{(limit of a finite linear combination)} \\
&= \sum_{k=1}^{\infty} \mu(E_j), && \text{(finite linear combination of limit of sum)}
\end{aligned}$$
so μ is countably additive. Thus μ is a measure. □

Exercise 1.18: Folland Exercise 1.8.

If $(X, \mathcal{M}, \mu)$ is a measure space and $\{E_j\}_{j=1}^{\infty} \subset \mathcal{M}$, then
$$\mu(\liminf_{j\to\infty} E_j) \leqslant \liminf_{j\to\infty} \mu(E_j).$$
Also,
$$\mu(\limsup_{j\to\infty} E_j) \geqslant \limsup_{j\to\infty} \mu(E_j)$$
provided that $\mu(\bigcup_{j=1}^{\infty} E_j) < \infty$.

Solution. Let $\{E_j\}_{j=1}^{\infty}$ be any countable collection of elements of $\mathcal{M}$, where $\mathcal{M}$ is the σ-algebra on which μ is defined.

- $\mu(\liminf_{j\to\infty} E_j) \leqslant \liminf_{j\to\infty} \mu(E_j)$: We have
$$\mu(\liminf_{j\to\infty} E_j) = \mu\Big(\bigcup_{n\in\mathbb{Z}_{\geqslant 1}} \bigcap_{j\geqslant n} E_j\Big) = \lim_{n\to\infty} \mu\Big(\bigcap_{j\geqslant n} E_j\Big) \leqslant \liminf_{j\to\infty} E_j,$$
where the third equality is by continuity from below and the final inequality can be argued as follows. Since
$$\bigcap_{j\geqslant n} E_j \subset E_j \text{ for all } j \geqslant n,$$
by monotonicity of μ, we have
$$\mu\Big(\bigcap_{j\geqslant n} E_j\Big) \subset \mu(E_j) \text{ for all } j \geqslant n.$$
Thus $\mu(\bigcap_{j\geqslant n} E_j)$ is a lower bound for $\{\mu(E_j) \mid j \geqslant n\}$. It follows that
$$\mu\Big(\bigcap_{j\geqslant n} E_j\Big) \leqslant \inf_{j\geqslant n} E_j.$$

Sending $n \to \infty$, we obtain

$$\mu\Big(\bigcap_{j\geqslant n} E_j\Big) \leqslant \lim_{n\to\infty} \inf_{j\geqslant n} E_j = \liminf_{n\to\infty} \mu(E_j).$$

- $\limsup_{n\to\infty} \mu(E_n) \leqslant \mu(\limsup_{j\to\infty} E_j)$ if $\mu(\bigcup_{j=1}^{\infty} E_j) < \infty$: Suppose $\mu(\bigcup_{j=1}^{\infty} E_j) < \infty$. Then

$$\limsup_{j\to\infty} \mu(E_j) = \lim_{n\to\infty} \sup_{j\geqslant n} \mu(E_j) \leqslant \mu(\limsup_{j\to\infty} E_j),$$

where the final inequality holds because

$$\mu(\limsup_{j\to\infty} E_j) = \mu\Big(\bigcap_{n\in\mathbb{Z}_{\geqslant 1}} \bigcup_{j\geqslant n} E_j\Big) = \lim_{n\to\infty} \mu\Big(\bigcup_{j\geqslant n} E_j\Big) \geqslant \lim_{n\to\infty} \sup_{j\geqslant n} \mu(E_j),$$

where the second last equality is by continuity from above since $\mu(\bigcup_{j=1}^{\infty} E_j) < \infty$ and the inequality is because $\mu(\bigcup_{j\geqslant n} E_j)$ is an upper bound for $\{\mu(E_j) \mid j \geqslant n\}$, for all $n \in \mathbb{Z}_{\geqslant 1}$. □

Exercise 1.19: Folland Exercise 1.9 (Strenthened Version).

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

(a) If $E, F \in \mathcal{M}$, then

$$\mu(E) + \mu(F) = \mu(E \cup F) + \mu(E \cap F).$$

(b) If $E \subset X$ is μ^*-measurable, then for every subset $A \subset X$ we have

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

Note that only one subset needs to be μ^*-measurable, unlike for measures!

Solution.

(a) First note that $E \cup F, E \cap F, E\backslash F \in \mathcal{M}$. Thus

$$\mu(E) + \mu(F) = \mu(E\backslash F) + \mu(E \cap F) + \mu(F) = \mu(E \cup F) + \mu(E \cap F),$$

where we have used the facts

$$(E\backslash F) \cap (E \cap F) = \varnothing = (E\backslash F) \cap F.$$

(b) Since E is μ^*-measurable, we can write

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

Applying that E is μ^*-measurable once again yields

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

Combining the last two equations proves that

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

Exercise 1.20: Folland Exercise 1.10.

Given a measure space $(X, \mathcal{M}, \mu)$ and $E \in \mathcal{M}$, define $\mu_E(\mathcal{M}) = \mu(A \cap E)$ for $A \in \mathcal{M}$. Then μ_E is a measure.

Solution. First observe that $\mu_E(\varnothing) = \mu(\varnothing \cap E) = \mu(\varnothing) = 0$. Next, given a sequence of disjoint sets in $\mathcal{M}$, $\{E_j\}_{j\in\mathbb{Z}_{\geqslant 1}}$, note that $\{E_j \cap E\}_{j\in\mathbb{Z}_{\geqslant 1}}$ are disjoint and

$$\mu_E\Big(\bigcup\nolimits_{j\in\mathbb{Z}_{\geqslant 1}} E_j\Big) = \mu\Big(\Big(\bigcup\nolimits_{j\in\mathbb{Z}_{\geqslant 1}} E_j\Big) \cap E\Big) = \mu\Big(\bigcup\nolimits_{j\in\mathbb{Z}_{\geqslant 1}} (E_j \cap E)\Big) = \sum\nolimits_{j\in\mathbb{Z}_{\geqslant 1}} \mu(E_j \cap E) = \sum\nolimits_{j\in\mathbb{Z}_{\geqslant 1}} \mu_E(E_j$$

Thus μ_E is a measure. □

Exercise 1.21: Folland Exercise 1.11.

A finitely additive measure μ is a measure if and only if it is continuous from below as in Theorem 14(c). If $\mu(X) < \infty$, μ is a measure if and only if it is continuous from above as in Theorem 14(d).

Exercise 1.22: Folland Exercise 1.12.

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

(a) If $E, F \in \mathcal{M}$ and $\mu(E\Delta F) = 0$, then $\mu(E) = \mu(F)$.
(b) Say that $E \sim F$ if $\mu(E \triangle F) = 0$; then $\sim$ is an equivalence relation on $\mathcal{M}$.
(c) For $E, F \in \mathcal{M}$, define $\rho(E, F) = \mu(E\Delta F)$. Then $\rho(E, G) \leqslant \rho(E, F) + \rho(F, G)$, and hence ρ defines a metric on the space $\mathcal{M}/\sim$ of equivalence classes.

Solution.

(a) First, recall that the symmetric difference is defined by $E\Delta F = (E\backslash F) \cup (F\backslash E)$. Since $E\backslash F$ and $F\backslash E$ are disjoint, we have

$$\mu(E\Delta F) = \mu(E\backslash F) + \mu(F\backslash E),$$

so that if $\mu(E\Delta F) = 0$ (since μ only takes on nonnegative values) we have

$$\mu(E\backslash F) = 0 = \mu(F\backslash E).$$

Therefore,

$$\mu(E) = \mu(E \cap F) + \mu(E\backslash F) = \mu(E \cap F) + \mu(F\backslash E) = \mu(F).$$

(b) Reflexive: As $\mu(E\Delta E) = \mu(\varnothing) = 0$ for all $E \in \mathcal{M}$, we have $E \sim E$ for all $E \in \mathcal{M}$.

Symmetric: Since $E\Delta F = F\Delta E$ by definition (as the name symmetric difference suggests!), if $E \sim F$ then $0 = \mu(E\Delta F) = \mu(F\Delta E)$ so that $F \sim E$ for all $E, F \in \mathcal{M}$.

Transitive: Suppose that $E \sim F$ and $F \sim G$ for $E, F, G \in \mathcal{M}$. Then we have that

$$\mu(E\Delta F) = 0 = \mu(F\Delta G).$$

As we want to show that $\mu(E\Delta G) = 0$, we first note that both $E\backslash G = E \cap G^c$ and $G\backslash E = G \cap E^c$ are measurable since E and G are. Furthermore, one sees

$$E\backslash G \subset (E\backslash F) \cup (F\backslash G), \quad G\backslash E \subset (G\backslash F) \cup (F\backslash E),$$

so that both $E\backslash G$ and $G\backslash E$ are null sets. Therefore, we conclude $\mu(E\Delta G) = 0$ so that $E \sim G$.

(c) By definition, we have that $\rho\colon \mathcal{M}/\sim \to [0,\infty]$. Furthermore, $\mu(E\Delta F) = \rho(E,F) = 0$ if and only if $E \sim F$, that is, E and F are in the same equivalence class. As we just proved, μ, and hence ρ, is symmetric. By our remarks in proving b., it is clear that

$$E\Delta G \subset (E\Delta F) \cup (F\Delta G).$$

Hence, by monotonicity we have

$$\rho(E\Delta G) = \mu(E\Delta G) \leqslant \mu(E\Delta F) + \mu(F\Delta G) = \rho(E\Delta F) + \rho(F\Delta G),$$

so that the triangle inequality holds. Therefore, ρ defines a metric on the space $\mathcal{M}/\sim$ of equivalence classes.

□

Exercise 1.23: Folland Exercise 1.13.

Every σ-finite measure is semifinite.

Solution. Let μ be σ-finite, and fix $E \in \mathfrak{M}$ with $\mu(E) = +\infty$. Since μ is σ-finite, it can be expressed as a countable union of sets $\{U_i\}_{i=1}^\infty$ of finite measure. We have $E \sqcup E^c = X = \bigcup_{i=1}^\infty U_i$, which implies

$$E = \left(\bigcup_{i=1}^\infty U_i\right) \setminus E^c = \bigcup_{i=1}^\infty (U_i \setminus E^c) = \bigcup_{i=1}^\infty (U_i \cap E).$$

Therefore,

$$+\infty = \mu(E) = \mu\left(\bigcup_{i=1}^\infty (U_i \cap E)\right) \leqslant \sum_{i=1}^\infty \mu(U_i \cap E).$$

Now, for each i, we know that $U_i \cap E \subset E$, and $\mu(U_i \cap E) < \infty$ since $\mu(U_i) < \infty$. It remains to show that $0 < \mu(U_k \cap E) < \infty$ for some k. Suppose there were no such k; then for all i, we would have $\mu(U_i \cap E) = \sum_{i=1}^\infty (0) = 0$, which contradicts our assumption that $\mu(E) = +\infty$. This completes the proof. □

Exercise 1.24: Folland Exercise 1.14.

If μ is a semifinite measure and $\mu(E) = \infty$, for any $C > 0$ there exists $F \subset E$ with $C < \mu(F) < \infty$.

Exercise 1.25: Folland Exercise 1.15.

Given a measure μ on $(X, \mathcal{M})$, define μ_0 on $\mathcal{M}$ by

$$\mu_0(E) = \sup\{\mu(F) \mid F \subset E \text{ and } \mu(F) < \infty\}.$$

(a) μ_0 is a semifinite measure. It is called the semifinite part of μ.
(b) If μ is semifinite, then $\mu = \mu_0$. (Use Folland Exercise 1.14.)
(c) There is a measure ν on $\mathcal{M}$ (in general, not unique) which assumes only the values 0 and ∞ such that $\mu = \mu_0 + \nu$.

Exercise 1.26: Folland Exercise 1.16.

Let $(X, \mathcal{M}, \mu)$ be a measure space. A set $E \subset X$ is called locally measurable if $E \cap A \in \mathcal{M}$ for all $A \in \mathcal{M}$ such that $\mu(A) < \infty$. Let $\tilde{\mathcal{M}}$ be the collection of all locally measurable sets. Clearly $\mathcal{M} \subset \tilde{\mathcal{M}}$; if $\mathcal{M} = \tilde{\mathcal{M}}$, then μ is called saturated.

(a) If μ is σ-finite, then μ is saturated.
(b) $\tilde{\mathcal{M}}$ is a σ-algebra.
(c) Define $\tilde{\mu}$ on $\tilde{\mathcal{M}}$ by $\tilde{\mu}(E) = \mu(E)$ if $E \in \mathcal{M}$ and $\tilde{\mu}(E) = \infty$ otherwise. Then $\tilde{\mu}$ is a saturated measure on $\tilde{\mathcal{M}}$, called the saturation of μ.
(d) If μ is complete, so is $\tilde{\mu}$.
(e) Suppose that μ is semifinite. For $E \in \tilde{\mathcal{M}}$, define $\underline{\mu}(E) = \sup\{\mu(A) \mid A \in \mathcal{M} \text{ and } A \subset E\}$. Then $\underline{\mu}$ is a saturated measure on $\tilde{\mathcal{M}}$ that extends μ.
(f) Let X_1, X_2 be disjoint uncountable sets, $X = X_1 \cup X_2$, and $\mathcal{M}$ the σ-algebra of countable or co-countable sets in X. Let μ_0 be counting measure on $\mathcal{M}(X_1)$, and define μ on $\mathcal{M}$ by $\mu(E) = \mu_0(E \cap X_1)$. Then μ is a measure on $\mathcal{M}$, $\tilde{\mathcal{M}} = \mathcal{M}(X)$, and in the notation of parts (c) and (e), $\tilde{\mu} \neq \underline{\mu}$.

Exercise 1.27: The Borel-Cantelli Lemma.

Let $(X, \mathfrak{M}, \mu)$ be a measure space, let $\{E_j\}_{j=1}^{\infty} \subset \mathfrak{M}$, and let $\limsup E_j$ denote the set of points that lie in infinitely many of the E_j. If

$$\sum_{j=1}^{\infty} \mu(E_j) < \infty,$$

then

$$\mu(\limsup E_j) = 0.$$

Solution. Notice that $\mu(\bigcup_{j=1}^{\infty} E_j) < +\infty$ since

$$\begin{aligned}\mu\Big(\bigcup_{j=1}^{\infty} E_j\Big) &\leqslant \sum_{j=1}^{\infty} \mu(E_j) && \text{(by monotonicity)} \\ &< +\infty && \text{(given).}\end{aligned}$$

So, by continuity from above, we have

$$\begin{aligned}\mu(\limsup E_j) &= \mu\Big(\bigcap_{k=1}^{\infty}\bigcup_{j=k}^{\infty} E_j\Big) && \text{(definition of } \limsup E_j) \\ &= \lim_{k\to\infty} \mu\Big(\bigcup_{j=k}^{\infty} E_j\Big) && \text{(continuity from below)} \\ &\leqslant \lim_{k\to\infty} \sum_{j=k}^{\infty} \mu(E_j) && \text{(monotonicity)}\end{aligned}$$

Since we know that $\sum_{j=1}^{\infty} \mu(E_j) < +\infty$, it follows that $\lim_{k\to\infty} \sum_{j=k}^{\infty} \mu(E_j) = 0$. Hence, the above limit is zero, as desired. □

1.4 Outer Measures

In this section we develop the tools we shall use to construct measures. To motivate the ideas, it may be useful to recall the procedure used in calculus to define the area of a bounded region E in the plane $\mathbb{R}^2$. One draws a grid of rectangles in the plane and approximates the area of E from below by the sum of the areas of the rectangles in the grid that are subsets of E, and from above by the sum of the areas of the rectangles in the grid that intersect E. The limits of these approximations as the grid is taken finer and finer give the "inner area" and "outer area" of E, and if they are equal, their common value is the "area" of E. (We shall discuss these matters in more detail in 2.6.) The key idea here is that of outer area, since if R is a large rectangle containing E, the inner area of E is just the area of R minus the outer area of $R \setminus E$.

The abstract generalization of the notion of outer area is as follows:

Definition 28. *An* ***outer measure*** *on a nonempty set X is a function $\mu^*: \mathcal{P}(X) \to [0,\infty]$ that satisfies*

- $\mu^*(\varnothing) = 0$,
- $\mu^*(A) \leqslant \mu^*(B)$ *if* $A \subset B$, *and*
- $\mu^*(\bigcup_1^{\infty} A_j) \leqslant \sum_1^{\infty} \mu^*(A_j)$.

The most common way to obtain outer measures is to start with a family $\mathcal{E}$ of "elementary sets" on which a notion of measure is defined (such as rectangles in the plane) and then to approximate arbitrary sets "from the outside" by countable unions of members of $\mathcal{E}$. The precise construction is as follows.

Proposition 1.29: 1.10.

Let $\mathcal{E} \subset \mathcal{E}(X)$ and $\rho\colon \mathcal{E} \to [0,\infty]$ be such that $\varnothing \in \mathcal{E}, X \in \mathcal{E}$, and $\rho(\varnothing) = 0$. For any $A \subset X$, define

$$\mu^*(A) = \inf\Big\{\sum_1^{\infty} \mu(E_j) \;\Big|\; E_j \in \mathcal{E} \text{ and } A \subset \bigcup_1^{\infty} E_j\Big\}$$

Then μ^* is an outer measure.

Proof. For any $A \subset X$ there exists $\{E_j\}_1^\infty \subset \mathcal{E}$ such that $A \subset \bigcup_1^\infty E_j$ (take $E_j = X$ for all j) so the definition of μ^* makes sense. Obviously $\mu^*(\varnothing) = 0$ (take $E_j = \varnothing$ for all j), and $\mu^*(A) \leqslant \mu^*(B)$ for $A \subset B$ because the set over which the infimum is taken in the definition of $\mu^*(A)$ includes the corresponding set in the definition of $\mu^*(B)$. To prove the countable subadditivity, suppose $\{A_j\}_1^\infty \subset \mathcal{E}(X)$ and $\varepsilon > 0$. For each j there exists $\{E_j^k\}_{k=1}^\infty \subset \mathcal{E}$ such that $A_j \subset \bigcup_{k=1}^\infty E_j^k$ and $\sum_{k=1}^\infty \rho(E_j^k) \leqslant \mu^*(A_j) + \varepsilon 2^{-j}$. But then if $A = \bigcup_1^\infty A_j$, we have $A \subset \bigcup_{j,k=1}^\infty E_j^k$ and $\sum_{j,k} \rho(E_j^k) \leqslant \sum_j \mu^*(A_j) + \varepsilon$, whence $\mu^*(A) \leqslant \sum_j \mu^*(A_j) + \varepsilon$. Since ε is arbitrary, we are done. □

The fundamental step that leads from outer measures to measures is as follows:

Definition 30. *If μ^* is an outer measure on X, a set $A \subset X$ is called **μ^*-measurable** if*

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

Of course, the inequality $\mu^*(E) \leqslant \mu^*(E \cap A) + \mu^*(E \cap A^c)$ holds for any A and E, so to prove that A is μ^*-measurable, it suffices to prove the reverse inequality. The latter is trivial if $\mu^*(E) = \infty$, so we see that A is μ^*-measurable iff

$$\mu^*(E) \geqslant \mu^*(E \cap A) + \mu^*(E \cap A^c) \text{ for all } E \subset X \text{ such that } \mu^*(E) < \infty.$$

Some motivation for the notion of μ^*-measurability can be obtained by referring to the discussion at the beginning of this section. If E is a "well-behaved" set such that $E \supset A$, the equation $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$ says that the outer measure of $A, \mu^*(A)$, is equal to the "inner measure" of $A, \mu^*(E) - \mu^*(E \cap A^c)$. The leap from "well-behaved" sets containing A to arbitrary subsets of X a large one, but it is justified by the following theorem.

Theorem 1.31: 1.11: Carathéodory's Theorem.

If μ^* is an outer measure on X, the collection $\mathcal{M}$ of μ^*-measurable sets is a σ-algebra, and the restriction of μ^* to $\mathcal{M}$ is a complete measure.

Proof. First, we observe that $\mathcal{M}$ is closed under complements since the definition of μ^*-measurability of A is symmetric in A and A^c. Next, if $A, B \in \mathcal{M}$ and $E \subset X$,

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

But $(A \cup B) = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$, so by subadditivity,

$$\mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B) \geqslant \mu^*(E \cap (A \cup B))$$

and hence

$$\mu^*(E) \geqslant \mu^*(E \cap (A \cup B)) + \mu^*(E \cap (A \cup B)^c)$$

It follows that $A \cup B \in \mathcal{M}$, so $\mathcal{M}$ is an algebra. Moreover, if $A, B \in \mathcal{M}$ and $A \cap B = \varnothing$,

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

so μ^* is finitely additive on $\mathcal{M}$.

To show that $\mathcal{M}$ is a σ-algebra, it will suffice to show that $\mathcal{M}$ is closed under countable disjoint unions. If $\{A_j\}_1^\infty$ is a sequence of disjoint sets in $\mathcal{M}$, let $B_n = \bigcup_1^n A_j$ and $B = \bigcup_1^\infty A_j$. Then for any $E \subset X$,

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

so a simple induction shows that $\mu^*(E \cap B_n) = \sum_1^n \mu^*(E \cap A_j)$. Therefore,

$$\mu^*(E) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geqslant \sum_1^n \mu^*(E \cap A_j) + \mu^*(E \cap B^c)$$

and letting $n \to \infty$ we obtain

$$\begin{aligned}\mu^*(E) &\geqslant \sum_1^\infty \mu^*(E \cap A_j) + \mu^*(E \cap B^c) \geqslant \mu^*\Big(\bigcup_1^\infty (E \cap A_j)\Big) + \mu^*(E \cap B^c)\\ &= \mu^*(E \cap B) + \mu^*(E \cap B^c) \geqslant \mu^*(E)\end{aligned}$$

All the inequalities in this last calculation are thus equalities. It follows that $B \in \mathcal{M}$ and—taking $E = B$—that $\mu^*(B) = \sum_1^\infty \mu^*(A_j)$, so μ^* is countably additive on $\mathcal{M}$. Finally, if $\mu^*(A) = 0$, for any $E \subset X$ we have

$$\mu^*(E) \leqslant \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E \cap A^c) \leqslant \mu^*(E)$$

so that $A \in \mathcal{M}$. Therefore $\mu^* \mid \mathcal{M}$ is a complete measure. □

Our first applications of Carathéodory's theorem will be in the context of extending measures from algebras to σ-algebras. More precisely, if $\mathcal{A} \subset \mathcal{A}(X)$ is an algebra, a function $\mu_0 \colon \mathcal{A} \to [0, \infty]$ will be called a premeasure if

- $\mu_0(\varnothing) = 0$,
- if $\{A_j\}_1^\infty$ is a sequence of disjoint sets in $\mathcal{A}$ such that $\bigcup_1^\infty A_j \in \mathcal{A}$, then $\mu_0(\bigcup_1^\infty A_j) = \sum_1^\infty \mu_0(A_j)$.

In particular, a premeasure is finitely additive since one can take $A_j = \varnothing$ for j large. The notions of finite and σ-finite premeasures are defined just as for measures. If μ_0 is a premeasure on $\mathcal{A} \subset \mathcal{A}(X)$, it induces an outer measure on X in accordance with Proposition 29, namely,

$$\mu^*(E) = \inf\Big\{\sum_1^\infty \mu_0(A_j) \,\Big|\, A_j \in \mathcal{A}, E \subset \bigcup_1^\infty A_j\Big\}$$

Proposition 1.32: 1.13.

If μ_0 is a premeasure on $\mathcal{A}$ and μ^* is defined by (1.12), then

(a) $\mu^* \mid \mathcal{A} = \mu_0$;
(b) every set in $\mathcal{A}$ is μ^* measurable.

Proof. (a) Suppose $E \in \mathcal{A}$. If $E \subset \bigcup_1^\infty A_j$ with $A_j \in \mathcal{A}$, let $B_n = E \cap (A_n \setminus \bigcup_1^{n-1} A_j)$.

Then the B_ns are disjoint members of $\mathcal{A}$ whose union is E, so $\mu_0(E) = \sum_1^\infty \mu_0(B_j) \leqslant \sum_1^\infty \mu_0(A_j)$. It follows that $\mu_0(E) \leqslant \mu^*(E)$, and the reverse inequality is obvious since $E \subset \bigcup_1^\infty A_j$ where $A_1 = E$ and $A_j = \varnothing$ for $j > 1$.

(b) If $A \in \mathcal{A}, E \subset X$, and $\varepsilon > 0$, there is a sequence $\{B_j\}_1^\infty \subset \mathcal{A}$ with $E \subset \bigcup_1^\infty B_j$ and $\sum_1^\infty \mu_0(B_j) \leqslant \mu^*(E) + \varepsilon$. Since μ_0 is additive on $\mathcal{A}$,

$$\mu^*(E) + \varepsilon \geqslant \sum_1^\infty \mu_0(B_j \cap A) + \sum_1^\infty \mu_0(B_j \cap A^c) \geqslant \mu^*(E \cap A) + \mu^*(E \cap A^c)$$

Since ε is arbitrary, A is μ^*-measurable. □

Theorem 1.33: 1.14.

Let $\mathcal{A} \subset \mathcal{A}(X)$ be an algebra, μ_0 a premeasure on $\mathcal{A}$, and $\mathcal{A}$ the σ-algebra generated by $\mathcal{A}$. There exists a measure μ on $\mathcal{A}$ whose restriction to $\mathcal{A}$ is μ_0—namely, $\mu = \mu^* \mid \mathcal{A}$ where μ^* is given by (1.12). If ν is another measure on $\mathcal{A}$ that extends μ_0, then $\nu(E) \leqslant \mu(E)$ for all $E \in \mathcal{A}$, with equality when $\mu(E) < \infty$. If μ_0 is σ-finite, then μ is the unique extension of μ_0 to a measure on $\mathcal{A}$.

Proof. The first assertion follows from Carathéodory's theorem and Proposition 32 since the σ-algebra of μ^*-measurable sets includes $\mathcal{A}$ and hence $\mathcal{A}$. As for the second assertion, if $E \in \mathcal{A}$ and $E \subset \bigcup_1^\infty A_j$ where $A_j \in \mathcal{A}$, then $\nu(E) \leqslant \sum_1^\infty \nu(A_j) = \sum_1^\infty \mu_0(A_j)$, whence $\nu(E) \leqslant \mu(E)$. Also, if we set $A = \bigcup_1^\infty A_j$, we have

$$\nu(A) = \lim_{n\to\infty} \nu\Big(\bigcup_1^n A_j\Big) = \lim_{n\to\infty} \mu\Big(\bigcup_1^n A_j\Big) = \mu(A)$$

If $\mu(E) < \infty$, we can choose the A_js so that $\mu(A) < \mu(E) + \varepsilon$, hence $\mu(A \setminus E) < \varepsilon$, and

$$\mu(E) \leqslant \mu(A) = \nu(A) = \nu(E) + \nu(A \setminus E) \leqslant \nu(E) + \mu(A \setminus E) \leqslant \nu(E) + \varepsilon$$

Since ε is arbitrary, $\mu(E) = \nu(E)$. Finally, suppose $X = \bigcup_1^\infty A_j$ with $\mu_0(A_j) < \infty$, where we can assume that the A_js are disjoint. Then for any $E \in \mathcal{M}$,

$$\mu(E) = \sum_1^\infty \mu(E \cap A_j) = \sum_1^\infty \nu(E \cap A_j) = \nu(E)$$

so $\nu = \mu$. □

The proof of this theorem yields more than the statement. Indeed, μ_0 may be extended to a measure on the algebra $\mathcal{M}^*$ of all μ^*-measurable sets. The relation between $\mathcal{M}$ and $\mathcal{M}^*$ is explored in Folland Exercise 1.22 (along with Folland Exercise 1.20(b), which ensures that the outer measures induced by μ_0 and μ are the same).

Exercise 1.34: Folland Exercise 1.17.

If μ^* is an outer measure on X and $\{A_j\}_{j=1}^\infty$ is a sequence of disjoint μ^*-measurable sets, then

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

for any $E \subset X$.

Solution. By subadditivity, we immediately have

$$\mu^*(E \cap (\cup_{j\in\mathbb{Z}_{\geqslant 1}} A_j)) = \mu^*(\cup_{j\in\mathbb{Z}_{\geqslant 1}} E \cap A_j) \leqslant \sum_{j\in\mathbb{Z}_{\geqslant 1}} \mu^*(E \cap A_j), \quad \text{for any } E \subset X.$$

On the other hand, since $E \cap (\cup_{j=1}^n A_j) \subset E \cap (\cup_{j\in\mathbb{Z}_{\geqslant 1}} A_j)$, it follows that

$$\sum_{j=1}^n \mu^*(E \cap A_j) = \mu(E \cap (\cup_{j=1}^n A_j)) \leqslant \mu(E \cap (\cup_{j\in\mathbb{Z}_{\geqslant 1}} A_j)), \quad \text{for all } n \text{ and any } E \subset X.$$

Therefore, taking the limit $n \to \infty$ yields the inequality, and combining the inequalities proves that

$$\mu^*(E \cap (\cup_{j\in\mathbb{Z}_{\geqslant 1}} A_j)) = \sum_{j\in\mathbb{Z}_{\geqslant 1}} \mu^*(E \cap A_j), \quad \text{for any } E \subset X$$

□

Exercise 1.35: Folland Exercise 1.18.

Let $\mathcal{A} \subset \mathcal{A}(X)$ be an algebra, $\mathcal{A}_\sigma$ the collection of countable unions of sets in $\mathcal{A}$, and $\mathcal{A}_{\sigma\delta}$ the collection of countable intersections of sets in $\mathcal{A}_\sigma$. Let μ_0 be a premeasure on $\mathcal{A}$ and μ^* the induced outer measure.

(a) For any $E \subset X$ and $\varepsilon > 0$, there exists $A \in \mathcal{A}_\sigma$ with $E \subset A$ and $\mu^*(A) \leqslant \mu^*(E)+\varepsilon$.

(b) If $\mu^*(E) < \infty$, then E is μ^*-measurable if and only if there exists $B \in \mathcal{A}_{\sigma\delta}$ with $E \subset B$ and $\mu^*(B \setminus E) = 0$.

(c) If μ_0 is σ-finite, the restriction $\mu^*(E) < \infty$ in (b) is superfluous.

Exercise 1.36: Folland Exercise 1.19.

Let μ^* be an outer measure on X induced from a finite premeasure μ_0. If $E \subset X$, define the inner measure of E to be $\mu_*(E) = \mu_0(X) - \mu^*(E^c)$. Then E is μ^*-measurable if and only if $\mu^*(E) = \mu_*(E)$. (Use Folland Exercise 1.18.)

Exercise 1.37: Folland Exercise 1.20.

Let μ^* be an outer measure on X, $\mathcal{M}^*$ the σ-algebra of μ^*-measurable sets, $\overline{\mu} = \mu^* \mid \mathcal{M}^*$, and μ^+ the outer measure induced by $\overline{\mu}$ as in (1.12) (with $\overline{\mu}$ and $\mathcal{M}^*$ replacing μ_0 and $\mathcal{M}$).

(a) If $E \subset X$, we have $\mu^*(E) \leqslant \mu^+(E)$, with equality if and only if there exists $A \in \mathcal{M}^*$ with $A \supset E$ and $\mu^*(A) = \mu^*(E)$.

(b) If μ^* is induced from a premeasure, then $\mu^* = \mu^+$. (Use Folland Exercise 1.18a.)

(c) If $X = \{0, 1\}$, there exists an outer measure μ^* on X such that $\mu^* \neq \mu^+$.

Exercise 1.38: Folland Exercise 1.21.

Let μ^* be an outer measure induced from a premeasure and $\overline{\mu}$ the restriction of μ^* to the μ^*-measurable sets. Then $\overline{\mu}$ is saturated. (Use Folland Exercise 1.18.)

Exercise 1.39: Folland Exercise 1.22.

Let $(X, \mathcal{M}, \mu)$ be a measure space, μ^* the outer measure induced by μ according to (1.12), $\mathcal{M}^*$ the σ-algebra of μ^*-measurable sets, and $\overline{\mu} = \mu^* \mid \mathcal{M}^*$.

(a) If μ is σ-finite, then $\overline{\mu}$ is the completion of μ. (Use Folland Exercise 1.18.)
(b) In general, $\overline{\mu}$ is the saturation of the completion of μ. (See Exercises 16 and 21.)

Exercise 1.40: Folland Exercise 1.23.

Let $\mathcal{A}$ be the collection of finite unions of sets of the form $(a, b] \cap \mathbb{Q}$ where $-\infty \leqslant a < b \leqslant \infty$.

(a) $\mathcal{A}$ is an algebra on $\mathbb{Q}$. (Use Proposition 7.)
(b) The σ-algebra generated by $\mathcal{A}$ is $\mathcal{A}(\mathbb{Q})$.
(c) Define μ_0 on $\mathcal{A}$ by $\mu_0(\varnothing) = 0$ and $\mu_0(A) = \infty$ for $A \neq \varnothing$. Then μ_0 is a premeasure on $\mathcal{A}$, and there is more than one measure on $\mathcal{A}(\mathbb{Q})$ whose restriction to $\mathcal{A}$ is μ_0.

Exercise 1.41: Folland Exercise 1.24.

Let μ be a finite measure on $(X, \mathcal{M})$, and let μ^* be the outer measure induced by μ. Suppose that $E \subset X$ satisfies $\mu^*(E) = \mu^*(X)$ (but not that $E \in \mathcal{M}$).

(a) If $A, B \in \mathcal{M}$ and $A \cap E = B \cap E$, then $\mu(A) = \mu(B)$.
(b) Let $\mathcal{M}_E = \{A \cap E \mid A \in \mathcal{M}\}$, and define the function ν on $\mathcal{M}_E$ defined by $\nu(A \cap E) = \mu(A)$ (which makes sense by (a)). Then $\mathcal{M}_E$ is a σ-algebra on E and ν is a measure on $\mathcal{M}_E$.

1.5 Borel Measures on the Real Line

We are now in a position to construct a definitive theory for measuring subsets of $\mathbb{R}$ based on the idea that the measure of an interval is its length. We begin with a more general (but only slightly more complicated) construction that yields a large family of measures on $\mathbb{R}$ whose domain is the Borel σ-algebra $\mathcal{B}_{\mathbb{R}}$; such measures are called Borel measures on $\mathbb{R}$.

To motivate the ideas, suppose that μ is a finite Borel measure on $\mathbb{R}$, and let $F(x) = \mu((-\infty, x])$. (F is sometimes called the distribution function of μ.) Then F is increasing by Theorem 14a and right continuous by Theorem 1.8 d since $(-\infty, x] = \bigcap_1^\infty (-\infty, x_n]$ whenever $x_n \searrow x$. (Recall the discussion of increasing functions in §0.5.) Moreover, if $b > a, (-\infty, b] = (-\infty, a] \cup (a, b]$, so $\mu((a, b]) = F(b) - F(a)$. Our procedure will be to turn this process around and construct a measure μ starting from an increasing, right-continuous function F. The special case $F(x) = x$ will yield the usual "length" measure.

The building blocks for our theory will be the left-open, right-closed intervals in $\mathbb{R}$—that is, sets of the form $(a, b]$ or (a, ∞) or $\varnothing$, where $-\infty \leqslant a < b < \infty$. In this section we shall refer to such sets as h-intervals (h for "half-open"). Clearly the intersection of two h-intervals is an h-interval, and the complement of an h-interval is an h-interval or the disjoint union of two h-intervals. By Proposition 7, the collection $\mathcal{A}$ of finite disjoint unions of h-intervals is an algebra, and by Proposition 2, the σ-algebra generated by $\mathcal{A}$ is $\mathcal{A}_{\mathbb{R}}$.

Proposition 1.42: 1.15.

Let $F\colon \mathbb{R} \to \mathbb{R}$ be increasing and right continuous. If $(a_j, b_j]$ $(j = 1, \ldots, n)$ are disjoint h-intervals, let

$$\mu_0\Big(\bigcup\nolimits_1^n (a_j, b_j]\Big) = \sum\nolimits_1^n [F(b_j) - F(a_j)]$$

and let $\mu_0(\varnothing) = 0$. Then μ_0 is a premeasure on the algebra $\mathcal{A}$.

Proof. First we must check that μ_0 is well defined, since elements of $\mathcal{A}$ can be represented in more than one way as disjoint unions of h-intervals. If $\{(a_j, b_j]\}_1^n$ are disjoint and $\bigcup_1^n (a_j, b_j] = (a, b]$, then, after perhaps relabeling the index j, we must have $a = a_1 < b_1 = a_2 < b_2 = \ldots < b_n = b$, so $\sum_1^n [F(b_j) - F(a_j)] = F(b) - F(a)$. More generally, if $\{I_i\}_1^n$ and $\{J_j\}_1^m$ are finite sequences of disjoint h-intervals such that $\bigcup_1^n I_i = \bigcup_1^n J_j$, this reasoning shows that

$$\sum\nolimits_i \mu_0(I_i) = \sum\nolimits_{i,j} \mu_0(I_i \cap J_j) = \sum\nolimits_j \mu_0(J_j)$$

Thus μ_0 is well defined, and it is finitely additive by construction.

It remains to show that if $\{I_j\}_1^\infty$ is a sequence of disjoint h-intervals with $\bigcup_1^\infty I_j \in \mathcal{A}$ then $\mu_0(\bigcup_1^\infty I_j) = \sum_1^\infty \mu_0(I_j)$. Since $\bigcup_1^\infty I_j$ is a finite union of h-intervals, the sequence $\{I_j\}_1^\infty$ can be partitioned into finitely many subsequences such that the union of the intervals in each subsequence is a single h-interval. By considering each subsequence separately and using the finite additivity of μ_0, we may assume that $\bigcup_1^\infty I_j$ is an h-interval $I = (a, b]$. In this case, we have

$$\mu_0(I) = \mu_0\Big(\bigcup\nolimits_1^n I_j\Big) + \mu_0\Big(I \setminus \bigcup\nolimits_1^n I_j\Big) \geqslant \mu_0\Big(\bigcup\nolimits_1^n I_j\Big) = \sum\nolimits_1^n \mu_0(I_j)$$

Letting $n \to \infty$, we obtain $\mu_0(I) \geqslant \sum_1^\infty \mu(I_j)$. To prove the reverse inequality, let us suppose first that a and b are finite, and let us fix $\varepsilon > 0$. Since F is right continuous, there exists $\delta > 0$ such that $F(a + \delta) - F(a) < \varepsilon$, and if $I_j = (a_j, b_j$, for each j there exists $\delta_j > 0$ such that $F(b_j + \delta_j) - F(b_j) < \varepsilon 2^{-j}$. The open intervals $(a_j, b_j + \delta_j)$ cover the compact set $[a + \delta, b]$, so there is a finite subcover. By discarding any $(a_j, b_j + \delta_j)$

that is contained in a larger one and relabeling the index j, we may assume that - the intervals $(a_1, b_1 + \delta_1), \dots, (a_N, b_N + \delta_N)$ cover$[a+\delta, b]$, - $b_j + \delta_j \in (a_{j+1}, b_{j+1} + \delta_{j+1})$ for $j = 1, \dots, N-1$.

But then

$$\begin{aligned}
\mu_0(I) &< F(b) - F(a+\delta) + \varepsilon \\
&\leqslant F(b_N + \delta_N) - F(a_1) + \varepsilon \\
&= F(b_N + \delta_N) - F(a_N) + \sum_1^{N-1} [F(a_{j+1}) - F(a_j)] + \varepsilon \\
&\leqslant F(b_N + \delta_N) - F(a_N) + \sum_1^{N-1} [F(b_j + \delta_j) - F(a_j)] + \varepsilon \\
&< \sum_1^{N} [F(b_j) + \varepsilon 2^{-j} - F(a_j)] + \varepsilon \\
&< \sum_1^{\infty} \mu(I_j) + 2\varepsilon
\end{aligned}$$

Since ε is arbitrary, we are done when a and b are finite. If $a = -\infty$, for any $M < \infty$ the intervals $(a_j b_j + \delta_j)$ cover $[-M, b]$, so the same reasoning gives $F(b) - F(-M) \leqslant \sum_1^\infty \mu_0(I_j) + 2\varepsilon$, whereas if $b = \infty$, for any $M < \infty$ we likewise obtain $F(M) - F(a) \leqslant \sum_1^\infty \mu_0(I_j) + 2\varepsilon$. The desired result then follows by letting $\varepsilon \to 0$ and $M \to \infty$. ☐

Theorem 1.43: 1.16.

If $F\colon \mathbb{R} \to \mathbb{R}$ is any increasing, right continuous function, 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. If G is another such function, we have $\mu_F = \mu_G$ if and only if $F - G$ is constant. Conversely, if μ is a Borel measure on $\mathbb{R}$ that 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 and right continuous, and $\mu = \mu_F$.

Proof. Each F induces a premeasure on $\mathcal{A}$ by Proposition 42. It is clear that F and G induce the same premeasure if and only if $F - G$ is constant, and that these premeasures are σ-finite (since $\mathbb{R} = \bigcup_{-\infty}^{\infty} (j, j+1]$). The first two assertions therefore follow from Theorem 33. As for the last one, the monotonicity of μ implies the monotonicity of F, and the continuity of μ from above and below implies the right continuity of F for $x \geqslant 0$ and $x < 0$. It is evident that $\mu = \mu_F$ on $\mathcal{A}$, and hence $\mu = \mu_F$ on $\mathcal{A}_{\mathbb{R}}$ by the uniqueness in Theorem 33. ☐

Several remarks are in order. First, this theory could equally well be developed by using intervals of the form $[a, b)$ and left continuous functions F. Second, if μ is a finite Borel measure on $\mathbb{R}$, then $\mu = \mu_F$ where $F(x) = \mu((-\infty, x])$ is the cumulative distribution function of μ; this differs from the F specified in Theorem 43 by the constant $\mu((-\infty, 0])$.

Third, the theory of Folland Section 1.4 gives, for each increasing and right continuous F, not only the Borel measure μ_F but a complete measure $\overline{\mu}_F$ whose domain includes $\mathcal{B}_{\mathbb{R}}$. In fact, $\overline{\mu}_F$ is just the completion of μ_F (Folland Exercise 1.22a or Theorem 46 below), and one can show that its domain is always strictly larger than $\mathcal{B}_{\mathbb{R}}$. We shall usually denote this complete measure also by μ_F; it is called the Lebesgue-Stieltjes measure associated to F.

Lebesgue-Stieltjes measures enjoy some useful regularity properties that we now investigate. In this discussion we fix a complete Lebesgue-Stieltjes measure μ on $\mathbb{R}$ associated to the increasing, right continuous function F, and we denote by $\mathcal{M}_\mu$ the domain of μ. Thus, for any $E \in \mathcal{M}_\mu$,

$$\begin{aligned}\mu(E) &= \inf\Big\{\sum_1^\infty [F(b_j) - F(a_j)] \,\Big|\, E \subset \bigcup_1^\infty (a_j, b_j]\Big\} \\ &= \inf\Big\{\sum_1^\infty \mu((a_j, b_j]) \,\Big|\, E \subset \bigcup_1^\infty (a_j, b_j]\Big\}.\end{aligned}$$

We first observe that in the second formula for $\mu(E)$ we can replace h-intervals by open h-intervals:

Lemma 1.44: 1.17.

For any $E \in \mathcal{M}_\mu$,

$$\mu(E) = \inf\Big\{\sum_1^\infty \mu((a_j, b_j)) \,\Big|\, E \subset \bigcup_1^\infty (a_j, b_j)\Big\}$$

Proof. Let us call the quantity on the right $\nu(E)$. Suppose $E \subset \bigcup_1^\infty (a_j, b_j)$. Each (a_j, b_j) is a countable disjoint union of h-intervals $I_j^k (k = 1, 2, \ldots)$; specifically, $I_j^k = (c_j^k, c_j^{k+1}]$ where $\{c_j\}$ is any sequence such that $c_j^1 = a_j$ and c_j^k increases to b_j as $k \to \infty$. Thus $E \subset \bigcup_{j,k=1}^\infty I_j^k$, so

$$\sum_1^\infty \mu((a_j, b_j)) = \sum_{j,k=1}^\infty \mu(I_j^k) \geqslant \mu(E)$$

and hence $\nu(E) \geqslant \mu(E)$. On the other hand, given $\varepsilon > 0$ there exists $\{(a_j, b_j]\}_1^\infty$ with $E \subset \bigcup_1^\infty (a_j, b_j]$ and $\sum_1^\infty \mu((a_j, b_j]) \leqslant \mu(E) + \varepsilon$, and for each j there exists $\delta_j > 0$ such that $F(b_j + \delta_j) - F(b_j) < \varepsilon 2^{-j}$. Then $E \subset \bigcup_1^\infty (a_j, b_j + \delta_j)$ and

$$\sum_1^\infty \mu((a_j, b_j + \delta_j)) \leqslant \sum_1^\infty \mu((a_j, b_j]) + \varepsilon \leqslant \mu(E) + 2\varepsilon$$

so that $\nu(E) \leqslant \mu(E)$. □

Theorem 1.45: 1.18.

If $E \in \mathcal{M}_\mu$, then

$$\begin{aligned}\mu(E) &= \inf\{\mu(U) \mid U \supset E \text{ and } U \text{ is open }\} \\ &= \sup\{\mu(K) \mid K \subset E \text{ and } K \text{ is compact }\}.\end{aligned}$$

Proof. By Lemma 44, for any $\varepsilon > 0$ there exist intervals (a_j, b_j) such that $E \subset \bigcup_1^\infty (a_j, b_j)$ and $\mu(E) \leqslant \sum_1^\infty \mu((a_j, b_j)) + \varepsilon$. If $U = \bigcup_1^\infty (a_j, b_j)$ then U is open, $U \supset E$, and $\mu(U) \leqslant \mu(E) + \varepsilon$. On the other hand, $\mu(U) \geqslant \mu(E)$ whenever $U \supset E$, so the first equality is valid. For the second one, suppose first that E is bounded. If E is closed, then E is compact and the equality is obvious. Otherwise, given $\varepsilon > 0$ we can choose an open $U \supset \overline{E} \setminus E$ such that $\mu(U) \leqslant \mu(\overline{E} \setminus E) + \varepsilon$. Let $K = \overline{E} \setminus U$. Then K is compact, $K \subset E$, and

$$\begin{aligned} \mu(K) &= \mu(E) - \mu(E \cap U) = \mu(E) - [\mu(U) - \mu(U \setminus E)] \\ &\geqslant \mu(E) - \mu(U) + \mu(\overline{E} \setminus E) \geqslant \mu(E) - \varepsilon \end{aligned}$$

If E is unbounded, let $E_j = E \cap (j, j+1]$. By the preceding argument, for any $\varepsilon > 0$ there exist compact $K_j \subset E_j$ with $\mu(K_j) \geqslant \mu(E_j) - \varepsilon 2^{-j}$. Let $H_n = \bigcup_{-n}^n K_j$. Then H_n is compact, $H_n \subset E$, and $\mu(H_n) \geqslant \mu(\bigcup_{-n}^n E_j) - \varepsilon$. Since $\mu(E) = \lim_{n\to\infty} \mu(\bigcup_{-n}^n E_j)$, the result follows. □

Theorem 1.46: 1.19.

If $E \subset \mathbb{R}$, the following are equivalent.

(a) $E \in \mathcal{M}_\mu$.
(b) $E = V \setminus N_1$ where V is aG_δ set and $\mu(N_1) = 0$.
(c) $E = H \cup N_2$ where H is an F_σ set and $\mu(N_2) = 0$.

Proof. Obviously (b) and (c) each imply (a) since μ is complete on $\mathcal{M}_\mu$. Suppose $E \in \mathcal{M}_\mu$ and $\mu(E) < \infty$. By Theorem 45, for $j \in \mathbb{Z}_{\geqslant 1}$ we can choose an open $U_j \supset E$ and a compact $K_j \subset E$ such that

$$\mu(U_j) - 2^{-j} \leqslant \mu(E) \leqslant \mu(K_j) + 2^{-j}.$$

Let $V = \bigcap_1^\infty U_j$ and $H = \bigcup_1^\infty K_j$. Then $H \subset E \subset V$ and $\mu(V) = \mu(H) = \mu(E) < \infty$, so $\mu(V \setminus E) = \mu(E \setminus H) = 0$. The result is thus proved when $\mu(E) < \infty$; the extension to the general case is left to the reader (Folland Exercise 1.25). □

The significance of Theorem 46 is that all Borel sets (or, more generally, all sets in $\mathcal{M}_\mu$) are of a reasonably simple form modulo sets of measure zero. This contrasts markedly with the machinations necessary to construct the Borel sets from the open sets when null sets are not excepted; see Proposition 60 below. Another version of the idea that general measurable sets can be approximated by "simple" sets is contained in the following proposition.

Proposition 1.47: 1.20.

If $E \in \mathcal{M}_\mu$ and $\mu(E) < \infty$, then for every $\varepsilon > 0$ there is a set A that is a finite union of open intervals such that $\mu(E \triangle A) < \varepsilon$.

Proof. See Folland Exercise 1.26. □

We now examine the most important measure on $\mathbb{R}$, namely, Lebesgue measure: This is the complete measure μ_F associated to the function $F(x) = x$, for which the measure of an interval is simply its length. We shall denote it by m. The domain of m is called the class of Lebesgue measurable sets, and we shall denote it by $\mathcal{L}$. We shall also refer to the restriction of m to $\mathcal{L}_{\mathbb{R}}$ as Lebesgue measure.

Among the most significant properties of Lebesgue measure are its invariance under translations and simple behavior under dilations. If $E \subset \mathbb{R}$ and $s, r \in \mathbb{R}$, we define

$$E + s = \{x + s \mid x \in E\}, \quad rE = \{rx \mid x \in E\}.$$

Theorem 1.48: 1.21.

If $E \in \mathcal{L}$, then $E + s \in \mathcal{L}$ and $rE \in \mathcal{L}$ for all $s, r \in \mathbb{R}$. Moreover, $m(E + s) = m(E)$ and $m(rE) = |r|m(E)$.

Proof. Since the collection of open intervals is invariant under translations and dilations, the same is true of $\mathcal{B}_{\mathbb{R}}$. For $E \in \mathcal{B}_{\mathbb{R}}$, let $m_s(E) = m(E + s)$ and $m^r(E) = m(rE)$. Then m_s and m^r clearly agree with m and $|r|m$ on finite unions of intervals, hence on $\mathcal{B}_{\mathbb{R}}$ by Theorem 33. In particular, if $E \in \mathcal{B}_{\mathbb{R}}$ and $m(E) = 0$, then $m(E + s) = m(rE) = 0$, from which it follows that the class of sets of Lebesgue measure zero is preserved by translations and dilations. It follows that $\mathcal{B}$ (the members of which are a union of a Borel set and a Lebesgue null set) is preserved by translation and dilations and that $m(E + s) = m(E)$ and $m(rE) = |r|m(E)$ for all $E \in \mathcal{B}$. □

The relation between the measure-theoretic and topological properties of subsets of $\mathbb{R}$ is delicate and contains some surprises. Consider the following facts. Every singleton set in $\mathbb{R}$ has Lebesgue measure zero, and hence so does every countable set. In particular, $m(\mathbb{Q}) = 0$. Let $\{r_j\}_1^\infty$ be an enumeration of the rational numbers in $[0, 1]$, and given $\varepsilon > 0$, let I_j be the interval centered at r_j of length $\varepsilon 2^{-j}$. Then the set $U = (0, 1) \cap \bigcup_1^\infty I_j$ is open and dense in $[0, 1]$, but $m(U) \leqslant \sum_1^\infty \varepsilon 2^{-j} = \varepsilon$; its complement $K = [0, 1] \setminus U$ is closed and nowhere dense, but $m(K) \geqslant 1 - \varepsilon$. Thus a set that is open and dense, and hence topologically "large," can be measuretheoretically small, and a set that is nowhere dense, and hence topologically "small," can be measure-theoretically large. (A nonempty open set cannot have Lebesgue measure zero, however.)

The Lebesgue null sets include not only all countable sets but many sets having the cardinality of the continuum. We now present the standard example, the Cantor set, which is also of interest for other reasons.

The Lebesgue null sets include not only all countable sets but many sets having the cardinality of the continuum. We now present the standard example, the Cantor set, which is also of interest for other reasons.

Each $x \in [0,1]$ has a base-3 decimal expansion $x = \sum_1^\infty a_j 3^{-j}$ where $a_j = 0, 1$, or 2. This expansion is unique unless x is of the form $p3^{-k}$ for some integers p, k, in which case x has two expansions: one with $a_j = 0$ for $j > k$ and one with $a_j = 2$ for $j > k$. Assuming p is not divisible by 3, one of these expansions will have $a_k = 1$ and the other will have $a_k = 0$ or 2. If we agree always to use the latter expansion, we see that

$$a_1 = 1 \text{ if and only if } \frac{1}{3} < x < \frac{2}{3},$$
$$a_1 \neq 1 \text{ and } a_2 = 1 \text{ if and only if } \frac{1}{9} < x < \frac{2}{9} \text{ or } \frac{7}{9} < x < \frac{8}{9},$$

and so forth. It will also be useful to observe that if $x = \sum a_j 3^{-j}$ and $y = \sum b_j 3^{-j}$, then $x < y$ if and only if there exists an n such that $a_n = b_n$ and $a_j = b_j$ for $j < n$.

The ***Cantor set*** C is the set of all $x \in [0,1]$ that have a base-3 expansion $x = \sum a_j 3^{-j}$ with $a_j \neq 1$ for all j. Thus C is obtained from $[0,1]$ by removing the open middle third $\left(\frac{1}{3}, \frac{2}{3}\right)$, then removing the open middle thirds $\left(\frac{1}{9}, \frac{2}{9}\right)$ and $\left(\frac{7}{9}, \frac{8}{9}\right)$ of the two remaining intervals, and so forth. The basic properties of C are summarized as follows:

Proposition 1.49: 1.22.

Let C be the Cantor set.

(a) C is compact, nowhere dense, and totally disconnected (i.e., the only connected subsets of C are single points). Moreover, C has no isolated points.

(b) $m(C) = 0$.

(c) $\text{card}(C) = \mathfrak{c}$.

Proof. We leave the proof of (a) to the reader (Folland Exercise 1.27). As for (b), C is obtained from $[0,1]$ by removing one interval of length $\frac{1}{3}$, two intervals of length $\frac{1}{9}$, and so forth. Thus

$$m(C) = 1 - \sum_0^\infty \frac{2^j}{3^{j+1}} = 1 - \frac{1}{3} \cdot \frac{1}{1-(2/3)} = 0.$$

Lastly, suppose $x \in C$, so that $x = \sum_0^\infty a_j 3^{-j}$ where $a_j = 0$ or 2 for all j. Let $f(x) = \sum_1^\infty b_j 2^{-j}$ where $b_j = a_j/2$. The series defining $f(x)$ is the base-2 expansion of a number in $[0,1]$, and any number in $[0,1]$ can be obtained in this way. Hence f maps C onto $[0,1]$, and (c) follows. $\square$

Let us examine the map f in the preceding proof more closely. One readily sees that if $x, y \in C$ and $x < y$, then $f(x) < f(y)$ unless x and y are the two endpoints of one of the intervals removed from $[0,1]$ to obtain C. In this case $f(x) = p2^{-k}$ for some integers p, k, and $f(x)$ and $f(y)$ are the two base-2 expansions of this number. We can therefore extend f to a map from $[0,1]$ to itself by declaring it to be constant on each interval missing from C. This extended f is still increasing, and since its range is all of $[0,1]$ it cannot have any jump discontinuities; hence it is continuous. f is called the Cantor function or ***Cantor-Lebesgue function***.

The construction of the Cantor set by starting with $[0, 1]$ and successively removing open middle thirds of intervals has an obvious generalization. If I is a bounded interval and $\alpha \in (0, 1)$, let us call the open interval with the same midpoint as I and length equal to α times the length of I the "open middle αth" of I. If $\{\alpha_j\}_1^\infty$ is any sequence of numbers in $(0, 1)$, then, we can define a decreasing sequence $\{K_j\}$ of closed sets as follows: $K_0 = [0, 1]$, and K_j is obtained by removing the open middle α_jth from each of the intervals that make up K_{j-1}. The resulting limiting set $K = \bigcap_1^\infty K_j$ is called a generalized Cantor set. Generalized Cantor sets all share with the ordinary Cantor set the properties (a) and (c) in Proposition 49. As for their Lebesgue measure, clearly $m(K_j) = (1 - \alpha_j)m(K_{j-1})$, so $m(K)$ is the infinite product $\prod_1^\infty (1 - \alpha_j) = \lim_{n\to\infty} \prod_1^n (1 - \alpha_j)$. If the α_j are all equal to a fixed $\alpha \in (0, 1)$ (for example, $\alpha = \frac{1}{3}$ for the ordinary Cantor set), we have $m(K) = 0$. However, if $\alpha_j \to 0$ sufficiently rapidly as $j \to \infty, m(K)$ will be positive, and for any $\beta \in (0, 1)$ one can choose α_j so that $m(K)$ will equal β; see Folland Exercise 1.32. This gives another way of constructing nowhere dense sets of positive measure.

Not every Lebesgue measurable set is a Borel set. One can display examples of sets in $\mathcal{L} \setminus \mathcal{L}_\mathbb{R}$ by using the Cantor function; see Folland Exercise 2.9.

Exercise 1.50: Folland Exercise 1.25.

Complete the proof of Theorem 46

Exercise 1.51: Folland Exercise 1.26.

Prove Proposition 47. (Use Theorem 45.)

Exercise 1.52: Folland Exercise 1.27.

Prove Proposition 49(a). (Show that if $x, y \in C$ and $x < y$, there exists $z \notin C$ such that $x < z < y$.)

Exercise 1.53: Folland Exercise 1.28.

Let F be increasing and right continuous, and let μ_F be the associated measure. Then $\mu_F(\{a\}) = F(a) - F(a-), \mu_F([a, b)) = F(b-) - F(a-), \mu_F([a, b]) = F(b) - F(a-)$, and $\mu_F((a, b)) = F(b-) - F(a)$.

Exercise 1.54: Folland Exercise 1.29 (Variant).

Let μ^* be an outer measure on X. If V is a μ^*-measurable subset of X and E is a μ^*-measurable subset of X contained in V, then

$$\mu^*(E \setminus V) > 0.$$

Proof. If $\mu^*(E\setminus V) = 0$, then $E\setminus V$ is μ^*-measurable. so $E\cap(E\setminus V)^c = E\setminus(V\setminus V) = V$, a contradiction. Thus $\mu^*(V \setminus A) > 0$. □

Remark 55. *Since every null set is μ^*-measurable and μ^* restricted to the σ-algebra of μ^*-measurable sets is a complete measure, it is clear that $\mu^*(E) > 0$ if E contains a μ^*-nonmeasurable set. Indeed, if V is a nonmeasurable set in E but $\mu^*(E) = 0$, then since any subsets of null sets are measurable when the measure is complete, this would mean V would be measurable, a contradiction. Thus $\mu^*(E) > 0$.*

Exercise 1.56: Folland Exercise 1.30.

If $E \in \mathcal{L}$ and $m(E) > 0$, for any $\alpha < 1$ there is an open interval I such that $m(E \cap I) > \alpha m(I)$.

Exercise 1.57: Folland Exercise 1.31.

If $E \in \mathcal{L}$ and $m(E) > 0$, the set $E - E = \{x - y \mid x, y \in E\}$ contains an interval centered at 0. (If I is as in Folland Exercise 1.30 with $\alpha > \frac{3}{4}$, then $E - E$ contains $\left(-\frac{1}{2}m(I), \frac{1}{2}m(I)\right)$.)

Exercise 1.58: Folland Exercise 1.32.

Suppose $\{\alpha_j\}_1^\infty \subset (0, 1)$.

(a) $\prod_1^\infty(1 - \alpha_j) > 0$ if and only if $\sum_1^\infty \alpha_j < \infty$. (Compare $\sum_1^\infty \log(1 - \alpha_j)$ to $\sum \alpha_j$.)
(b) Given $\beta \in (0, 1)$, exhibit a sequence $\{\alpha_j\}$ such that $\prod_1^\infty(1 - \alpha_j) = \beta$.

Exercise 1.59: Folland Exercise 1.33.

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

[a] *Hint: Every subinterval of $[0, 1]$ contains Cantor-type sets of positive measure.*

Proposition 1.60: 1.23.

$\mathcal{M}(E) = \bigcup_{\alpha\in\Omega} \mathcal{M}_\alpha$, where Ω is the set of countable ordinals.

Proof. Transfinite induction shows that $\mathcal{E}_\alpha \subset \mathcal{E}(E)$ for all $\alpha \in \Omega$, and hence $\bigcup_{\alpha\in\Omega} \mathcal{E}_\alpha \subset \mathcal{E}(E)$. The reverse inclusion follows from the fact that any sequence in Ω has a supremum in Ω (see Folland Proposition 20): If $E_j \in \mathcal{E}_{\alpha_j}$ for $j \in \mathbb{Z}_{\geqslant 1}$ and $\beta = \sup\{\alpha_j\}$, then $E_j \in \mathcal{E}_\alpha$ for all j and hence $\bigcup_1^\infty E_j \in \mathcal{E}_\beta$ where β is the successor of α. □

Combining Proposition 60 with Folland Proposition 15, we see that if $\text{card}(\mathbb{Z}_{\geqslant 1}) \leqslant \text{card}(\mathcal{E}) \leqslant \mathfrak{c}$, then $\text{card}(\mathcal{E}(\mathcal{E})) = \mathfrak{c}$. (See Folland Exercise 1.3.)

2 Integration

2.1 Measurable Functions

We first study the category of measurable spaces, whose morphisms are measurable mappings. Recall that any set map $f\colon X \to Y$ induces a mapping $f^{-1}\colon \mathcal{P}(Y) \to \mathcal{P}(X)$, given by the preimage of f, that preserves unions, intersections, and complements. (Check!) Thus, if $\mathcal{N}$ is a σ-algebra on Y, then $\mathcal{M} := f^{-1}(\mathcal{N})$ is a σ-algebra on X.

Definition 1. *A morphism of measurable spaces is called a* ***measurable function****. That is, given measurable spaces $(X, \mathcal{M}), (Y, \mathcal{N})$ and a set map $f\colon X \to Y$, f is called* ***$(\mathcal{M}, \mathcal{N})$-measurable****, or simply* ***measurable*** *when $\mathcal{M}$ and $\mathcal{N}$ are understood, if $f^{-1}(E) \in \mathcal{M}$ whenever $E \in \mathcal{N}$. One can check that this in fact makes the collection of measurable spaces with measurable mappings into a category.*

Proposition 2.2: 2.1.

If $\mathcal{N} = \mathcal{M}(\mathcal{E})$, then $f\colon X \to Y$ is $(\mathcal{M}, \mathcal{N})$-measurable if and only if $f^{-1}(E) \in \mathcal{M}$ for all $E \in \mathcal{E}$.

Proof. The forward implication is trivial. Conversely, if $\{E \subset Y \mid f^{-1}(E) \in \mathcal{M}\}$ is a σ-algebra containing $\mathcal{E}$, then it contains $\mathcal{M}$. □

Corollary 2.3: 2.2.

If X and Y are topological spaces, then every continuous $f\colon X \to Y$ is $(\mathcal{B}_X, \mathcal{B}_Y)$-measurable.

Proof. This is almost trivial, and becomes so when noting f is continuous if and only if $f^{-1}(U)$ is open for all open subsets U of Y. □

If $(X, \mathcal{M})$ is a measurable space, a real- or complex-valued function f on X is called ***$\mathcal{M}$-measurable***, or simply ***measurable*** when $\mathcal{M}$ is understood, if f is $(\mathcal{M}, \mathcal{B}_{\mathbb{R}})$- or $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$-measurable.

$\mathcal{B}_{\mathbb{R}}$ or $\mathcal{B}_{\mathbb{C}}$ are *always* understood as the σ-algebra on the codomain unless otherwise specified. In particular, $f: \mathbb{R} \to \mathbb{C}$ is ***Lebesgue measurable*** (resp. ***Borel measurable***) if it is $(\mathcal{L}, \mathcal{B}_{\mathbb{C}})$- (resp.$(\mathcal{B}_{\mathbb{R}}, \mathcal{B}_{\mathbb{C}})$-)measurable; likewise for $f: \mathbb{R} \to \mathbb{R}$.

Warning 2.4.

If f, g are Lebesgue measurable, it is *not* necessarily the case that $f \circ g$ is Lebesgue measurable, even if g is assumed to be continuous.

Proposition 2.5: 2.3.

If $(X, \mathcal{M})$ is a measurable space and $f: X \to \mathbb{R}$, the following are equivalent:

- f is $\mathcal{M}$-measurable.
- $\{f > a\} := f^{-1}((a, \infty)) \in \mathcal{M}$ for all $a \in \mathbb{R}$.
- $\{f \geqslant a\} := f^{-1}([a, \infty)) \in \mathcal{M}$ for all $a \in \mathbb{R}$.
- $\{f < a\} := f^{-1}((-\infty, a)) \in \mathcal{M}$ for all $a \in \mathbb{R}$.
- $\{f \leqslant a\} := f^{-1}((-\infty, a]) \in \mathcal{M}$ for all $a \in \mathbb{R}$.

Proof. This is an immediate consequence of Proposition 2. □

Sometimes we wish to consider measurability on subsets of X. If $(X, \mathcal{M})$ is a measurable space, f is a function on X, and $E \in \mathcal{M}$, we say that f is ***measurable on E*** if $f^{-1}(B) \cap E \in \mathcal{M}$ for all Borel sets B. (Equivalently, $f|_E$ is $\mathcal{M}_E$-measurable, where $\mathcal{M}_E = \{F \cap E \mid F \in \mathcal{M}\}$.)

Given a set X, if $\{(Y_\alpha, \mathcal{N}_\alpha)\}_{\alpha \in A}$ is a family of measurable spaces, and $f: X \to Y_\alpha$ is a map for each $\alpha \in A$, there is a unique smallest σ-algebra on X with respect to which the f_αs are all measurable, namely, the σ-algebra generated by the sets $f_\alpha^{-1}(E_\alpha)$ with $E_\alpha \in \mathcal{N}_\alpha$ and $\alpha \in A$. It is called the ***σ-algebra generated by $\{f_\alpha\}_{\alpha \in A}$***. In particular, if $X = \prod_{\alpha \in A} Y_\alpha$, we see that the product σ-algebra on X is the σ-algebra generated by the coordinate maps $\pi_\alpha^{-1}: X \to Y_\alpha$.

Proposition 2.6: 2.4.

Let $(X, \mathcal{M})$ and $(Y_\alpha, \mathcal{N}_\alpha)(\alpha \in A)$ be measurable spaces, $Y = \prod_{\alpha \in A} Y_\alpha$, $\mathcal{N} = \bigotimes_{\alpha \in A} \mathcal{N}_\alpha$, and $\pi_\alpha: Y \to Y_\alpha$ the coordinate maps. Then $f: X \to Y$ is $(\mathcal{M}, \mathcal{N})$-measurable if and only if $f_\alpha = \pi_\alpha \circ f$ is $(\mathcal{M}, \mathcal{N}_\alpha)$-measurable for all α.

Proof. If f is measurable, so is each f_α since the composition of measurable maps is measurable. Conversely, if each f_α is measurable, then for all $E_\alpha \in \mathcal{N}_\alpha$, $f^{-1}(\pi_\alpha^{-1}(E_\alpha)) = f_\alpha^{-1}(E_\alpha) \in \mathcal{M}$, for which f is measurable by Proposition 2. □

Corollary 2.7: 2.5.

A function $f\colon X \to \mathbb{C}$ is $\mathcal{M}$-measurable if and only if $\operatorname{Re} f$ and $\operatorname{Im} f$ are $\mathcal{M}$-measurable.

Proof. This follows from Proposition 6 since $\mathcal{B}_{\mathbb{C}} = \mathcal{B}_{\mathbb{R}^2} = \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}}$ by Proposition 5. ☐

It is sometimes convenient to consider functions with values in the extended real number system $\overline{\mathbb{R}} = [\infty, \infty]$. We define Borel sets in $\overline{\mathbb{R}}$ by $\mathcal{B}_{\overline{\mathbb{R}}} = \{E \subset \overline{\mathbb{R}} \mid E \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}\}$. (This coincides with the usual definition of the Borel σ-algebra if we make $\overline{\mathbb{R}}$ into a metric space with metric $\rho(x, y) = |A(x) - A(y)|$, where $A(x) = \arctan x$.) It is easily verified as in Proposition 5 that $\mathcal{B}_{\overline{\mathbb{R}}}$ is generated by the rays $(a, \infty]$ or $[-\infty, a)(a \in \mathbb{R})$, and we define $f\colon X \to \overline{\mathbb{R}}$ to be $\mathcal{M}$-measurable if it is $(\mathcal{M}, \mathcal{B}_{\overline{\mathbb{R}}})$-measurable.

We now establish that measurability is preserved under the familiar algebraic and limiting operations.

Proposition 2.8: Extended Version of 2.6.

Let $(X, \mathcal{M})$ be a measurable space and suppose f, g, and f_k are $\mathcal{M}$-measurable for all k.

(1) The sets $\{f < g\}$, $\{f \leqslant g\}$, and $\{f = g\}$ are in $\mathcal{M}$ for all k,

(2) The restriction of f to any $E \in \mathcal{M}$ is $\mathcal{M}$-measurable,

(3) $f + g$ is $\mathcal{M}$-measurable,

(4) λf for any constant λ is measurable,

(5) fg is $\mathcal{M}$-measurable, and

(6) $\sup_k f_k$, $\inf_k f_k$, $\limsup_k f_k$, $\liminf_k f_k$ are all $\mathcal{M}$-measurable.

Proof. For (1), write

$$\{f < g\} = \bigcup_{r \in \mathbb{Q}} \Big(\underbrace{\{f < r\}}_{\in \mathcal{M} \text{ by } (5)} \cap \overbrace{\{r < g\}}^{\in \mathcal{M} \text{ by } (5)} \Big) \in \mathcal{M},$$

so $\{f \leqslant g\} = \{g < f\}^c$ and $\{f = g\} = \{f \leqslant g\} \cap \{g \leqslant f\}$ are also in $\mathcal{M}$. For (2), note that $\{f|_E < a\} = \{f < a\} \cap E$ for any $a \in \mathbb{R}$. For (3), write $\{f + g > a\} = \bigcup_{r \in \mathbb{Q}} \{f > r\} \cap \{g > a - r\}$. To see (4), first note if $c = 0$ then $cf = 0$ is $\mathcal{M}$-measurable because constant functions are measurable. If $c > 0$ then $\{cf > a\} = \{f > a/c\}$. If $c < 0$ then $\{f < a/c\}$. (5) follows from the fact $fg = (f + g)^2 - (f - g)^2$ together with the previous points. To see (6), write $\{\sup_k f_k > a\} = \bigcup_{k=1}^{\infty}\{f_k > a\}$, and $\inf_k f_k = -\sup(-f_k)$, $\limsup_{k\to\infty} = \inf_k \sup_{n\geqslant k} f_n = \inf_k \sup_n f_{n+k}$, and similarly for $\liminf_{k\to\infty} f_k$. □

Remark 9. *Proposition 8 holds for $\overline{\mathbb{R}}$-valued functions. (Check!)*

To prove the following two corollaries, apply Corollary 7.

Corollary 2.10: 2.8.

If $f, g\colon X \to \mathbb{R}$ is measurable, then so are $\max\{f, g\}$ and $\min\{f, g\}$.

Corollary 2.11: 2.9.

If $\{f_n\colon X \to \mathbb{C}\}_{n=1}^{\infty}$ is a sequence of measurable functions and $f_n \to f$ pointwise, then f is measurable.

By Corollary 11, measurability is closed under pointwise limits. We will soon define an integral on the set of nonnegative measurable functions, which will mean that if we have a pointwise limit of integrable functions, then the limit is integrable. This is something that we didn't necessarily have before measures, as shown in the following example.

2.2 Constructing Measurable Functions from Simple Functions

We now build functions from building blocks. Let $(X, \mathcal{M})$ be a measurable space.

Definition 12. *Given $E \subset X$, define a map $\chi_E\colon X \to \mathbb{R}$, called the* ***characteristic function*** *of E, by*

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

The ***standard representation*** *of a simple function f is*

$$f = \sum_{j=1}^{n} z_j \chi_{E_j},$$

where $E_j = f^{-1}(\{z_j\})$, $\operatorname{im} f = \{z_1, \ldots, z_n\}$. Note that the E_j are disjoint sets. We can write it this way because any simple function is a finite linear combination of characteristic functions. Note that we allow $z_j = 0$, and indeed these sets may play an important role when multiplying or composing functions.

Remark 13. *Note that χ_E is measurable if and only if $E \in \mathcal{M}$. (Check!)*

Remark 14. *It is useful to know how to construct or disassemble characteristic functions of given measurable sets. To that end, here are some useful characteristic function identities. Let $E, F \in \mathcal{M}$.*

$$\chi_{E^c} = 1 - \chi_E,$$

$$\begin{aligned} \chi_{E\cup F} &= \chi_E + \chi_F - \chi_E\chi_F, \\ \chi_{E\cap F} &= \chi_E\chi_F, \\ \chi_{E\smallsetminus F} &= \chi_E(1-\chi_F), \\ \chi_{E\triangle F} &= \chi_E + \chi_F - 2\chi_E\chi_F, \\ \mu(E\triangle F) &= \int |\chi_E - \chi_F|\, d\mu. \end{aligned}$$

Definition 15. *A* ***simple function*** $f\colon X \to \mathbb{C}$ *is a finite linear combination of characteristic functions.*

Remark 16. *It is useful to note that* $f\colon X\to\mathbb{C}$ *is simple if and only if* f *is measurable and* $\operatorname{im} f$ *is a finite set.*

Notation 17. *Given any sequence* $\{f_n\colon X \to \mathbb{R}\}$ *of set functions and a set function* $f\colon X\to\mathbb{R}$*, we will write* $f_n \nearrow f$ *to mean*

$$f_1 \leqslant f_2 \leqslant \cdots \leqslant f$$

and $f_n \to f$ *pointwise.*

Theorem 2.18: 2.10.

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

(a) If $f\colon X \to [0,\infty]$ is measurable, then there exist simple functions $\{\phi_n\}_{n=1}^{\infty}$ such that $\phi_n \nearrow f$, and this convergence is uniform on any set where f is bounded.

(b) If $f\colon X\to\mathbb{C}$ is measurable, then there exists a sequence $\{\phi\}_{n=1}^{\infty}$ of simple functions such that $|\phi_n| \nearrow |f|$ and $\phi_n \to f$ pointwise, and the latter convergence is uniform on any set where f is bounded.

Proof. This is a constructive proof. We will prove (a) since (b) will then follow. Suppose $f\colon X\to[0,\infty]$ is measurable. Let

$$E_n^k = f^{-1}\left(\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right) \qquad \text{and} \qquad F_n = f^{-1}((2^n,\infty]),$$

where $0 \leqslant k \leqslant 2^{2n}-1$ and n ranges over all nonnegative integers. Define

$$\phi_n = \sum_{k=0}^{2^{2n}-1} \frac{k}{2^n}\chi_{E_n^k} + 2^n\chi_{F_n}$$

One can check $f_n \nearrow f$ by induction. Further note that $0 \leqslant f - \phi_n \leqslant 1/2^n$ on a set where $f \leqslant 2^n$, which again follows from an induction argument, and completes the proof of (a). To get part (b) from part (a), let $f = g + ih$, where g and h are real functions, so that $f = g^+ - g^- + i(h^+ - h^-)$, where $g^+, g^-, h^+, h^-\colon X \to [0,\infty]$. Then the result follows from part (a) and the triangle inequality. □

Proposition 2.19: Slight Alteration of 2.11.

- If f is measurable and $f = g$ a.e., then g is measurable.
- If $\{f_n\}_{n=1}^{\infty}$ are measurable and $f_n \to f$ a.e., then f is measurable.

Proof. (a) If $f = g$ μ-a.e., then $\mu(\{f = g\}^c) = 0$. Then since μ is complete, any subset of $\{f = g\}^c$ is measurable. We want to show $\{g > a\}$ is measurable for any $a \in \mathbb{R}$. Since

$$\{g > a\} = (\{g > a\} \cap \{f = g\}) \cup \underbrace{(\{g > a\} \cap \{f = g\}^c)}_{\substack{=:N \subset \{f=g\},\\ \mu(N)=0}},$$

we have

$$\underbrace{\{g > a\}}_{\text{measurable}} = \underbrace{\{f > a\}}_{\text{measurable}} \cup \underbrace{N}_{\text{measurable}}.$$

For (b), let $E = \{f_n \to f\}$. Since $f_n \to f$ μ-a.e., $\mu(E^c) = 0$. By Proposition 8, f is measurable on E^c, so $\{f > a\} \cap E^c$ is measurable. Then

$$\underbrace{\{f > a\}}_{\text{measurable}} = \underbrace{\{f > a\} \cap E^c}_{\text{measurable}} \cup \underbrace{\{f > a\} \cap E}_{\substack{\subset \text{ null set } E\\ \Rightarrow \text{ measurable}}}.$$

□

Remark 20. *Proposition 19(a) implies μ is complete. In fact, Proposition 19(b) also implies μ is complete. (Check!)*

Exercise 2.21.

If $X = A \cup B$ where $A, B \in \mathcal{M}$, a function on X is measurable if and only if f is measurable on A and B.

Solution. This proof can be found here. Let $f_A = f|_A$ and $f_B = f|_B$. If f is measurable, then for each $C \in \mathcal{B}_{\mathbb{R}}$ we have for $J \in \{A, B\}$ that

$$(f_J)^{-1}(C) = f^{-1}(C) \cap J \in \mathcal{M}.$$

Hence f_A and f_B are measurable. Now for the converse, note that

$$f^{-1}(C) = (f^{-1}(C) \cap A) \cup (f^{-1}(C) \cap B) = (f_A)^{-1}(C) \cup (f_B)^{-1}(C) \in \mathcal{M}$$

and thus f is measurable. □

Proposition 2.22: 2.12.

Let $(X, \mathcal{M}, \mu)$ be a measure space and $(X, \overline{\mathcal{M}}, \overline{\mu})$ its completion. If f is $\overline{\mathcal{M}}$-measurable on X, then there exists a $\mathcal{M}$-measurable function g such that $f = g$ $\overline{\mu}$-a.e.

Proof. First we show this for simple functions. If $f = \chi_E$, where $E \in \overline{\mathcal{M}}$, then $E = A \cup N$, where $A \in \mathcal{M}$ and $\mu(N) = 0$. Let $g = \chi_A$. Then $\{f \neq g\} \subset N \implies f = g$ $\overline{\mu}$-a.e.

Now suppose f is a simple function, say $f = \sum_{j=1}^{n} z_j \chi_{E_j}$, where $E_j = f^{-1}(\{z_j\})$. One can then use induction to prove there exist $\mathcal{M}$-measurable g such that $\phi = \psi$ μ-a.e. (Check!). By Theorem 18, there exists a sequence $\{\phi_n\}_{n=1}^{\infty}$ such that $\phi_n \to f$. Then there exists a sequence $\{\psi_n\}_{n=1}^{\infty}$ of $\mathcal{M}$-measurable simple functions such that $\phi_n = f$ except on a set E_n, where $\overline{\mu}(E_n) = 0$. We fix $\mathcal{M} \ni N = \bigcup_{n=1}^{\infty} E_n$ such that $\mu(N) = 0$. Then set

$$g = \lim_{n\to\infty} \underbrace{\chi_{(X \setminus N)} \cdot \psi_n}_{\text{measurable}},$$

so that $g = f$ $\overline{\mu}$-a.e. □

2.3 Integration of Nonnegative Functions

Fix a measure space $(X, \mathcal{M}, \mu)$ and define

$$L^+(\mu) = \{\mathcal{M}\text{-measurable functions } f\colon X \to [0,\infty]\}.$$

If the measure space $(X, \mathcal{M}, \mu)$ is understood, then we simply write L^+ to mean $L^+(\mu)$. Note that writing $L^+(\mu)$ specifies not only the measure but the whole measure space $(X, \mathcal{M}, \mu)$, because given a measure $\mu\colon \mathcal{M} \to X$ the σ-algebra $\mathcal{M}$ is specifies since it is the domain of μ and the underlying set X is specified since it is the unique maximal set in $\mathcal{M}$.

Given a simple function $\phi \in L^+$ with standard representation $\phi = \sum_{j=1}^{n} a_j \chi_{E_j}$. We define

$$\int \phi \, d\mu = \sum_{j=1}^{n} a_j \mu(E_j).$$

We may also write this as $\int \phi(x)\, d\mu(x)$, or even simply $\int \phi$ when the measure μ is understood. Here we are using the convention $0 \cdot \infty = 0$. Given $E \in \mathcal{M}$, we define

$$\int_E \phi \, d\mu = \int \phi \chi_E \, d\mu$$

Proposition 2.23: 2.13.

Let ϕ, ψ be simple functions in L^+.

(a) If $c \geqslant 0$, then $\int c\phi = c \int \phi$.
(b) $\int (\phi + \psi) = \int \phi + \int \psi$.
(c) If $\phi \leqslant \psi$, then $\int \phi \leqslant \int \psi$.
(d) The map $\nu_1\colon \mathcal{M} \to [0,\infty]$ given by $\nu_1(E) = \int_E d\mu$ is a measure on $\mathcal{M}$.

Proof. (a), (b), and (c) are immediate by definition of the integral for simple functions. To see (d), note that for any $E \in \mathcal{M}$, $\nu_1(E) = \int_E d\mu = \mu(E)$. Thus $\nu_1 = \mu$, which and we already know μ is a measure, so ν_1 is a measure. □

We now extend the definition of the integral to all $f \in L^+$. We define

$$\int f \, d\mu = \sup\left\{ \int \phi \, d\mu \,\middle|\, 0 \leqslant \phi \leqslant f, \ \phi \text{ simple} \right\}.$$

By Proposition 23, this definition satisfies

$$\int cf = c \int f \text{ for all } c \geqslant 0,$$

and

$$f \leqslant g \implies \int f \leqslant \int g.$$

Theorem 2.24: 2.14: Monotone Convergence Theorem.

If $\{f_n\} \subset L^+(\mu)$ and $f_n \nearrow f$ as $n \to \infty$, then

$$\int f = \lim_{n\to\infty} \int f_n.$$

Proof. Since $f_n \leqslant f$ for all $n \in \mathbb{Z}_{\geqslant 1}$ and the integral is monotone, by taking the limit as $n \to \infty$ we obtain

$$\lim_{n\to\infty} \int f_n \leqslant \int f.$$

It remains to show the reverse inequality. Let ϕ be a simple function with $0 \leqslant \phi \leqslant f$, and let $E_n = \{x \mid f_n(x) \geqslant \alpha\phi(x)\}$ for a fixed $\alpha \in (0,1)$. Then the E_n form an increasing sequence in the sense that $E_1 \subset E_2 \subset \cdots$. The E_j are measurable (Check!) and $\bigcup_{j=1}^{\infty} E_j = X$ (Check!). Now $\int f_n \geqslant \int_{E_n} \geqslant \alpha \int_{E_n} \phi$ by Proposition 23 and continuity from below of the measure $A \mapsto \int_A d\mu$. Since this is true for all $\alpha < 1$, taking the limit as $\alpha \to 1$ from below gives that it also holds for $\alpha = 1$, that is, that $\int f_n \geqslant \int_{E_n} \phi$. Taking the supremum over all simple functions $0 \leqslant \phi \leqslant f$, we obtain

$$\int f \leqslant \lim_{n\to\infty} \int f_n,$$

which completes the proof. □

Before continuing, we introduce an extremely powerful technique for proving results with the theory we have so far developed: an induction principle on measurable functions:

Theorem 2.25: Induction Principle for Measurable Functions.

Let $(X, \mathcal{M})$ be a measurable space, $\mathcal{F}$ the set of measurable functions (resp. measurable nonnegative functions) with property P such that the following hold.

(a) For all $E \in \mathcal{M}$, $\chi_E \in \mathcal{F}$.

(b) For all $f, g \in \mathcal{F}$, $af + bg \in \mathcal{F}$ for all $a, b \in \mathbb{R}$ (resp. all nonnegative $a, b \in \mathbb{R}$).

(c) For all $\{f_n\}_{n=1}^{\infty} \subset \mathcal{F}$ such that $f_n \nearrow f$, $f \in \mathcal{F}$.

Then $\mathcal{F}$ contains *all* measurable functions (resp. nonnegative measurable functions).

Proof. Prove the case of nonnegative measurable functions first as follows: Note (a) with (b) implies simple functions are in $\mathcal{F}$. Then approximate any $f \in L^+$ by simple functions, and then apply (c). To prove the case for all $a, b \in \mathbb{R}$, then $\mathcal{F}$ satisfying (a), (b'), and (c) contain all measurable functions by applying the nonnegative case above to each of f^+, f^-, then using part (b) to the sum $f = f^+ - f^-$ to get the result, since $f^+, f^- \geqslant 0$ and are measurable. ☐

Theorem 2.26: 2.15: MCT for Series.

If $\{f_n\} \subset L^+$, then

$$\int \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int f_n.$$

Proof. Suppose $f_1, f_2 \in L^+$. Then there exist simple functions ϕ_n, ψ_n such that $\phi_n \nearrow f_1$ and $\psi_n \nearrow f_2$. Then $\phi_n + \psi_n \nearrow f_1 + f_2$. Then

$$\int (f_1 + f_2) \overset{\text{(MCT)}}{=} \int \lim_{n\to\infty} (\phi_n + \psi_n) = \lim_{n\to\infty} \int \phi_n + \lim_{n\to\infty} \int \psi_n \overset{\text{(MCT)}}{=} \int f + \int g.$$

Continuing similarly, we can see N,

$$\int \sum_{n=1}^{N} f_n = \sum_{n=1}^{N} \int f_n. \tag{2.26.1}$$

for all positive integers N. Sending $n \to \infty$, we obtain

$$\sum_{n=1}^{\infty} \int f_n = \lim_{N\to\infty} \sum_{n=1}^{N} \int f_n \overset{(2.26.1)}{=} \lim_{N\to\infty} \int \sum_{n=1}^{N} f_n \overset{\text{(MCT)}}{=} \int \sum_{n=1}^{\infty} f_n.$$ ☐

Proposition 2.27: 2.16.

Let $f \in L^+$. Then

$$\int f = 0 \qquad \Longleftrightarrow \qquad f = 0 \text{ a.e.}$$

Proof. ($\Rightarrow$) If f is simple, say with standard representation $\sum_{j=1}^{n} a_j \chi_{E_j}$, then

$$0 = \int f = \sum_{j=1}^{n} a_j \mu(E_j) \iff a_j = 0 \text{ or } \mu(E_j) = 0 \text{ for all } j = 1, \dots, n,$$

so f can only be nonzero on null sets, and hence $f = 0$ a.e.

Now suppose f is any function and $f = 0$ a.e. Then any simple function ϕ with $0 \leqslant \phi \leqslant f$ must also equal 0 a.e., and since $\int f$ is defined as the supremum of such simple functions we conclude $\int f = 0$.

($\Leftarrow$) We will show that if $f \neq 0$ a.e., then $\int f \neq 0$. Suppose $f \neq 0$ a.e. Then $\mu(\{f > 0\})0$. We can write

$$\{f > 0\} = \bigcup_{n=1}^{\infty} \underbrace{\{f(x) > 1/n\}}_{:=E_n}.$$

Then $\mu(E_n) > 0$ for some $n \in \mathbb{Z}_{\geqslant 1}$. Then $f > \chi_{E_n}/n$, so $\int f > \mu(E_n)/n > 0$, which completes the proof. □

Corollary 2.28: 2.17: MCT for Convergence Almost Everywhere.

If $\{f_n\} \subset L^+$, $f \in L^+$ and $f_n \nearrow f$ a.e., then

$$\int f = \lim_{n\to\infty} \int f_n.$$

Proof. Suppose f_n increases to $f(x)$ for all $x \in E$ and $\mu(E^c) = 0$. Then

$$f\chi_E(x) = \begin{cases} f(x) & \text{if } x \in E, \\ 0 & \text{if } x \in E^c. \end{cases}$$

So, $f - f\chi_E = 0$ a.e. Similarly, $f_n - f_n\chi_E = 0$ a.e. Then

$$\int f \overset{(24)}{=} \int f\chi_E \overset{(\text{MCT})}{=} \lim_{n\to\infty} f_n\chi_E \overset{(27)}{=} \lim \int f_n.$$ □

Remark 29. *The assumption that $\{f_n\}$ be increasing to f is a crucial one. For example, consider $f_n = n\chi_{(0,1/n)}$ for each x. But, with respect to the Lebesgue measure, we have*

$$\lim_{n\to\infty} f_n = 0 \neq 1 = n\left(\frac{1}{n}\right) = \int f_n.$$

Theorem 2.30: 2.18: Fatou's Lemma.

If $\{f_n\}_{n=1}^{\infty} \subset L^+$, then

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

Proof. For each fixed $k \in \mathbb{Z}_{\geqslant 1}$, we have $\inf_{n \geqslant k} f_n \leqslant f_j$ whenever $j \geqslant k$. Thus $\int \inf_{n\geqslant k} f_n \leqslant \int f_j$ whenever $j \geqslant k$. Taking the infimum of both sides over all $j \geqslant k$, we obtain

$$\int \inf_{n\geqslant k} f_n \leqslant \inf_{j\geqslant k} \int f_j \tag{2.30.1}$$

Letting $k \to \infty$, we find

$$\int \lim_{k\to\infty} \inf_{n\geqslant k} f_n \leqslant \lim_{k\to\infty} \int \inf_{n\geqslant k} f_n \overset{(2.30.1)}{\leqslant} \liminf_{n\to\infty} \int f_n. \qquad \square$$

Corollary 2.31: 2.19.

If $\{f_n\} \subset L^+$, $f \in L^+$, and $f_n \to f$ a.e., then

$$\int f \leqslant \liminf_{n\to\infty} \int f_n.$$

Proof. This is almost *verbatim* the proof of Corollary 28. □

Proposition 2.32: 2.20.

If $f \in L^+$ and $\int f < \infty$, then $\{f = \infty\}$ is a null set, and $\{f > 0\}$ is σ-finite.

Proof. $\{f = +\infty\} = \bigcap_{n=1}^{\infty}\{f \geqslant n\}$. We have for each n that

$$\mu(\{f = +\infty\}) \leqslant \mu(\{f \geqslant n\}) = \int_{\{f\geqslant n\}} d\mu \leqslant \frac{1}{n}\int f < \frac{1}{n}M$$

for some $M > 0$, where such an M exists because $\int f < \infty$. Sending $n \to \infty$, the right-hand side vanishes. Thus $\{f = \infty\}$ has measure zero.

We now show $\{f > 0\}$ is σ-finite. Write

$$\{f > 0\} = \bigcup_{n=1}^{\infty} \underbrace{\{f > 1/n\}}_{=:\, E_n},$$

so if $F_n = E_n \setminus \bigcup_1^{n-1} E$ then $\{f > 0\}$ is the countable disjoint union of F_n. If some F_n has infinite measure, then

$$\int f = \underbrace{\int_{F_n} f}_{=\infty} + \underbrace{\int_{F_n^c} f}_{\substack{\geqslant\, 0,\text{ since} \\ f\in L^+}} = \infty,$$

contradicting $\int f < \infty$. □

Exercise 2.33: Folland Exercise 2.13.

Suppose $\{f_n\} \subset L^+$, $\int f < \infty$. $f_n \to f$, and $\int f_n \to \int f < \infty$, then $\int_E f_n \to \int_E f$ for all $E \in \mathcal{M}$.

Solution. Let $E \in \mathcal{M}$. Since $f_n \to f$ pointwise, $f \in L^+$. Then

$$\int_E f = \int f\chi_E \overset{(30)}{\leqslant} \liminf_{n\to\infty} \int f_n\chi_E = \liminf_{n\to\infty} f_n,$$

so we only need to show $\limsup_{n\to\infty} \int_E f_n \leqslant \int f$. Since $f - f\chi_E \in L^+$, we can write

$$\int f - \int_E f \overset{(30)}{\leqslant} \liminf_{n\to\infty} \int (f_n - f_n\chi_E) = \liminf_{n\to\infty} \int f_n - \limsup_{n\to\infty} \int_E f_n = \int f - \limsup_{n\to\infty} \int_E f$$

Since $\int f < \infty$, we can subtract it from both sides to obtain $\limsup_{n\to\infty} \int_E f_n \leqslant \int f$. ☐

Remark 34. *The above claim fails if $\int_E f_n \to \int_E f = \infty$. To see this, consider the Lebesgue measure space $(\mathbb{R}, \mathcal{L}, m)$, $f = \chi_{[2,\infty)}$, $f_n = \chi_{[2,\infty)} + n\chi_{[0,1/n)}$ and $E = [0,1)$. In this case $f_n \to f$ pointwise, $\int f = \int f_n = \infty$ and hence $\int f = \lim_{n\to\infty} \int f_n$, but*

$$\int_E f = 0 \neq 1 = \lim_{n\to\infty} 1 = \lim_{n\to\infty} n \cdot m([0,1/n)) = \lim_{n\to\infty} \int_E f_n.$$

Exercise 2.35: Folland Exercise 2.16.

If $f \in L^+$ and $\int f < \infty$, then for all $\varepsilon > 0$ there exists $E \in \mathcal{M}$ such that $\mu(E) < \infty$ and $\int_E f > \left(\int f\right) - \varepsilon$.

Solution. Let $\varepsilon > 0$. Since $\sup\left\{\int \phi \,\middle|\, 0 \leqslant \phi \leqslant f, \text{ where } \phi \text{ is simple}\right\} = \int f < \infty$, by definition of the supremum there exists a simple function ϕ such that $0 \leqslant \phi \leqslant f$ and

$$\int \phi > \left(\int f\right) - \varepsilon. \tag{2.35.1}$$

Write can ϕ as $\phi = \sum_{j=1}^n a_j \chi_{E_j}$, where $a_j \geqslant 0$ and the E_j are disjoint elements of $\mathcal{M}$. Let $E = \bigcup_{j=1}^n E_j$. Then

$$\int \phi = \int_E \phi \leqslant \int_E f \tag{2.35.2}$$

by monotonicity of the integral. Combining Equations (2.35.1) and (2.35.2), we conclude

$$\int_E f \geqslant \int_E f \geqslant \int \phi > \left(\int f\right) - \varepsilon.$$ ☐

Exercise 2.36: Folland Exercise 2.17.

Assume Fatou's Lemma and deduce the MCT from it.

Solution. Let $\{f_n\}_{n=1}^\infty \subset L^+$, $f_n \nearrow f$ (so $f \in L^+$). We want to show $\int f = \lim_{n\to\infty} \int f_n$. By Fatou's Lemma, $\int f \leqslant \liminf_{n\to\infty} \int f_n$, so we need to show $\limsup_n \int f_n \leqslant \int f$. But $f_n \leqslant f$ for all n, so taking the limsup of both sides gives $\limsup_{n\to\infty} \int f_n \leqslant \limsup f = \int f$, as desired. ☐

2.4 Integration of Complex Functions

We again fix a measure space $(X, \mathcal{M}, \mu)$. The integral can be extended from L^+ to all complex-value d measurable functions in the following way. We first extend the integral to all real-valued functions by considering the positive part f^+ and the negative part f^- of a given measurable real valued function $f\colon X \to \mathbb{R}$, and define

$$\int f = \int f^+ - \int f^-.$$

whenever at least one of $\int f^+$ and $\int f^-$ is finite. If *both* $\int f^+$ and $\int f^-$ are both $< \infty$, or equivalently if $\int |f| < \infty$, we say f is ***integrable***. We say a measurable complex-valued function $f\colon X \to \mathbb{C}$ is ***integrable*** if $\int |f| < \infty$. More generally, if $E \in \mathcal{M}$, f is ***integrable on E*** if $\int_E |f| < \infty$. Since $|f| \leqslant |\operatorname{Re} f| + |\operatorname{Im} f| \leqslant 2|f|$, f is integrable if and only if $\operatorname{Re} f$ and $\operatorname{Im} f$ are both integrable, and in this case we define

$$\int f = \int \operatorname{Re} f + i \int \operatorname{Im} f$$

Proposition 2.37: 2.21.

The set of real-valued integrable functions is a complex valued vector space and the integral is a linear functional on it.

Proof.
This follows from the fact that $|af + bg| \leqslant |a||f| + |b||g|$, and it is easy to check that $\int af = a \int f$ for any $a \in \mathbb{R}$. (Check!) To show additivity, suppose that f and g are integrable and let $h = f + g$. Then $h^+ - h^- = f^+ - f^- + g^+ - g^-$, so $h^+ + f^- + g^- = h^- + f^+ + g^+$. By Theorem 26,

$$\int h^+ + \int f^- + \int g^- = \int h^- + \int f^+ + \int g^+,$$

and regrouping then yields the desired result:

$$\int h = \int h^+ - \int h^- = \int f^+ - \int f^- + \int g^+ - \int g^- = \int f + \int g. \qquad \square$$

Showing the analogous statement for the complex case is left as an exercise.

The superscript 1 is standard notation, but it will not assume any significance for us until Chapter 6.

Proposition 2.38: 2.22.

If $f\colon X \to \mathbb{C}$ is integrable, then $\left|\int f\right| \leqslant \int |f|$.

Proof. The claim is true if $\int f = 0$, and if f is real valued then

$$\left|\int f\right| = \left|\int f^+ - \int f^-\right| \leqslant \int f^+ + \int f^- = \int |f|,$$

which also affirms the claim. If f is complex-valued and $\int f \neq 0$, let $\alpha = \overline{\operatorname{sgn}(\int f)}$. Then $|\int f| = \alpha \int f = \int \alpha f$, so $\int \alpha f$ is real, which means

$$\left|\int f\right| = \operatorname{Re}\int \alpha f = \int \operatorname{Re}(\alpha f) \leqslant \int |\operatorname{Re}(\alpha f)| \leqslant \int |\alpha f| = \int |f|. \qquad \square$$

Proposition 2.39: 2.23.

(a) If $f\colon X \to \mathbb{C}$ is integrable, then $\{f \neq 0\}$ is σ-finite and $\{|f| = +\infty\}$ is a null set.
(b) If $f, g\colon X \to \mathbb{C}$ are integrable, then

$$\int_E f = \int_E g \text{ for all } E \in \mathcal{M} \iff \int |f-g| = 0 \iff f = g \text{ a.e.}$$

Proof. (a) and the second equivalence in (b) follow from Propositions 27 and 32. If $\int |f-g| = 0$, then for any $E \in \mathcal{M}$,

$$\left|\int_E f - \int_E g\right| \overset{(38)}{\leqslant} \int \chi_E |f-g| \leqslant \int |f-g| = 0,$$

so that $\int_E f = \int_E g$. Conversely, if $f \neq g$ a.e., then at least one of the positive or negative parts of the functions $u = \operatorname{Re}(f-g)$ or $v = \operatorname{Im}(f-g)$ are nonzero on a set of positive measure. We may assume $u^+ > 0$ to is nonzero on a set E of positive measure, since the other cases are similar. In this case we have $\operatorname{Re}(\int_E f - \int_E g) = \int_E u^+ > 0$ since $u^- = 0$ on E, affirming the claim. $\square$

Proposition 39 shows that for the purposes of integration it makes no difference if we alter functions on null sets. Indeed, one can integrate functions f that are only defined on a measurable set E whose complement is null simply by defining f to be zero (or anything else) on E^c. In this fashion we can treat $\overline{\mathbb{R}}$-valued functions that are finite a.e. as real-valued functions for the purposes of integration.

With this in mind, it is convenient to define $L^1(\mu)$ (or $L^1(X,\mu)$, or $L^1(X)$, or simply L^1, depending on the context) as the set of equivalence classes of a.e.-defined integrable functions on X, where f and g are considered equivalent if and only if $f = g$ a.e. Then $L^1(\mu)$ is a complex vector space under pointwise a.e. addition and scalar multiplication, and we will write "$f \in L^1(\mu)$" to mean that f is an a.e.-defined integrable function.

Remark 40 (Very important remark)**.** *Here we will make some critical observations.*

(1) If $\overline{\mu}$ is the completion of μ, Proposition 22 yields a natural one-to-one correspondence between $L^1(\overline{\mu})$ and $L^1(\mu)$, so we can and will identify these spaces. In other words, when discussing the complex vector space of complex-valued integrable functions with respect to a measure space $(X, \mathcal{M}, \mu)$, we may assume μ is complete.

(2) L^1 is a metric space with distance function $\rho(f,g) = \int |f-g|$. (The triangle inequality is easily verified, and obviously $\rho(f,g) = \rho(g,f)$; but to obtain the condition that $\rho(f,g) = 0$ only when $f = g$, one must identify functions that are equal a.e., according

to Proposition 39(b).) We shall refer to convergence with respect to this metric as ***convergence in*** $\boldsymbol{L^1}$*; thus* $f_n \to f$ *in* L^1 *if and only if* $\int |f_n - f| \to 0$.

Exercise 2.41.

The integral is absolutely continuous on L^1. In other words, if $f \in L^1$, then for all $\varepsilon > 0$, there exists $\delta > 0$ such that for all $E \in \mathcal{M}$,

$$\mu(E) < \delta \quad \Longrightarrow \quad \int_E |f| < \varepsilon.$$

Solution. Since $f \in L^1$, $\mu(\{|f| = \infty\}) = 0$. Thus there exists some M such that $\mu(\{|f| > M\}) < \varepsilon/2$. Let $\delta = \varepsilon/(2M)$ and suppose $E \in \mathcal{M}$ has $\mu(E) < \delta$. Then

$$\int_E |f| = \int_{\{f \geqslant M\} \cap E} |f| + \int_{\{f < M\} \cap E} |f| \leqslant \int_{\{f \geqslant M\}} |f| + M \int_{\{f < M\} \cap E} d\mu < \frac{\varepsilon}{2} + \cancel{M} \frac{\varepsilon}{2\cancel{M}} = \varepsilon,$$

as desired. □

Theorem 2.42: Folland Exercise 2.18: Strenghtened Fatou's Lemma.

Let $\{f_n \colon X \to \overline{\mathbb{R}}\}$ be any sequence of measurable functions.

(a) If for all n, $-f_n \leqslant g$ for some $g \in L^1 \cap L^+$, then

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

(b) If for all n, $f_n \leqslant g$ for some $g \in L^1 \cap L^+$, then

$$\limsup_{n\to\infty} \int f_n \leqslant \int \limsup_{n\to\infty} f_n.$$

Proof. By Proposition 39(b), it suffices to replace the "a.e." with "everywhere". We first show (a). The hypotheses imply $\{g + f_n\}_{n=1}^\infty \subset L^+$, so

$$\int g + \int \liminf_{n\to\infty} f_n = \int \liminf_{n\to\infty} (g + f_n) \overset{(30)}{\leqslant} \liminf_{n\to\infty} \int (g + f_n) = \int g + \liminf_{n\to\infty} \int f_n.$$

Since $\int g$ is finite, we can subtract it from both sides to obtain the desired result.

For (b), the hypotheses imply $\{g - f_n\}_{n=1}^\infty \subset L^+$, so

$$\int g - \int \liminf_{n\to\infty} (-f_n) = \int \liminf_{n\to\infty} (g - f_n) \overset{(30)}{\leqslant} \liminf_{n\to\infty} \int (g - f_n) = \int g - \limsup_{n\to\infty} \int f_n.$$

Since $\int g$ is finite, we can subtract it from both sides and use that $\int \liminf_{n\to\infty}(-f_n) = \int \limsup_{n\to\infty} f_n$ to obtain the desired result. □

The following theorem has an intuitive explanation. In the context of integration on $\mathbb{R}$ with Lebesgue measure as in the discussion preceding Fatou's Lemma, the idea behind this theorem is that if $f_n \to f$ a.e. and the graph of $|f_n|$ is confined to a region of the plane with finite area so that the area beneath it cannot escape to infinity, then $\int f_n \to \int f$.

Theorem 2.43: 2.24: Dominated Convergence Theorem (DCT).

If $\{f_n : X \to \mathbb{C}\}_{n=1}^{\infty}$ is a sequence of measurable functions such that $f_n \to f$ a.e. and there is some $g \in L^1$ such that $|f_n| \leqslant g$ a.e. for all n, then $f \in L^1$ and

$$\int f = \lim_{n\to\infty} \int f_n.$$

Proof. Note $f \in L^1$ since otherwise $|f(x)| \geqslant g$ for all x in some set of positive measure, so since $f_n \to f$ a.e. we also have $|f_n(x)| \geqslant g$ for large enough n, contradicting the hypothesis. We may assume f is real-valued, since we can show convergence in real and imaginary parts. Since $|f_n| \leqslant g$ implies $-g \leqslant f_n \leqslant g$, that is, $-f_n \leqslant g$ and $f_n \leqslant -g$, we obtain

$$\limsup_{n\to\infty} \int f_n \overset{(42(b))}{\leqslant} \int \limsup_{n\to\infty} f_n = \int f = \int \liminf_{n\to\infty} f_n \overset{(42(a))}{\leqslant} \int \liminf_{n\to\infty} \int f_n,$$

so $\lim_{n\to\infty} \int f_n$ exists and equals $\int f$. □

Exercise 2.44: Folland Exercise 2.19.

Suppose $\{f_n\}_{n=1}^{\infty} \subset L^1$ and $f_n \to f$ uniformly.

(a) If $\mu(X) < \infty$, then $f \in L^1$ and $\int f_n \to \int f$.

(b) If $\mu(X) = \infty$, then the conclusions of (a) can fail.

Solution. (a) If $f_n \to f$ a.e., then for all sufficiently large n we have $|f_n - f| < 1$, and hence that $|f_n| \leqslant |f| + 1, |f| \leqslant |f_n| + 1$, so $f \in L^1$. And $1 \in L^1$ because $\int |1| = \mu(X) < \infty$, so $|f_n| \leqslant |f| + 1 \in L^1$. Then by the DCT, $f \in L^1$ and $\int f_n \to f$.

(b) If $f_n = \chi_{[0,n]}/n$. Then $f_n \to 0$ uniformly, but $\lim_{n\to\infty} \int f_n = \lim_{n\to\infty} 1 = 1 \neq 0$. □

Exercise 2.45: Folland Exercise 2.26.

Let m be the Lebesgue measure on $\mathbb{R}$. If $f \in L^1(m)$, then

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

is continuous on $\mathbb{R}$.

Solution. Let $x_n \to x$. Then $|f\chi_{(-\infty,x_n]}| \leqslant |f| \in L^1(m)$ and $f\chi_{(-\infty,x_n]} \to f\chi_{(-\infty,x]}$ a.e, so by the DCT we have

$$\lim_{j\to\infty} F(x_n) = \lim_{j\to\infty} \int f\chi_{(-\infty,x_n]} = \int f\chi_{(-\infty,x]} = F(x).$$

Hence F is continuous. □

Exercise 2.46: Folland Exercise 2.20: Generalized DCT.

If $g \in L^1$, $f_n \to f$ a.e., $g_n \to g$ a.e., $|f_n| \leqslant g_n$, and $\int g_n \to \int g$, then $f \in L^1$ and

$$\lim_{n\to\infty} \int f_n = \int f.$$

Solution. Since $|f_n| \leqslant g_n$, $\{g_n + f_n\}_{n=1}^{\infty}, \{g_n - f_n\}_{n=1}^{\infty} \subset L^1 \cap L^+$. Applying Fatou's Lemma to both, we obtain

$$\int g + \int f \leqslant \liminf_{n\to\infty} \int (g_n + f_n) = \int g + \liminf_{n\to\infty} \int f_n,$$
$$\int g - \int f \leqslant \liminf_{n\to\infty} \int (g_n - f_n) = \int g - \limsup_{n\to\infty} \int f_n.$$

Since $g \in L^1$, we can subtract $\int g$ from both sides to obtain

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

Thus $\lim_{n\to\infty} \int f_n = \int f$. □

Exercise 2.47: Folland Exercise 2.21.

Suppose $f_n, f \in L^1$ and $f_n \to f$ a.e. Then

$$\int |f_n - f| \to 0 \qquad \Longleftrightarrow \qquad \int |f_n| \to \int |f|.$$

(In other words, if $f_n, f \in L^1$ and $f_n \to f$ a.e., then $f_n \to f$ in L^1 if and only if $\int |f_n| \to \int |f|$.)

Solution. If $\int |f_n - f| \to 0$, then by the triangle inequality $0 \leqslant \int |f_n| - \int |f| \leqslant \int |f_n - f| \to 0$, so $\int |f_n| \to \int |f|$ by the squeeze theorem. (Note the a.e. convergence hypothesis was not used).

Conversely, if $\int |f_n| \to \int |f|$, then $g_n := |f| + |f_n| \geqslant 0$ and $h_n := |f_n - f|$ satisfy

- $h_n, g_n \in L^1$,
- $|h_n| = |f_n - f| \leqslant |f_n| + |f| = g_n$, and
- $\int g_n = \int |f_n| + \int |f| \to 2 \int |f| \in L^1$ (by hypothesis),

so by Folland Exercise 2.20 we conclude $\int h_n \to \int \lim_{n\to\infty} h_n = \int 0 = 0$, that is, $\int |f_n - f| \to 0$. □

Theorem 2.48: 2.25: DCT for Series.

Suppose that $\{f_n\}$ is a sequence in L^1 such that $\sum_{n=1}^{\infty} \int |f_n| < \infty$. Then $\sum_{n=1}^{\infty} f_n$

converges a.e. to a function in L^1, and

$$\int \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int f_n.$$

Proof. By Theorem 26, $\int \sum_1^\infty |f_j| = \sum_1^\infty \int |f_j| < \infty$, so the function $g = \sum_1^\infty |f_j|$ is in L^1. In particular, by Proposition 32 $\sum_1^\infty |f_j(x)|$ is finite for a.e. x, and for each such x the series $\sum_1^\infty f_j(x)$ converges. Moreover, $|\sum_1^n f_j| \leqslant g$ for all n, so we can apply the DCT to the sequence of partial sums to obtain $\int \sum_1^\infty f_j = \sum_1^\infty \int f_j$. □

Theorem 2.49: 2.26.

(a) The integrable simple functions are dense in L^1 in the L^1 metric. More precisely, if $f \in L^1(\mu)$ and $\varepsilon > 0$, there is an integrable simple function $\phi = \sum_{j=1}^n a_j \chi_{E_j}$ such that $\int |f - \phi| < \varepsilon$.

(b) If μ is a Lebesgue-Stieltjes measure on $\mathbb{R}$, the sets E_j in the definition of ϕ can be taken to be finite unions of open intervals; moreover, there is a continuous function g that vanishes outside a bounded interval such that $\int |f - g| d\mu < \varepsilon$.

Proof. Let $\{\phi_n\}$ be as in Theorem 18(b); then $\int |\phi_n - f| < \varepsilon$ for n sufficiently large by the DCT, since $|\phi_n - f| \leqslant 2|f|$. If $\phi_n = \sum a_j \chi_E$, where the E_j are disjoint and the a_j are nonzero, we observe that $\mu(E_j) = |a_j|^{-1} \int_{E_j} |\phi_n| \leqslant |a_j|^{-1} \int |f| < \infty$. Moreover, if E and F are measurable sets, we have $\mu(E \triangle F) = \int |\chi_E - \chi_F|$. Thus if μ is a Lebesgue-Stieltjes measure on $\mathbb{R}$, by Proposition 47 we can approximate χ_{E_j} arbitrarily closely in the L^1 metric by finite sums of functions χ_{I_k} where the I_ks are open intervals. Finally, if $I_k = (a, b)$ we can approximate χ_{I_k} in the L^1 metric by continuous functions that vanish outside (a, b). (For example, given $\varepsilon > 0$, take g to be the continuous function that equals 0 on $(-\infty, a]$ and $[b, \infty)$, equals 1 on $[a + \varepsilon, b - \varepsilon]$, and is linear on $[a, a + \varepsilon]$ and $[b - \varepsilon, b]$.) Putting these facts together, we obtain the desired assertions. □

The next theorem gives a criterion, less restrictive than those found in most advanced calculus books, for interchanging limits or derivatives with integrals.

Theorem 2.50: 2.27.

Suppose for each $t \in [a, b]$ $(-\infty < a < b < \infty)$, $f_t \colon X \to \mathbb{C}$ is L^1.

(a) (Interchanging integrals with limits). If for all t, $|f_t| \leqslant g \in L^1(\mu)$, then

$$\lim_{t \to t_0} \int f_t = \int \lim_{t \to t_0} f_t,$$

whenever $\lim_{t \to t_0} f_t = f_{t_0}$.

(b) (Interchanging integrals with derivatives). If for all t, $|\partial f_t/\partial t| \leqslant g \in L^1(\mu)$, then
$$\frac{d}{dt}\int f_t = \int \frac{\partial f_t}{\partial t},$$
whenever $\partial f/\partial t$ exists.

Proof. For (a), apply the DCT to $f_n(x) = f(x, t_n)$ where $\{t_n\}$ is any sequence in $[a, b]$ converging to t_0. For (b), observe that
$$\frac{\partial f}{\partial t}(x, t_0) = \lim h_n(x) \text{ where } h_n(x) = \frac{f(x, t_n) - f(x, t_0)}{t_n - t_0},$$
$\{t_n\}$ again being any sequence converging to t_0. It follows that $\partial f/\partial t$ is measurable, and by the mean value theorem,
$$|h_n(x)| \leqslant \sup_{t\in[a,b]}\left|\frac{\partial f}{\partial t}(x, t)\right| \leqslant g(x),$$
so the DCT can be invoked again to give
$$F'(t_0) = \lim \frac{F(t_n) - F(t_0)}{t_n - t_0} = \lim \int h_n(x)d\mu(x) = \int \frac{\partial f}{\partial t}(x, t)d\mu(x)$$
Note that the device of using sequences converging to t_0 in this proof is technically necessary because the DCT deals only with sequences of functions. However, in such situations we shall usually just say "let $t \to t_0$" with the understanding that sequential convergence is underlying the argument. □

It is important to note that in Theorem 50 the interval $[a, b]$ on which the estimates on f or $\partial f/\partial t$ hold might be a proper subinterval of an open interval I (perhaps $\mathbb{R}$ itself) on which $f(x, \cdot)$ is defined. If the hypotheses of (a) or (b) hold for all $[a, b] \subset I$, perhaps with the dominating function g depending on a and b, one obtains the continuity or differentiability of the integrated function F on all of I, as these properties are local in nature.

Example 51 (Folland Exercise 2.28). *We will compute the limits*

(b) $\lim_{n\to\infty} \int_0^1 \frac{1+nx^2}{(1+x^2)^{-n}}\, dx$ *and*

(c) $\lim_{n\to\infty} \int_0^\infty \frac{n\sin(x/n)}{x(1+x^2)}\, dx$.

For (b), note that by the Bernoulli inequality we have $(1+x^2)^n \geqslant 1+nx^2$, *so* $(1+nx^2)/(1+x^2) \leqslant 1 \in L^1(m, [0, 1])$. *Then by the DCT,*
$$\lim_{n\to\infty} \int_0^1 \frac{1 + nx^2}{(1 + x^2)^n}\, dx = \int_0^1 \left(\lim_{n\to\infty} \frac{1 + nx^2}{(1 + x^2)^n}\right) = \int_0^1 0\, dx = 0.$$
For (c), first estimate by writing
$$\left|\frac{n\sin(x/n)}{x(1 + x^2)}\right| \leqslant \frac{1}{1 + x^2} \qquad \text{and} \qquad \int_0^\infty \frac{dx}{1 + x^2} = \frac{\pi}{2} < \infty,$$

(where the first inequality is for all sufficiently large n) so by the DCT

$$\lim_{n\to\infty}\int_0^\infty \frac{n\sin(x/n)}{x(1+x^2)}\,dx = \int_0^\infty \lim_{n\to\infty}\frac{n\sin(x/n)}{x(1+x^2)}\,dx = \int_0^\infty \frac{dx}{1+x^2}\,dx = \frac{\pi}{2}.$$

2.4.1 Comparing the Riemann and Lebesgue Integrals

In the special case where the measure μ is Lebesgue measure m on $\mathbb{R}$, the integral we have developed is called the ***Lebesgue integral***. Let $[a,b]$ be a compact interval in $\mathbb{R}$. A ***partition*** of $[a,b]$ is a finite sequence $P = \{t_j\}_{j=0}^n$ such that $a = t_0 < t_1 < \cdots < t_n = b$. Let f be an arbitrary bounded real-valued function on $[a,b]$. For each partition P we define

$$S_P f = \sum\nolimits_1^n M_j(t_j - t_{j-1}) \qquad \text{and} \qquad s_P f = \sum\nolimits_1^n m_j(t_j - t_{j-1}),$$

where M_j and m_j are the supremum and infimum of f on $[t_{j-1}, t_j]$. Then we define

$$\overline{I}_a^b(f) = \inf_P S_P f \qquad \text{and} \qquad \underline{I}_a^b(f) = \sup\nolimits_P s_P f$$

where the infimum and supremum are taken over all partitions P. If $\overline{I}_a^b(f) = \underline{I}_a^b(f)$, their common value is the Riemann integral $\int_a^b f(x)\,dx$, and f is called ***Riemann integrable***.

Theorem 2.52: 2.28.

Let $f\colon [a,b] \to \mathbb{R}$ be any bounded function.

(a) If f is Riemann integrable, then f is Lebesgue measurable (and hence integrable on $[a,b]$ since it is bounded), and

$$\int_a^b f(x)\,dx = \int_{[a,b]} f\ dm.$$

(b) f is Riemann integrable if and only if the set

$$D(f) = \{x \in [a,b] \mid f \text{ is discontinuous at } x\}$$

has Lebesgue measure zero.

Proof. We adopt the notation from above. Suppose f is Riemann integrable. For each partition P of $[a,b]$, define simple functions

$$G_P = \sum\nolimits_{j=1}^n M_j \chi_{(t_{j-1},t_j]} \qquad \text{and} \qquad g_P = \sum\nolimits_1^n m_j \chi_{(t_{j-1},t_j]},$$

so that $S_P f = \int G_P\,dm$ and $s_P f = \int g_P\ dm$. There is a sequence $\{P_k\}$ of partitions whose mesh (i.e., $\max_j(t_j - t_{j-1})$) tends to zero, each of which includes the preceding one (so that g_{P_k} increases with k while G_{P_k} decreases), such that $S_{P_k} f$ and $s_{P_k} f$ converge to $\int_a^b f(x)\,dx$. Let $G = \lim_{k\to\infty} G_{P_k}$ and $g = \lim_{k\to\infty} g_{P_k}$. Then $g \leqslant f \leqslant G$, and by the DCT, $\int G\,dm = \int g\,dm = \int_a^b f(x)\,dx$. Hence $\int (G-g)\,dm = 0$, so by Proposition 27 $G = g$ a.e., and thus $G = f$ a.e. Since G is measurable (being the limit of a sequence of simple functions) and m is complete, f is measurable and $\int_{[a,b]} f\,dm = \int G\,dm = \int_a^b f(x)\,dx$. This proves (a).

To prove (b), we first prove the following lemma.

Lemma 2.53: Folland Exercise 2.23.

Given a bounded function $f\colon [a,b] \to \mathbb{R}$, the following hold for the functions given by

$$H(x) = \lim_{\delta\to 0} \sup_{|y-x|\leqslant\delta} f(y) \qquad \text{and} \qquad h(x) = \lim_{\delta\to 0} \inf_{|y-x|\leqslant\delta} f(y).$$

(i) $H(x) = h(x)$ if and only if f is continuous at x.
(ii) In the notation of the proof of part (a) above, $H = G$ a.e. and $h = g$ a.e.
(iii) H and h are Lebesgue measurable, and $\int_{[a,b]} H\,dm = \overline{I}_a^b(f)$ and $\int_{[a,b]} h\,dm = \underline{I}_a^b(f)$. □

Proof. To prove (i), suppose $\varepsilon > 0$. Since f is continuous at x, there exists $\delta > 0$ such that $f(x) - \varepsilon \leqslant f(y) \leqslant f(y) + \varepsilon$ whenever $|x-y| < \delta$. Thus

$$f(x) - \varepsilon \leqslant \inf_{|x-y|<\delta} f(y) \leqslant f(x) \leqslant \sup_{|x-y|<\delta} \leqslant f(y) + \varepsilon.$$

Sending $\varepsilon \to 0$ (and hence we may assume $\delta \to 0$), we obtain

$$H(x) = \lim_{\delta\to 0} \sup_{|x-y|<\delta} f(y) = f(x) = \lim_{\delta\to 0} \inf_{|x-y|<\delta} f(y) = h(x),$$

as desired. Conversely, suppose $H(x) = h(x)$. Then

$$\lim_{\delta\to 0}\left(\sup_{|x-y|<\delta} f(y) - \inf_{|x-y|<\delta} f(y)\right) = 0,$$

so given $\varepsilon > 0$, there exists $\delta_0 > 0$ such that for all $|x-y| < \delta_0$,

$$(|f(x) - f(y)| \leqslant) \sup_{|x-y|<\delta_0} f(y) - \inf_{|x-y|<\delta_0} f(y) < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we conclude f is continuous at x. This proves (i).

The proof of (ii) goes as follows, and can be found here. Consider the sequence of partitions $\{P_k\}$ used in the proof of part (a) above. Now set $E = \{\text{points of } P_k \text{ for all } k\}$. Since each P_k has a finite number of points, E is countable and hence has Lebesgue measure zero.

We will show that $H = G$ in $[a,b] \smallsetminus E$. If $x \in [a,b] \smallsetminus E$, then $G_{P_k}(x) \geqslant H(x)$, since if $x \in (t_{j-1}, t_j)$ we have $G_{P_k}(x) = \sup_{y\in(t_{j-1},t_j]} f(y) \geqslant H(x)$. Hence $G(x) \geqslant H(x)$.

If $H(x) < G(x)$, choose $a \in \mathbb{R}$ such that $H(x) < a < G(x)$. By definition of H, there exists $\delta_0 > 0$ such that if $0 < \delta < \delta_0$ we have $f(y) < a$ if $|y-x| < \delta$. But since the mesh of the partitions P_k tends to zero, for large k, $x \in (t_{j-1}, t_j)$ and $t_j - t_{j-1} < \delta$, hence

$$G_{P_k}(x) = M_j = \sup_{y\in(t_{j-1},t_j]} f(y) \leqslant a$$

Since the sequence $\{G_{P_k}\}$ is decreasing, we have $G(x) \leqslant G_{P_k}(x) \leqslant a < G(x)$, which gives us a contradiction, hence $H = G$ in $[a,b] \smallsetminus E$, therefore $H = G$ a.e. The proof that $h = g$ a.e. is similar.

It remains to prove (iii). Since G and g are measurable, $H = G, h = g$ a.e. and since m is a complete measure, H and h are measurable. Moreover,

$$\int_{[a,b]} H\,dm = \int_{[a,b]} G\,dm = \lim \int_{[a,b]} G_{P_k}\,dm = \lim S_{P_k} f = \overline{I}_a^b(f),$$

and

$$\int_{[a,b]} h\,dm = \int_{[a,b]} g\,dm = \lim \int_{[a,b]} g_{P_k}\,dm = \lim s_{P_k} f = \underline{I}_a^b(f). \qquad \square$$

We can now prove part (b) of the theorem. If f is Riemann integrable, then by Folland Exercise 2.23,

$$\int_{[a,b]} H\,dm = \int_a^b f(x)\,dx = \int_{[a,b]} h\,dm.$$

Hence $H = h$ a.e. by Proposition 39. Thus the set of discontinuity points of f has zero Lebesgue measure by Folland Exercise 2.23.

Conversely, if $D(f)$ has zero Lebesgue measure, $H = h$ a.e. by Folland Exercise 2.23(i) and Folland Exercise 2.23(ii). Then by Proposition 39 and Folland Exercise 2.23(iii), we obtain

$$\overline{I}_a^b(f) = \int_{[a,b]} H\,dm = \int_{[a,b]} h\,dm = \underline{I}_a^b(f),$$

so f is Riemann integrable.

The (proper) Riemann integral is thus subsumed in the Lebesgue integral. Some improper Riemann integrals (the absolutely convergent ones) can be interpreted directly as Lebesgue integrals, but others still require a limiting procedure. For example, if f is Riemann integrable on $[0, b]$ for all $b > 0$ and Lebesgue integrable on $[0, \infty)$, then $\int_{[0,\infty)} f\,dm = \lim_{b\to\infty} \int_0^b f(x)\,dx$ (by the DCT), but the limit on the right can exist even when f is not integrable. (Example: $f = \sum_1^\infty n^{-1}(-1)^n \chi_{(n,n+1]}$.)

Notation 54. *Henceforth we shall tend to use the notation* $\int_a^b f(x)\,dx$ *for Lebesgue integrals.*

The Lebesgue theory offers two real advantages over the Riemann theory.

(1) First, much more powerful convergence theorems, such as the monotone and DCTs, are available. These not only yield results previously unobtainable but also reduce the labor in proving classical theorems.
(2) Second, a wider class of functions can be integrated. For example, if R is the set of rational numbers in $[0, 1], \chi_R$ is not Riemann integrable, being everywhere discontinuous on $[0, 1]$, but it is Lebesgue integrable, and $\int \chi_R\,dm = 0$. (Actually, this is in some sense a trivial example since χ_R agrees a.e. with the constant function 0. For a more interesting example, see Folland Exercise 2.25.)

2.5 Modes of Convergence

If X is a measure space, one can speak of a.e. convergence or convergence in L^1. Of course, uniform convergence implies pointwise convergence, which in turn implies a.e. convergence (and not conversely, in general), but these modes of convergence do not imply L^1 convergence or vice versa. It will be useful to keep in mind the following examples on $\mathbb{R}$ (with Lebesgue measure):

$$f_n \coloneqq \tfrac{1}{n}\chi_{[0,n]}, \qquad g_n \coloneqq \chi_{(n,n+1)}, \qquad h_n \coloneqq n\chi_{[0,1/n]}, \qquad k_n = \chi_{[j/2^k,(j+1)/2^k]} \left(\begin{smallmatrix}\text{where } n = 2^k+j \\ \text{and } 0 \leqslant j < 2^k\end{smallmatrix}\right).$$

- f_n: uniformly flattens to 0 ($\searrow 0$ uniformly and in L^1)
- g_n: "floating carpet" flying into the distance ($\searrow 0$ pointwise and in L^1)
- h_n: spikes to infinity at 0 ($\searrow 0$ a.e. and in L^1)
- k_n: "wanders" around the unit interval [0,1] ($\searrow 0$ in measure and in L^1 but *not* a.e. or uniformly)

Note $f_n \to 0$ uniformly, $g_n \to 0$ pointwise, and $h_n \to 0$ a.e. (namely, everywhere except zero), but none of these converge to 0 in L^1. In fact $\int f_n = \int g_n = \int h_n = 1$ for all n for all n. But $k_n \to 0$ in L^1 since $\int |k_n| = 2^{-k}$ for $2^k \leqslant n < 2^{k+1}$, but $k_n(x)$ does not converge at *any* $x \in [0,1]$, since there are infinitely many n for which $k_n(x) = 0$ and infinitely many for which $k_n(x) = 1$.

On the other hand, if f_n is arbitrary, $f_n \to f$ a.e. and $|f_n| \leqslant g \in L^1$ for all n, then $f_n \to f$ in L^1. (This is clear from the DCT since $|f_n - f| \leqslant 2g$.) Also, we shall see below that if $f_n \to f$ in L^1 then some subsequence converges to f a.e. Another mode of convergence that is frequently useful is convergence in measure. We say that a sequence $\{f_n\}$ of measurable complex-valued functions on $(X, \mathcal{M}, \mu)$ is ***Cauchy in measure*** if for every $\varepsilon > 0$,

$$\mu(\{|f_n(x) - f_m(x)| \geqslant \varepsilon\}) \to 0 \text{ as } m, n \to \infty,$$

and that $\{f_n\}$ ***converges in measure*** to f if for every $\varepsilon > 0$,

$$\mu(\{|f_n(x) - f(x)| \geqslant \varepsilon\}) \to 0 \text{ as } n \to \infty.$$

For example, the sequences the spike f_n, h_n, and k_n above converge to zero in measure, but g_n is not Cauchy in measure.

Proposition 2.55: 2.29.

If $f_n \to f$ in L^1, then $f_n \to f$ in measure.

Proof. Let $E_{n,\varepsilon} = \{|f_n - f| \geqslant \varepsilon\}$. Then

$$\int |f_n - f| \geqslant \int_{E_{n,\varepsilon}} |f_n - f| \geqslant \varepsilon\mu(E_{n,\varepsilon}),$$

so

$$0 \leqslant \mu(E_{n,\varepsilon}) \leqslant \frac{1}{\varepsilon}\int |f_n - f| \to 0 \text{ as } n \to \infty, \qquad \square$$

so $\mu(E_{n,\varepsilon}) \to 0$ as $\nu \to \infty$.

The converse of Proposition 55 is false, as examples f_n (which uniformly flattens to 0 everywhere) and h_n (which spikes to infinity at $x = 0$) show.

Theorem 2.56: 2.30.

- A sequence $\{f_n\}_{n=1}^{\infty}$ that is Cauchy in measure converges in measure to some (measurable) function in measure, and that limit is unique up to null sets.
- If $f_n \to f$ in measure, then there exists a subsequence f_{n_j} such that $f_{n_j} \to f$ a.e.

Proof. This proof can be found here. Since $\{f_n\}$ is Cauchy in measure, we can find n_1 such that

$$\mu(\{|f_n - f_{n_1}| \geqslant 1/2\}) \leqslant 1/2 \text{ for all } n, m \geqslant n_1. \tag{2.56.1}$$

Set $g_1 \coloneqq f_{n_1}$. Likewise, we can choose $n_2 \geqslant n_1$ such that $\mu(\{|f_n - f_m| \geqslant 1/4\}) \leqslant 1/4$ for all $n, m \geqslant n_2$. Set $g_2 = f_{n_2}$ and $E_1 = \{g_1(x) - g_2| \geqslant 1/2\}$. Then by Equation (2.56.1) we have $\mu(E_1) \leqslant 2^{-1}$. Inductively we can choose $n_{j+1} \geqslant n_j, g_j = f_{n_j}$ and $E_j = \{x \in X \mid |g_j(x) - g_{j+1}(x)| \geqslant 2^{-j}\}$ with $\mu(E_j) \leqslant 2^{-j}$.

Now for each k, set $F_k = \bigcup_{j=k}^{\infty} E_j$ then $\mu(F_k) \leqslant \sum_{j=k}^{\infty} 2^{-j} = 2^{1-k}$, and for $x \notin F_k$ and $i \geqslant j \geqslant k$ we have

$$|g_j(x) - g_i(x)| \leqslant \sum_{p=j}^{i-1} |g_p(x) - g_{p+1}(x)| \leqslant \sum_{p=j}^{i-1} 2^{-p} \leqslant 2^{1-j}, \tag{2.56.2}$$

and thus $\{g_j\}$ is pointwise Cauchy on F_k^c. If $F = \bigcap_{k=1}^{\infty} F_k = \limsup E_j$, then $\mu(F) = \lim_{j\to\infty} \mu(E_j) = 0$, and $\{g_j\}$ is pointwise Cauchy on F^c. Set $f(x) = \lim g_j(x)$ for $x \in F^c$ and $f(x) = 0$ for $x \in F$ (by Folland Exercise 2.3,Folland Exercise 2.5, f is measurable). Hence $g_j \to f$ a.e.

Using Equation (2.56.2) and sending $\to \infty$ for each $x \in F_k^c$, we have $|g_j(x) - f(x)| \leqslant 2^{1-j}$ and since $\mu(F_k) \to 0$ as $k \to \infty, g_j \to f$ in measure. Now

$$\{x \in X \mid |f_n(x) - f(x)| \geqslant \varepsilon\} \supset \{x \in X \mid |f_n(x) - g_j(x)| \geqslant \varepsilon/2\} \cup \{x \in X \mid |g_j(x) - f(x)| \geqslant \varepsilon/2\}.$$

and thus $f_n \to f$ in measure, since the measure of both sets on the right side converge to zero as $n, j \to \infty$. Now assume that $f_n \to g$ in measure and fix $k \in \mathbb{Z}_{\geqslant 0}$. We have

$$\{x \in X | f(x) - g(x)| \geqslant k^{-1}\} \subset \{x \in X \mid |f(x) - f_n(x)| \geqslant k^{-1}/2\} \cup \{x \in X \mid |f_n(x) - g(x)| \geqslant k^{-1}/2\}$$

for all n, and making $n \to \infty$ we obtain $\mu(\{x \in X \mid |f(x) - g(x)| \geqslant k^{-1}\}) = 0$. Thus, since $\{x \in X \mid f(x) \neq g(x)\} = \bigcap_{n=1}^{\infty} \{x \in X \mid |f(x) - g(x)| \geqslant k^{-1}\}$, we have $\mu(\{x \in X \mid f(x) \neq g(x)\}) = 0$ and hence $f = g$ a.e. □

Exercise 2.57: Folland Exercise 2.32.

Suppose $\mu(X) < \infty$. If $f, g\colon X \to \mathbb{C}$ are measurable functions. Define

$$\rho(f, g) = \int u(|f - g|)\, d\mu,$$

where $u\colon \mathbb{R} \to \mathbb{R}$ via $u(x) = x/(1+x)$. Then ρ is a metric on the space of measurable functions on X where we identify functions f and g if $f = g$ a.e., and $f_n \to f$ with respect to ρ if and only if $f_n \to f$ in measure.

Proof. We only show $f_n \to f$ with respect to ρ if and only if $f_n \to f$ in measure. Suppose $\rho(f_n, f) \to 0$, $\varepsilon > 0$, and $E = \{|f_n - f| \geqslant \varepsilon\}$. Then $|f_n - f| \geqslant \varepsilon \implies u(\varepsilon) \leqslant u(|f_n - f|)$ since u is increasing, so

$$u(\varepsilon)\mu(E) = \int_E u(\varepsilon) \leqslant \int_E u(|f_n - f|) = \rho(f_n, f) \to 0,$$

so $\mu(E) \to 0$ as $n \to \infty$. Hence $f_n \to f$ in measure.

Conversely, suppose $f_n \to f$ in measure. Since $|f_n - f| \geqslant u(|f_n - f|)$, we can write $u(|f_n - f|) \geqslant \varepsilon \implies |f_n - f| \geqslant \varepsilon$, which implies $\{u(|f_n - f|) \geqslant \varepsilon\} \subset \{|f_n - f| \geqslant \varepsilon\}$, so

$$\mu(\{u(|f_n - f|) \geqslant \varepsilon\}) \leqslant \mu(E).$$

$$\begin{aligned}
\rho(f_n, f) &= \int u(|f_n - f|) \\
&= \int_{\{u(|f_n - f|) < \varepsilon\}} u(|f_n - f|) + \int_{\{u(|f_n - f|) \geqslant \varepsilon\}} u(|f_n - f|) \\
&\leqslant \varepsilon\mu(X) + \int_{\{u(|f_n - f|) \geqslant \varepsilon\}} \frac{\varepsilon}{1 + |f_n - f|} \\
&\leqslant \varepsilon\mu(X) + \underbrace{\varepsilon\mu(\{u(|f_n - f|) \geqslant \varepsilon\})}_{\substack{\to 0 \text{ as } n \to \infty, \text{ since} \\ f_n \to f \text{ in measure}}}. \\
&\leqslant \varepsilon\mu(X) + \varepsilon\mu(E).
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, we conclude $\rho(f_n, f) \to 0$ as $n \to \infty$. □

Exercise 2.58: Folland Exercise 2.33.

If $f_n \in \mathbb{Z}_{\geqslant 0}$ and $f_n \to f$ in measure, then $\int f \leqslant \liminf_{n\to\infty} \int f_n$

Solution. The limit infimum of a sequence is a limit point of that sequence by definition, so there exists a subsequence f_{n_k} of f_n such that $\int f_{n_k} \to \liminf \int f_n$. Then $f_{n_k} \to f$ in measure, so there is a subsubsequence $f_{n_{k_j}} \to f$ a.e. Since $f \geqslant 0$, by Fatou's Lemma,

$$\int f = \int \liminf_{j\to\infty} f_{n_{k_j}} \leqslant \liminf_{j\to\infty} \int f_{n_{k_j}} = \liminf_{j\to\infty} \int f_n. \qquad \square$$

In general, if $f_n \to f$ a.e. does *not* imply $f_n \to f$ in measure. However, if $\mu(X) < \infty$, then this does hold.

Definition 59. *If $\{f\colon X \to \mathbb{C}\}$ is a sequence of measurable functions such that for all $\varepsilon > 0$, there exists $E \subset X$ such that $\mu(E) < \varepsilon$ and $f_n \to f$ uniformly on $\mu(E^c)$, then we say $f_n \to f$* ***almost uniformly****.*

Exercise 2.60: Folland Exercise 2.39.

If $f_n \to f$ almost uniformly, then $f_n \to f$ a.e. and in measure.

Solution.

- $f_n \to f$ *in measure*: Let $\varepsilon > 0$. Since $f_n \to f$ almost uniformly, there exists $E \in \mathcal{M}$ such that $\mu(E) < \varepsilon$ and $f_n \to f$ uniformly on E^c. This means given $\varepsilon' > 0$, $|f_n - f| < \varepsilon'$ for all sufficiently large n. But this means $\{|f_n - f| \geqslant \varepsilon\} \subset E$, so $\mu(\{|f_n - f| \geqslant \varepsilon\}) \leqslant \mu(E) < \varepsilon$. Thus $f_n \to f$ in measure.
- $f_n \to f$ *a.e.*: Let $\varepsilon > 0$ and let $F = \{x \in X \mid f_n(x) \nrightarrow f(x)\}$. Since $f_n \to f$ almost uniformly, there exists $E \in \mathcal{M}$ such that $\mu(E) < \varepsilon$ and $f_n \to f$ uniformly on E^c. Since uniform convergence implies a.e. convergence, $E^c \subset F^c$. Thus $F \subset E$. Hence $\mu(F) \subset \mu(E) < \varepsilon$. Since ε was arbitrary and independent of F, we conclude $\mu(F) = 0$. Thus $f_n \to f$ a.e. □

Theorem 2.61: 2.33: Egoroff's Theorem.

If $\mu(X) < \infty$ and $\{f_n : X \to \mathbb{C}\}_{n=1}^{\infty}$ is a sequence of measurable functions converging a.e. to f, then $f_n \to f$ almost uniformly.

Proof. This proof can be found here. Assume first that $f_n \to f$ pointwise on X. For $k, n \in \mathbb{Z}_{\geqslant 0}$ define

$$E_n(k) = \bigcup_{j=n}^{\infty} \{x \in X \mid |f_j(x) - f(x)| \geqslant k^{-1}\}$$

If k is fixed, then $\{E_n(k)\}_n$ is a decreasing sequence and since $f_j(x) \to f(x)$ as $j \to \infty$ for each $x \in X$, we have $\bigcap_{n=1}^{\infty} E_n(k) = \varnothing$. Since $\mu(X) < \infty$, from the continuity from above, we have $\mu(E_n(k)) \to 0$ as $n \to \infty$. Given $\varepsilon > 0$ and $k \in \mathbb{Z}_{\geqslant 0}$, choose n_k such that $\mu(E_{n_k}(k)) < \varepsilon 2^{-k}$ and let $E = \bigcup_{k=1}^{\infty} E_{n_k}(k)$. Then $\mu(E) < \varepsilon$ and $|f_n(x) - f(x)| < k^{-1}$ for $n \geqslant n_k$ and $x \notin E$. Thus $f_n \to f$ uniformly on E^c.

Now if $f_n \to f$ a.e., let $F \subset X$ be the set with $\mu(F) = 0$ such that $f_n \to f$ everywhere on F^c. Thus, from the previous result (with F^c instead of X), given $\varepsilon > 0$ there exists a set $E \subset F^c$ with $\mu(E) < \varepsilon$ and $f_n \to f$ uniformly on E^c. Thus taking $A = E \cup F$ then $\mu(A) = \mu(E) + \mu(F) = \mu(E) < \varepsilon$ and $A^c = E^c \cap F^c = E^c$, hence $f_n \to f$ uniformly on A^c. □

Exercise 2.62: Folland Exercise 2.40: Strengthened Egoroff's Theorem.

If $\{f_n : X \to \mathbb{C}\}_{n=1}^{\infty}$ is a sequence of measurable functions such that $f_n \to f$ a.e. and there exists $g \in L^1 \cap L^+$ such that $|f_n| \leqslant g$ for all n, then $f_n \to f$ almost uniformly.

Solution. This proof can be found here. From the DCT $\int |f| \leqslant \int g$ and $f \in L^1(\mu)$. As in the proof of Egoroff's Theorem, we can assume without loss of generality that $f_n \to f$

pointwise, and we set

$$E_n(k) = \bigcup_{j=n}^{\infty} \{x \in X \mid |f_j(x) - f(x)| \geqslant k^{-1}\}.$$

If we can prove that $\mu(E_1(k)) < \infty$ for all k, then as in the proof of Egoroff's Theorem, it will follow that $\mu(E_n(k)) \to 0$ as $n \to \infty$ and the rest of the proof remains unchanged. Now, if $x \in E_1(k)$ then there exists $j \in \mathbb{Z}_{\geqslant 0}$ such that $|f_j(x) - f(x)| \geqslant k^{-1}$. Hence

$$\begin{aligned}\mu(E_1(k)) = \int \chi_{E_1(k)} d\mu = \int_{E_1(k)} d\mu \leqslant k \int_{E_1(k)} |f_j(x) - f(x)| d\mu \\ \leqslant k \int_{E_1(k)} (|f_j(x)| + |f(x)|) d\mu \leqslant 2k \int_{E_1(k)} g \, d\mu \leqslant 2k \int g \, d\mu < \infty,\end{aligned}$$

since $g \in L^1(\mu)$. Therefore the result follows. □

The following corollary is then immediate.

Corollary 2.63.

If $\{f_n\}_{n=1}^{\infty}$ is a sequence of measurable functions and there exists $g \in L^1 \cap L^+$ such that $|f_n| \leqslant g$ a.e. and $f_n \to f$ a.e., then $f_n \to f$ in measure.

Exercise 2.64: Folland Exercise 2.38.

If $\{f_n\}_{n=1}^{\infty}$ and $\{g_n\}_{n=1}^{\infty}$ are sequences of measurable functions such that $f_n \to f$ and $g_n \to g$ in measure, then the following hold.

(1) $|f_n| \to |f|$ in measure.
(2) $f_n + g_n \to f + g$ in measure.
(3) $f_n^2 \to f^2$ in measure if $\mu(X) < \infty$.
(4) $f_n g_n \to fg$ in measure if $\mu(X) < \infty$, but not necessarily if $\mu(X) = \infty$.

Solution.

(1) Fix $\varepsilon, \varepsilon' > 0$. We want some N such that for all $n \geqslant N$, $\mu(\{||f_n| - |f|| \geqslant \varepsilon\}) < \varepsilon'$. Since $f_n \to f$ in measure, there exists N_0 such that for all $n \geqslant N_0$,

$$\mu(\{|f_n - f| \geqslant \varepsilon\}) < \varepsilon'.$$

Since $|f_n - f| \geqslant ||f_n| - |f||$, if $||f_n(x)| - |f(x)|| \geqslant \varepsilon$ then $|f_n(x) - f(x)| \geqslant \varepsilon$. Hence $\{||f_n| - |f|| \geqslant \varepsilon\} \subset \{|f_n - f| \geqslant \varepsilon\}$, so $\mu(\{||f_n| - |f|| \geqslant \varepsilon\}) \leqslant \mu(\{|f_n - f| \geqslant \varepsilon\})$. Since the left-hand side of this inequality is nonnegative and the right-hand side vanishes as $n \to \infty$, the left-hand side also vanishes, so $|f_n| \to |f|$ in measure.

(2) Fix $\varepsilon, \varepsilon' > 0$. Since $|f_n + g_n - f - g| \leqslant |f_n - f| + |g_n - g|$, we can write $(|f_n + g_n - f - g| \geqslant \varepsilon) \subset \{|f_n - f| \geqslant \varepsilon\}$ and $\{|g_n - g| \geqslant \varepsilon\}$. Thus for all sufficiently large n,

$$\begin{aligned}\mu(\{|f_n + g_n - f - g| \geqslant \varepsilon\}) \leqslant \mu(\{|f_n - f| \geqslant \varepsilon\}) + \mu(\{|g_n - g| \geqslant \varepsilon\}) \\ < \varepsilon'/2 + \varepsilon'/2 = \varepsilon'.\end{aligned}$$

(3) To see this, first suppose $f_n \to f$ in measure and let $\varepsilon, \varepsilon', M > 0$. Now

$$\begin{aligned}\mu\{|f_n^2 - f^2| \geqslant \varepsilon\} &= \mu\{|f_n - f| \cdot |f_n + f| \geqslant \varepsilon\} \\ &= \mu\{|f_n - f| \cdot |f_n + f| \geqslant \varepsilon \text{ and } |f_n + f| \geqslant M\} \\ &\quad + \mu\{|f_n - f| \cdot |f_n + f| \geqslant \varepsilon \text{ and } |f_n + f| < M\} \\ &\qquad\qquad \text{(by disjont additivity of } \mu) \\ &\leqslant \mu\{|f_n - f| \cdot |f_n + f| \geqslant \varepsilon \text{ and } |f_n + f| \geqslant M\} + \mu\{|f_n - f| \geqslant \varepsilon/M\},\end{aligned}$$

where the last step is by monotonicity of μ because $|f_n + f| \leqslant M$ and $|f_n - f| \cdot |f_n + f| \geqslant \varepsilon$ together imply $k|f_n - f| \geqslant \varepsilon$, and hence that $|f_n - f| \geqslant \varepsilon/M$. Since $f_n \to f$ in measure, there exists N_0 such that for all $n \geqslant N_0$, $\mu\{|f_n - f| \geqslant \varepsilon/M\} < \varepsilon'/2$. We now claim that there exists N_1 such that for all $n \geqslant N_1$, $\mu\{|f_n - f| \cdot |f_n + f| \geqslant \varepsilon \text{ and } |f_n + f| \geqslant M\} < \varepsilon'/2$, which will complete the proof since then by the above we can conclude $\mu\{|f_n^2 - f| \geqslant \varepsilon\} < \varepsilon'$ for all $n \geqslant \max\{N_0, N_1\}$. To see why such an N_1 exists, note that since f is complex valued,

$$\bigcap_{M=1}^{\infty} \{|f_n - f| \geqslant M\} = \varnothing.$$

Thus

$$0 = \mu\Big(\bigcap_{M=1}^{\infty} \{|f_n + f| \geqslant M\}\Big) = \lim_{M\to\infty} \mu(\underbrace{\{|f_n - f| \geqslant M\}}_{:= E_{n,M}})$$

where we used continuity of μ from above (noting there is no issue with any set here being measurable), as

$$\mu(E_{n,1}) = \mu\{|f_n - f| \geqslant 1\} \leqslant \mu(X) < \infty.$$

Then in particular, there exists an N_1 such that for all $n \geqslant N_1$, $\mu(E_{n,M}) < \varepsilon'/2$. Since

$$\mu\{|f_n - f|, |f_n + f| \geqslant \varepsilon \text{ and } |f_n + f| \geqslant M\} \subset E_{n,M},$$

we conclude by monotonicity of μ that

$$\mu(\{|f_n - f| \cdot |f_n + f| \geqslant \varepsilon \text{ and } |f_n + f|\}) \geqslant M\} \leqslant \mu(E_{n,M}) < \varepsilon'/2,$$

which proves $f_n^2 \to f$ in measure, and hence completes the proof by our previous remarks.

(4) We first provide a counterexample to the statement of (3) in the case $\mu(X) = \infty$. Consider the Lebesgue measure space $(\mathbb{R}, \mathcal{L}, m)$ and the nonnegative measurable functions $f_n, g_n \colon \mathbb{R} \to \mathbb{R}$ given by $f_n(x) = x + 1/n$ and $g_n(x) = x$. Then the f_n, g_n are nonnegative measurable functions, $f_n(x), g_n(x) \to x$ in measure (so in this case $f(x) = g(x) = x$, $f(x)g(x) \to x^2$ in measure, but $f_n(x)g_n(x) = x^2 + x/n$, which does not converge to x^2 in measure: indeed, $|x^2 + x/n - x^2| = |x/n| > \varepsilon'$ for all x with $|x| \geqslant n\varepsilon'$; since the Lebesgue measure of all such x is infinite, $f_n g_n$ does not converge to fg in measure.

Now instead assume $\mu(X) < \infty$. By writing

$$fg = \frac{1}{2}((f+g)^2 - f^2 - g^2),$$

we observe that

$$\begin{aligned} |f_n g_n - fg| &= \left|\frac{1}{2}((f_n+g_n)^2 - f_n^2 - g_n^2) - \frac{1}{2}((f+g)^2 - f^2 - g^2)\right| \\ &\leqslant \frac{1}{2}|(f_n+g_n)^2 - (f+g)^2| + \frac{1}{2}|f_n^2 - f^2| + \frac{1}{2}|g_n^2 - g^2|. \end{aligned}$$

Since $f_n + g_n \to f + g$ in measure by (2), it suffices to show that $f_n \to f$ in measure implies $f_n^2 \to f^2$ in measure. But this is (3), so we are done. □

Exercise 2.65: Folland Exercise 2.34.

If $|f_n| \leqslant g \in L^1$ and $f_n \to f$ in measure, then
(a) $\int f = \lim_{n\to\infty} \int f_n$ and
(b) $f_n \to f$ in L^1.

Solution.

(a) Suppose $\int f_n \nrightarrow \int f$. Then there exists $\varepsilon > 0$ and a subsequence $\{\int f_{n_k}\}_{k=1}^{\infty}$ such that

$$\int |f_{n_k} - f| \geqslant \left|\int f_n - \int f\right| \geqslant \varepsilon. \tag{2.65.1}$$

But $f_n \to f$ in measure, so $f_{n_k} \to f$ in measure, so there exists a subsequence $\{f_{n_{k_\ell}}\}_{\ell=1}^{\infty}$ converging a.e to f. But $|f_{n_{k_\ell}}| \leqslant g \in L^1$ for all ℓ, so by the DCT $f \in L^1$ and $\int f_{n_{k_\ell}} \to \int f$, contradicting Equation (2.65.1).

(b) By Folland Exercise 2.21, it suffices to show $\int |f_n| \to \int |f|$. Since $f_n \to f$ in measure, $|f_n| \to |f|$ in measure by Folland Exercise 2.38. Then by part (a), $\int |f_n| \to \int |f|$, as desired. □

2.6 Product Measures

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be measure spaces. We define a ***(measurable) rectangle*** to be a set of the form $A \times B$, where $A \in \mathcal{M}$ and $B \in \mathcal{N}$. Note that

$$(A \times B) \cap (E \times F) = (A \cap E) \times (B \cap F) \quad \text{and} \quad (A \times B)^c = (X \times B^c) \cup (A^c \times B),$$

so by Proposition 7 the collection $\mathcal{A}$ of finite disjoint unions of rectangles is an algebra. The σ-algebra it generates is $\mathcal{M} \otimes \mathcal{N}$.

Suppose $A \times B$ is a rectangle that is a countable disjoint union of rectangles $\{A_j \times B_j\}_{j=1}^{\infty} \subset \mathcal{M} \times \mathcal{N}$. Then for any $x \in X$ and $y \in Y$,

$$\chi_A(x)\chi_B(y) = \chi_{A\times B}(x,y) = \sum_{j=1}^{\infty} \chi_{A_j \times B_j}(x,y) = \sum_{j=1}^{\infty} \chi_{A_j}(x)\chi_{B_j}(y).$$

If we integrate with respect to x and apply the MCT for series (Theorem 26), we obtain

$$\mu(A)\chi_B(y) = \int \chi_A(x)\chi_B(y)\,d\mu(x) = \sum_{j=1}^{\infty}\int \chi_{A_j}(x)\chi_{B_j}(y)\,d\mu(x) = \sum_{j=1}^{\infty}\mu(A_j)\chi_{B_j}(y).$$

Similarly, integration with respect to y yields

$$\mu(A)\nu(B) = \sum_{j=1}^{\infty}\mu(A_j)\nu(B_j).$$

It follows that if $E \in \mathcal{A}$ is the finite disjoint union of rectangles $A_1 \times B_1, \ldots, A_n \times B_n$, and we define

$$\pi(E) \coloneqq \sum_{j=1}^{n}\mu(A_j)\nu(E_j)$$

(with the usual convention that $0 \cdot \infty = 0$), then π is well-defined on $\mathcal{A}$ since any two representations of E as a finite disjoint union of rectangles have a common refinement, and π is a premeasure on $\mathcal{A}$. According to Theorem 33, therefore, π generates an outer measure on $X \times Y$ whose restriction to $\mathcal{M} \times \mathcal{N}$ is a measure that extends π. We call this measure the ***product of μ and ν*** and denote it by $\mu \times \nu$. Moreover, if μ and ν are σ-finite—say, $X = \bigcup_1^\infty A_j$ and $Y = \bigcup_1^\infty B_k$ with $\mu(A_j), \nu(B_k) < \infty$—then $X \times Y = \bigcup_{j,k=1}^\infty A_j \times B_k$ and $\mu \times \nu(A_j \times B_k) < \infty$, so $\mu \times \nu$ is also σ-finite. In this case, by Theorem 33, $\mu \times \nu$ is the unique measure on $\mathcal{M} \otimes \mathcal{N}$ such that $\mu \times \nu(A \times B) = \mu(A)\nu(B)$ for all rectangles $A \times B$.

The same construction works for any finite number of factors. That is, suppose $(X_j, \mathcal{M}_j, \mu_j)$ are measure spaces for $j = 1, \ldots, n$. If we define a rectangle to be a set of the form $A_1 \times \cdots \times A_n$ with $A_j \in \mathcal{M}_j$, then the collection $\mathcal{A}$ of finite disjoint unions of rectangles is an algebra, and the same procedure as above produces a measure $\mu_1 \times \cdots \times \mu_n$ on $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ such that

$$\mu_1 \times \cdots \times \mu_n(A_1 \times \cdots \times A_n) = \prod_1^n \mu_j(A_j).$$

Moreover, if the μ_js are σ-finite so that the extension from $\mathcal{A}$ to $\bigotimes_{j=1}^n \mathcal{M}_j$ is uniquely determined, the obvious associativity properties hold. For example, if we identify $X_1 \times X_2 \times X_3$ with $(X_1 \times X_2) \times X_3$, we have $\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3 = (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3$ (the former being generated by sets of the form $A_1 \times A_2 \times A_3$ with $A_j \in \mathcal{M}_j$, and the latter by sets of the form $B \times A_3$ with $B \in \mathcal{M}_1 \otimes \mathcal{M}_2$ and $A_3 \in \mathcal{M}_3$), and $\mu_1 \times \mu_2 \times \mu_3 = (\mu_1 \times \mu_2) \times \mu_3$ (since they agree on sets of the form $A_1 \times A_2 \times A_3$, and hence in general by uniqueness). Details are left to the reader (Folland Exercise 2.45). All of our results below have obvious extensions to products with n factors, but we shall stick to the case $n = 2$ for simplicity.

Definition 66. *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be any two measure spaces. For any $E \subset X \times Y$ and any $x \in X$ and $y \in Y$ we define the **x-section of E**, and **y-section of E**, denoted E_x and E^y, respectively, by*

$$E_x = \{y \in Y \mid (x, y) \in E\} \qquad \text{and} \qquad E^y = \{x \in X \mid (x, y) \in E\}.$$

*Also, if f is a function on $X \times Y$ we define the **x-section of f** and the **y-section of f**, denoted f_x and f^y, respectively, by*

$$f_x(y) = f^y(x) = f(x, y).$$

Example 67. *We can write* $(\chi_E)_x = \chi_{E_x}$ *and* $(\chi_E)^y = \chi_{E^y}$.

Proposition 2.68: 2.34.

(a) If $E \in \mathcal{M} \otimes \mathcal{N}$, then $E_x \in \mathcal{N}$ for all $x \in X$ and $E^y \in \mathcal{M}$ for all $y \in Y$.
(b) If f is $\mathcal{M} \otimes \mathcal{N}$-measurable, then f_x is $\mathcal{N}$-measurable for all $x \in X$ and f^y is $\mathcal{M}$-measurable for all $y \in Y$.

Proof. Let $\mathcal{R}$ be the collection of all subsets E of $X \times Y$ such that $E_x \in \mathcal{N}$ for all x and $E^y \in \mathcal{M}$ for all y. Then $\mathcal{R}$ obviously contains all rectangles (e.g., $(A \times B)_x = B$ if $x \in A$, and $(A \times B)_x = \varnothing$ otherwise). Since $(\bigcup_1^\infty E_j)_x = \bigcup_1^\infty (E_j)_x$ and $(E^c)_x = (E_x)^c$, and likewise for y-sections, $\mathcal{R}$ is a σ-algebra. Therefore $\mathcal{R} \supset \mathcal{M} \otimes \mathcal{N}$, which proves (a). (b) follows from (a) because $(f_x)^{-1}(B) = (f^{-1}(B))_x$ and $(f^y)^{-1}(B) = (f^{-1}(B))^y$. □

Before proceeding further we need a technical lemma. We define a ***monotone class*** on a space X to be a subset $\mathcal{C}$ of $\mathcal{P}(X)$ that is closed under countable increasing unions and countable decreasing intersections (that is, if $E_j \in \mathcal{C}$ and $E_1 \subset E_2 \subset \cdots$, then $\bigcup E \in \mathcal{C}$, and likewise for intersections). Clearly every σ-algebra is a monotone class. Also, the intersection of any family of monotone classes is a monotone class, so for any $\mathcal{P} \subset \mathcal{P}(X)$ there is a unique smallest monotone class containing $\mathcal{P}$, called the ***monotone class generated by $\mathcal{P}$***.

Theorem 2.69: 2.35: The Monotone Class Lemma.

If $\mathcal{A}$ is an algebra of subsets of X, then the monotone class $\mathcal{C}$ generated by $\mathcal{A}$ coincides with the σ-algebra $\mathcal{M}$ generated by $\mathcal{A}$.

Proof. Since $\mathcal{M}$ is a monotone class, we have $\mathcal{C} \subset \mathcal{M}$; and if we can show that $\mathcal{C}$ is a σ-algebra, we will have $\mathcal{M} \subset \mathcal{C}$. To this end, for $E \in \mathcal{C}$ let us define

$$\mathcal{C}(E) = \{F \in \mathcal{C} \mid E \setminus F, F \setminus E, \text{ and } E \cap F \text{ are in } \mathcal{C}\}.$$

Clearly $\varnothing$ and E are in $\mathcal{C}(E)$, and $E \in \mathcal{C}(F)$ if and only if $F \in \mathcal{C}(E)$. Also, it is easy to check that $\mathcal{C}(E)$ is a monotone class. If $E \in \mathcal{A}$, then $F \in \mathcal{C}(E)$ for all $F \in \mathcal{A}$ because $\mathcal{A}$ is an algebra; that is, $\mathcal{A} \subset \mathcal{C}(E)$, and hence $\mathcal{C} \subset \mathcal{C}(E)$. Therefore, if $F \in \mathcal{C}$, then $F \in \mathcal{C}(E)$ for all $E \in \mathcal{A}$. But this means that $E \in \mathcal{C}(F)$ for all $E \in \mathcal{A}$, so that $\mathcal{A} \subset \mathcal{C}(F)$ and hence $\mathcal{C} \subset \mathcal{C}(F)$. Conclusion: If $E, F \in \mathcal{C}$, then $E \setminus F$ and $E \cap F$ are in $\mathcal{C}$. Since $X \in \mathcal{A} \subset \mathcal{C}$, $\mathcal{C}$ is therefore an algebra. But then if $\{E_j\}_1^\infty \subset \mathcal{C}$, we have $\bigcup_1^n E_j \in \mathcal{C}$ for all n, and since $\mathcal{C}$ is closed under countable increasing unions it follows that $\bigcup_1^\infty E_j \in \mathcal{C}$. In short, $\mathcal{C}$ is a σ-algebra, and we are done. □

We now come to the main results of this section, which relates integrals on $X \times Y$ to integrals on X and Y.

Theorem 2.70: 2.36.

Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces. If $E \in \mathcal{M} \otimes \mathcal{N}$, then the functions $x \mapsto \nu(E_x)$ and $y \mapsto \mu(E^y)$ are measurable on X and Y, respectively, and

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

Proof.

- *Case 1: μ and ν are finite.* Let $\mathcal{C}$ be the set of all $E \in \mathcal{M} \otimes \mathcal{N}$ for which the conclusions of the theorem are true. If $E = A \times B$, then $\nu(E_x) = \chi_A(x)\nu(B)$ and $\mu(E^y) = \mu(A)\chi_B(y)$, so $E \in \mathcal{C}$. By countable additivity it follows that finite disjoint unions of rectangles are in $\mathcal{C}$, so by Theorem 69 it will suffice to show that $\mathcal{C}$ is a monotone class. If $\{E_n\}$ is an increasing sequence in $\mathcal{C}$ and $E = \bigcup_1^\infty E_n$, then the functions $f_n(y) = \mu((E_n)^y)$ are measurable and increase pointwise to $f(y) = \mu(E^y)$. Hence f is measurable, so by the MCT,
$$\int \mu(E^y)\, d\nu(y) = \lim_{n\to\infty} \int \mu((E_n)^y) d\nu(y) = \lim_{n\to\infty} \mu \times \nu(E_n) = \mu \times \nu(E).$$
Likewise $\mu \times \nu(E) = \int \nu(E_x) d\mu(x)$, so $E \in \mathcal{C}$. Similarly, if $\{E_n\}$ is a decreasing sequence in $\mathcal{C}$ and $\bigcap_1^\infty E_n$, the function $y \mapsto \mu((E_1)^y)$ is in $L^1(\nu)$ because $\mu((E_1)^y) \leqslant \mu(X) < \infty$ and $\nu(Y) < \infty$, so the DCT can be applied to show that $E \in \mathcal{C}$. Thus $\mathcal{C}$ is a monotone class, and the proof is complete for the case of finite measure spaces.
- *Case 2: μ and ν are σ-finite.* Then we can write $X \times Y$ as the union of an increasing sequence $\{X_j \times Y_j\}$ of rectangles of finite measure. If $E \in \mathcal{M} \otimes \mathcal{N}$, the preceding argument applies to $E \cap (X_j \times Y_j)$ for each j to give
$$\mu \times \nu(E \cap (X_j \times Y_j)) = \int \chi_{X_j}(x)\nu(E_x \cap Y_j) d\mu(x) = \int \chi_{Y_j}(y)\mu(E^y \cap X_j) d\nu(y),$$
and a final application of the MCT then yields the desired result. □

Theorem 2.71: 2.37: The Fubini-Tonelli Theorem.

Suppose that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces.

(a) (Tonelli) If $f \in L^+(X \times Y)$, then the functions $g(x) = \int f_x\, d\nu$ and $h(y) = \int f^y\, d\mu$ are in $L^+(X)$ and $L^+(Y)$, respectively, and
$$\int f\, d(\mu \times \nu) = \int \left[\int f(x,y)\, d\nu(y) \right] d\mu(x) = \int \left[\int f(x,y)\, d\mu(x) \right] d\nu(y). \quad (2.71.1)$$

(b) (Fubini) If $f \in L^1(\mu \times \nu)$, then $f_x \in L^1(\nu)$ for a.e. $x \in X$,, $f^y \in L^1(\mu)$ for a.e. $y \in Y$, the a.e-defined functions $g(x) = \int f_x\, d\nu$ and $h(x) = \int f^y\, d\nu$ are in $L^1(\mu)$ and $L^1(\nu)$, respectively, and Equation (2.71.1) holds.

Proof. Tonelli's theorem reduces to Theorem 70 in case f is a characteristic function, and it therefore holds for nonnegative simple functions by linearity. If $f \in L^+(X \times Y)$, let $\{f_n\}$ be a sequence of simple functions that increase pointwise to f as in Theorem 18. The MCT implies, first, that the corresponding g_n and h_n increase to g and h (so that g and h are measurable), and, second that

$$\int g\,d\mu = \lim \int g_n\,d\mu = \lim \int f_n d(\mu \times \nu) = \int f d(\mu \times \nu),$$
$$\int h\,d\nu = \lim \int h_n\,d\nu = \lim \int f_n d(\mu \times \nu) = \int f d(\mu \times \nu),$$

which is Equation (2.71.1). This establishes Tonelli's theorem and also shows that if $f \in L^+(X \times Y)$ and $\int f d(\mu \times \nu) < \infty$, then $g < \infty$ a.e. and $h < \infty$ a.e., that is, $f_x \in L^1(\nu)$ for a.e. x and $f^y \in L^1(\mu)$ for a.e. y. If $f \in L^1(\mu \times \nu)$, then, the conclusion of Fubini's theorem follows by applying these results to the positive and negative parts of the real and imaginary parts of f. □

Notation 72. *We shall usually omit the brackets in the iterated integrals in Equation* (2.71.1)*, thus:*

$$\int \left[\int f(x,y) d\mu(x) \right] d\nu(y) = \iint f(x,y) d\mu(x) d\nu(y) = \iint f\,d\mu\,d\nu.$$

A few remarks are in order:

- The hypothesis of σ-finiteness is necessary; see Folland Exercise 2.46.
- The hypothesis $f \in L^+(X \times Y)$ or $f \in L^1(\mu \times \nu)$ is necessary in the following two ways.
  - (i) First, it is possible for f_x and f^y to be measurable for all x, y and for the iterated integrals $\iint f\,d\mu\,d\nu$ and $\iint f\,d\nu\,d\mu$ to exist even if f is not $\mathcal{M} \otimes \mathcal{N}$-measurable. However, the iterated integrals need not then be equal; see Folland Exercise 2.47.
  - (ii) Second, if f is not nonnegative, it is possible for f_x and f^y to be integrable for all x, y and for the iterated integrals $\iint f\,d\mu\,d\nu$ and $\iint f\,d\nu\,d\mu$ to exist even if $\int |f| d(\mu \times \nu) = \infty$. But again, the iterated integrals need not be equal; see Folland Exercise 2.48
- The Fubini and Tonelli theorems are frequently used in tandem. Typically one wishes to reverse the order of integration in a double integral $\iint f\,d\mu\,d\nu$. First one verifies that $\int |f| d(\mu \times \nu) < \infty$ by using Tonelli's theorem to evaluate this integral as an iterated integral; then one applies Fubini's theorem to conclude that $\iint f\,d\mu\,d\nu = \iint f\,d\nu\,d\mu$. For examples, see the exercises in Folland Section 2.6.
- Even if μ and ν are complete, $\mu \times \nu$ is almost never complete. Indeed, suppose that there is a nonempty $A \in \mathcal{M}$ with $\mu(A) = 0$ and that $\mathcal{N} \neq \mathcal{P}(Y)$. (This is the case with $\mu = \nu =$ Lebesgue measure on $\mathbb{R}$, for example.) If $E \in \mathcal{P}(Y) \setminus \mathcal{N}$, then $A \times E \notin \mathcal{M} \otimes \mathcal{N}$ by Proposition 68, but $A \times E \subset A \times Y$, and $\mu \times \nu(A \times Y) = 0$.

Notice that we did not assume $\nu(Y) < \infty$, hence, we employed the convention that $0 \cdot (\pm\infty) = 0$ in measure theory (unless stated otherwise). In fact, this convention needs to be taken in order to call $\mu \times \nu$ a measure as the following elementary observation regarding continuity from below shows:

Consider $X = Y = \mathbb{R}$ and the Lebesgue measure, m, on each space. Then in the above example, let $E = \{0\}$ and choose any nonmeasurable subset of $[-1, 1]$ (note that this subset will not be a subset of a null set since m is complete). Then we clearly have that

$$E \times N \subset E \times [-n, n] \subset E \times \mathbb{R}, \text{ for each } n \in \mathbb{Z}_{\geqslant 0}.$$

As each $E \times [-n, n]$ is a rectangle, we have $m \times m(E \times [-n, n]) = 0 \cdot (2n) = 0$ for each $n \in \mathbb{Z}_{\geqslant 0}$. Furthermore, $E_n = E \times [-n, n]$ is a increasing, nested set such that $\cup_{n \in \mathbb{Z}_{\geqslant 0}} E_n = E \times \mathbb{R}$, and continuity from below applied to the measure $m \times m = \lambda$ shows that we must have

$$\lambda(E \times \mathbb{R}) = \lambda(\cup_{n \in \mathbb{Z}_{\geqslant 0}} E_n) = \lim_{n \to \infty} \lambda(E_n) = \lim_{n \to \infty} m \times m(E \times [-n, n]) = \lim_{n \to \infty} 0 = 0.$$

As continuity from below holds for all measures, if we took any other convention than $0 \cdot (\pm\infty) = 0$ in this setting, the product measure would not be a measure!

Although the σ-finite assumption is needed for Tonelli (see Folland Exercise 2.46), it turns out that is *not* needed for Fubini (see Tao's Remark 1.7.22 in his notes on measure theory). But by Proposition 32 the support of an L^1 function is σ-finite, so there is no harm in assuming σ-finiteness in general. (But this is not true in general for $f \in L^+$ only.) Nevertheless, the hypotheses $f \in L^+$ or $f \in L^1$ is needed (see Folland Exercise 2.47,Folland Exercise 2.48).

If one wishes to work with complete measures, of course, one can consider the completion of $\mu \times \nu$. In this setting the relationship between the measurability of a function on $X \times Y$ and the measurability of its x-sections and y-sections is not so simple. However, the Fubini-Tonelli theorem is still valid when suitably reformulated:

Theorem 2.73: 2.39: The Fubini-Tonelli Theorem for Complete Measures.

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be complete, σ-finite measure spaces, and let $(X \times Y, \mathcal{L}, \lambda)$ be the completion of $(X \times Y, \mathcal{M} \otimes \mathcal{N}, \mu \times \nu)$. If f is $\mathcal{L}$-measurable and either (a) $f \geqslant 0$ or (b) $f \in L^1(\lambda)$, then f_x is $\mathcal{N}$-measurable for a.e. x and f^y is $\mathcal{M}$-measurable for a.e. y, and in case $(b) f_x$ and f^y are also integrable for a.e. x and y. Moreover, $x \mapsto \int f_x \, d\nu$ and $y \mapsto \int f^y \, d\mu$ are measurable, and in case (b) also integrable, and

$$\int f \, d\lambda = \iint f(x, y) \, d\mu(x) \, d\nu(y) = \iint f(x, y) \, d\nu(y) \, d\mu(x).$$

This theorem is a fairly easy corollary of Theorem 71; the proof is outlined in Folland Exercise 2.49.

Exercise 2.74: Folland Exercise 2.45.

If $(X_j, \mathcal{M}_j)$ is a measurable space for $j = 1,2,3$, then $\otimes_1^3 \mathcal{M}_j = (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3$. Moreover, if μ_j is a σ-finite measure on $(X_j, \mathcal{M}_j)$, then $\mu_1 \times \mu_2 \times \mu_3 = (\mu_1 \times \mu_2) \times \mu_3$.

Exercise 2.75: Folland Exercise 2.46.

Let $X = Y = [0,1], \mathcal{M} = \mathcal{N} = \mathcal{B}_{[0,1]}$, $\mu =$ Lebesgue measure, and $\nu =$ counting measure. If $D = \{(x,x) \mid x \in [0,1]\}$ is the diagonal in $X \times Y$, then $\iint \chi_D \, d\mu \, d\nu$, $\iint \chi_D \, d\nu \, d\mu$, and $\int \chi_D d(\mu \times \nu)$ are all unequal. (To compute $\int \chi_D d(\mu \times \nu) = \mu \times \nu(D)$, go back to the definition of $\mu \times \nu$.)

Solution. First we note that D is measurable. Indeed, given $n \in \mathbb{Z}_{\geqslant 0}$, define $I_{n,k} = \left[\frac{k}{n}, \frac{k+1}{n}\right]$ for $k = 0, \cdots, n-1$ and $E_n = \bigcup_{k=0}^{n-1} (I_{n,k} \times I_{n,k})$. Thus $D = \bigcap_{n=1}^{\infty} E_n \in \mathcal{M} \otimes \mathcal{M}$. We have

$$\iint \chi_D d\mu d\nu = \int_{[0,1]} \int_0^1 \chi_D(x,y) dx d\nu(y),$$

but for each fixed $y \in [0,1]$, we have $\chi_D(x,y) = 0$ if $x \neq y$ and $\chi_D(x,y) = 1$ if $x = y$, hence $\chi_D(\cdot, y) = 0 \mu$-a.e., and thus

$$\iint \chi_D d\mu d\nu = \int_{[0,1]} 0 d\nu(y) = 0 \cdot \nu([0,1]) = 0 \cdot \infty = 0.$$

Now since $\nu(\{x\}) = 1$ for each $x \in [0,1]$ we have

$$\iint \chi_D d\nu d\mu = \int_0^1 \int_{[0,1]} \chi_D(x,y) d\nu(y) dx = \int_0^1 1 dx = 1.$$

To compute $\mu \times \nu(D)$, we will use the outer measure π^*, the restriction of which to $\mathcal{M} \otimes \mathcal{N}$ we recall is the definition of $\mu \times \nu$. Assume that $D \subset \bigcup_{j=1}^{\infty} (A_j \times B_j)$ where $A_j, B_j \in \mathcal{B}_{[0,1]}$ for all j. Since $D = \bigcup_{j=1}^{\infty} D \cap (A_j \times B_j)$, then given $x \in [0,1]$ we have $(x,x) \in (A_j \times B_j)$ for some j, that is, $x \in A_j \cap B_j$, and hence $\bigcup_{j=1}^{\infty} A_j \cap B_j = [0,1]$. Therefore there exists $j \in \mathbb{Z}_{\geqslant 0}$ such that $\mu(A_j \cap B_j) > 0$, thus $\mu(A_j) \geqslant \mu(A_j \cap B_j) > 0$ and $\mu(B_j) \geqslant \mu(A_j \cap B_j) > 0$, and in particular, $\nu(B_j) = \infty$ (since if $\nu(B_j) < \infty$ implies that $\mu(B_j) = 0$). Hence $\mu \times \nu(A_j \times B_j) = \infty$, and thus $\sum_{j=1}^{\infty} \mu \times \nu(A_j \times B_j) = \infty$. Since this is true for any cover of D by rectangles, we have $\mu \times \nu(D) = \infty$. □

Exercise 2.76: Folland Exercise 2.47.

Let $X = Y$ be an uncountable linearly ordered set such that for each $x \in X$, $\{y \in X \mid y < x\}$ is countable. (Example: the set of countable ordinals.) Let $\mathcal{M} = \mathcal{N}$ be the σ-algebra of countable or co-countable sets, and let $\mu = \nu$ be defined on $\mathcal{M}$ by $\mu(A) = 0$ if A is countable and $\mu(A) = 1$ if A is co-countable. Let $E = \{(x,y) \in X \times X \mid y < x\}$. Then E_x and E^y are measurable for all x, y, and $\iint \chi_E \, d\mu \, d\nu$ and $\iint \chi_E \, d\nu \, d\mu$ exist but are not equal. (If one believes in the continuum hypothesis, one can take $X = [0,1]$ [with a nonstandard ordering] and thus obtain a set $E \subset [0,1]^2$ such that E_x is countable and E^y is co-countable [in particular, Borel] for all x, y, but E is not Lebesgue measurable.)

Exercise 2.77: Folland Exercise 2.48.

Let $X = Y = \mathbb{Z}_{\geqslant 0}, \mathcal{M} = \mathcal{N} = \mathcal{P}(\mathbb{Z}_{\geqslant 0}), \mu = \nu =$ counting measure. Define $f(m,n) = 1$ if $m = n$, $f(m,n) = -1$ if $m = n+1$, and $f(m,n) = 0$ otherwise. Then $\int |f| \, d(\mu \times \nu) = \infty$, and $\iint f \, d\mu \, d\nu$ and $\iint f \, d\nu \, d\mu$ exist and are unequal.

Exercise 2.78: Folland Exercise 2.49.

Prove Theorem 73 by using Theorem 71 and Proposition 22, together with the following lemmas.

(a) If $E \in \mathcal{M} \otimes \mathcal{N}$ and $\mu \times \nu(E) = 0$, then $\nu(E_x) = \mu(E^y) = 0$ for a.e. x and y.

(b) If f is $\mathcal{L}$-measurable and $f = 0$ λ-a.e., then f_x and f^y are integrable for a.e. x and y, and $\int f_x \, d\nu = \int f^y \, d\mu = 0$ for a.e. x and y. (Here the completeness of μ and ν is needed.)

Exercise 2.79: Folland Exercise 2.50.

Suppose $(X, \mathcal{M}, \mu)$ is a σ-finite measure space and $f \in L^+(X)$. Let

$$G_f = \{(x,y) \in X \times [0,\infty] \mid y \leqslant f(x)\}.$$

Then G_f is $\mathcal{M} \times \mathcal{B}_{\mathbb{R}}$-measurable and $\mu \times m(G_f) = \int f \, d\mu$. The same is also true if the inequality $y \leqslant f(x)$ in the definition of G_f is replaced by $y < f(x)$. (To show measurability of G_f, note that the map $(x,y) \mapsto f(x) - y$ is the composition of $(x,y) \mapsto (f(x), y)$ and $(z,y) \mapsto z - y$.)

This is the definitive statement of the familiar theorem from calculus which states the integral of a function is the area under its graph.

Solution. G_f is measurable because the map $(x,y) \mapsto (f(x),y) \mapsto f(x)-y$ is a composition of measurable functions, under which G_f is the preimage of the measurable set $[0,\infty)$ (or $(0,\infty)$ in the case of '$<$' instead of '$\leqslant$'). Then $\chi_{G_f} \in L^+(\mu \times m)$, so by Tonelli's Theorem we have

$$\mu \times m(G_f) = \int \chi_{G_f} \, d(\mu \times m) = \int_X \left(\int_0^\infty \chi_{G_g}^x(y) \, dy \right) d\mu(x) = \int_X f(x) \, d\mu(x),$$

as desired. □

Exercise 2.80: Folland Exercise 2.51.

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be arbitrary measure spaces (not necessarily σ-finite).

(a) If $f : X \to \mathbb{C}$ is $\mathcal{M}$-measurable, $g : Y \to \mathbb{C}$ is $\mathcal{N}$-measurable, and $h(x, y) = f(x)g(y)$, then h is $\mathcal{M} \otimes \mathcal{N}$-measurable.

(b) If $f \in L^1(\mu)$ and $g \in L^1(\nu)$, then $h \in L^1(\mu \times \nu)$, and

$$\int h\, d(\mu \times \nu) = \left(\int f\, d\mu\right)\left(\int g\, d\nu\right).$$

Solution.

(a) h is the composition $(x, y) \mapsto (f(x), g(y)) \mapsto f(x)g(y)$, which is a composition of measurable functions, and hence measurable.

(b) If $f = \chi_A$ and $g = \chi_B$ for $A \in \mathcal{M}$ and $B \in \mathcal{M}$, then $h = \chi_A \times \chi_B = \chi_{A \times B}$, so

$$\begin{aligned}\int h\, d(\mu \times \nu) &= \mu \times \nu(A \times B)\\ &= \mu(A) \times \nu(B) = \left(\int \chi_A\, d\mu\right)\left(\int \chi_B\, d\nu\right) = \left(\int f\, d\mu\right)\left(\int g\, d\nu\right).\end{aligned}$$

where we used that $\mu \times \nu(A \times B) = \mu(A) \times \nu(B)$ by definition of the product measure.

Now suppose f is a simple nonnegative function $f = \sum_{i=1}^n c_i \chi_{A_i}$ and $g = \chi_B$. Then $h = \sum_{i=1}^n c_i \chi_{A_i \times B}$ and thus

$$\int h d(\mu \times \nu) = \sum_{i=1}^n c_i \mu \times \nu(A_i \times B) = \sum_{i=1}^n c_i \mu(A_i)\nu(B) = \left(\int f d\mu\right)\left(\int g d\nu\right).$$

Now suppose th f and g are $\mathcal{M}$ and $\mathcal{M}$ measurable (respectively) simple nonnegative functions. Then $g = \sum_{k=1}^m d_k \chi_{B_k}$ and we set $h_k = d_k f \chi_{B_k}$ for each $k = 1, \cdots, m$. Then by the previous case, we have

$$\int h_k d(\mu \times \nu) = d_k \int f \chi_{B_k} d(\mu \times \nu) = d_k \left(\int f d\mu\right)\left(\int \chi_{B_k} d\nu\right) = \left(\int f\right)\left(\int d_k \chi_{B_k} d\nu\right).$$

Summing over $k \in \{1, \ldots, m\}$, we obtain the result.

Now suppose $f \in L^+(X)$, $g \in L^+(Y)$. Then there exist sequences $\{s_n\}$ and $\{r_n\}$ of nonnegative simple $\mathcal{M}$ and $\mathcal{M}$-measurable functions that increase pointwise to f and g respectively. Then $h_n := f_n g_n \in L^+(X \times Y)$ as a product of measurable functions, and $\{h_n\}$ increases to fg. Then by the MCT,

$$\int h d(\mu \times \nu) = \lim_{n \to \infty} \int h_n d(\mu \times \nu) = \lim_{n \to \infty} \left(\int f_n d\mu\right)\left(\int g_n d\mu\right) = \left(\int f d\mu\right)\left(\int g d\nu\right).$$

For $f \in L^1(\mu)$ and $g \in L^1(\nu)$ real functions, the result follows by applying the previous case to $f^+g^+, f^+g^-, f^-g^+$and f^-g^-. For complex functions, just apply the real L^1 case to $\operatorname{Re} f \operatorname{Re} g, \operatorname{Re} f \operatorname{Im} g, \operatorname{Im} f \operatorname{Re} g$ and $\operatorname{Im} f \operatorname{Im} g$. □

Exercise 2.81: Folland Exercise 2.52.

The Fubini-Tonelli theorem is valid when $(X, \mathcal{M}, \mu)$ is an arbitrary measure space and Y is a countable set, $\mathcal{N} = \mathcal{P}(Y)$, and ν is counting measure on Y. (See Theorems 26 and 48.)

2.7 The n-Dimensional Lebesgue Integral

The ***n-dimensional Lebesgue measure*** m^n on $\mathbb{R}^n$ is the completion of the n-fold product of Lebesgue measure on $\mathbb{R}$ with itself, that is, the completion of $m \times \cdots \times m$ on $\mathcal{B}_\mathbb{R} \otimes \cdots \otimes \mathcal{B}_\mathbb{R} = \mathcal{B}_\mathbb{R}^n$, or equivalently the completion of $m \times \cdots \times m$ on $\mathcal{L} \otimes \cdots \otimes \mathcal{L}$ of m^n is the class of Lebesgue measurable sets in $\mathbb{R}^n$; we will denote this complete σ-algebra by $\mathcal{L}^n$. (Sometimes we shall also consider m^n as a measure on the smaller domain $\mathcal{L}_{\mathbb{R}^n}$.) When there is no danger of confusion, we shall usually omit the superscript n and write m for m^n, and as in the case $n = 1$, we shall usually write $\int f(x)\,dx$ for $\int f\,dm$.

We begin by establishing the extensions of some of the results in Folland Section 1.5 to the n-dimensional case. In what follows, if $E = \prod_1^n E_j$ is a rectangle in $\mathbb{R}^n$, we shall refer to the sets $E_j \subset \mathbb{R}$ as the ***sides*** of E.

Theorem 2.82: 2.40.

Suppose $E \in \mathcal{L}^n$.

(a) We can write

$$m(E) = \inf\{m(U) \mid U \text{ is an open set containing } E\}$$
$$= \sup\{m(K) \mid K \text{ is a compact set inside } E\}.$$

(b) $E = A_1 \cup N_1 = A_2 \setminus N_2$ where A_1 is an F_σ set, A_2 is a G_δ set, and $m(N_1) = m(N_2) = 0$.

(c) If $m(E) < \infty$, for any $\varepsilon > 0$ there is a finite collection $\{R_j\}_1^N$ of disjoint rectangles whose sides are intervals such that $m(E \triangle \bigcup_1^N R_j) < \varepsilon$.

Proof. By the definition of product measures, if $E \in \mathcal{L}^n$ and $\varepsilon > 0$ there is a countable family $\{T_j\}$ of rectangles such that $E \subset \bigcup_1^\infty T_j$ and $\sum_1^\infty m(T_j) \leqslant m(E) + \varepsilon$. For each j, by applying Theorem 45 to the sides of R_j we can find a rectangle $U_j \supset F_j$ whose sides are open sets such that $m(U_j) < m(T_j) + \varepsilon 2^{-j}$. If $U = \bigcup_1^\infty U_j$, then U is open and $m(U) \leqslant \sum_1^\infty m(U_j) \leqslant m(E) + 2\varepsilon$. This proves the first equation in part (a); the second one, and part (b), then follow as in the proofs of Theorems 45 and 46. Next, if $m(E) < \infty$, then $m(U_j) < \infty$ for all j. Since the sides of U_j are countable unions of open intervals, by taking suitable finite subunions we obtain rectangles $V_j \subset U_j$ whose sides are finite unions of intervals such that $m(V_j) \geqslant m(U_j) - \varepsilon 2^{-j}$. If N is sufficiently large, then, we have

$$m\Big(E \setminus \bigcup_1^N V_j\Big) \leqslant m\Big(\bigcup_1^N U_j \setminus V_j\Big) + m\Big(\bigcup_{N+1}^\infty U_j\Big) < 2\varepsilon$$

and

$$m\Big(\bigcup_1^N V_j \smallsetminus E\Big) \leqslant m\Big(\bigcup_1^\infty U_j \smallsetminus E\Big) < \varepsilon,$$

so that $m(E \triangle \bigcup_1^N V_j) < 3\varepsilon$. Since $\bigcup_1^N V_j$ can be expressed as a finite disjoint union of rectangles whose sides are intervals, we have proved (c). □

Theorem 2.83: 2.41.

If $f \in L^1(m)$ and $\varepsilon > 0$, there is a simple function $\phi = \sum_1^N a_j \chi_{R_j}$, where each R_j is a product of intervals, such that $\int |f - \phi| < \varepsilon$, and there is a continuous function g that vanishes outside a bounded set such that $\int |f - g| < \varepsilon$.

Proof. We first sketch the argument we will give. As in the proof of Theorem 49, we can approximate f by simple functions. We will then use Theorem 82(c) to approximate the latter by functions ϕ of the desired form. Finally, we will approximate such ϕs by continuous functions by applying an obvious generalization of the argument in the proof of Theorem 49.

We now give the argument. Fix $\varepsilon > 0$. By Theorem 18, there exists a sequence of simple functions $\{\phi_j\}_{j=1}^\infty$ such that $|\phi_j| \nearrow |f|$ and $\phi_j \to f$ pointwise as $j \to \infty$. Since $|\phi_j - f| \leqslant |\phi_j| + |f| \leqslant 2|f| \in L^1(m)$ for each j, we can apply the dominated convergence theorem to obtain $\lim_{j\to\infty} \int |\phi_j - f| = \int 0 = 0$. Thus there exists a simple function ϕ_0 such that $|\phi_0| \leqslant |f|$ and $\int |f - \phi_0| < \varepsilon/2$. Write ϕ_0 in its standard representation $\phi_0 = \sum_{j=1}^n a_j \chi_{E_j}$, where $a_j \in \mathbb{C}$ and the $E_j \in \mathcal{M}$ are disjoint. We may assume $a_j \neq 0$, since it makes no difference in the proof, as will be seen. First note that the E_j are m-finite, because

$$\sum_{j=1}^n |a_j| \mu(E_j) = \int |\phi_0| \leqslant \int |f| + \int |\phi_0 - f| < \varepsilon/2 + \int |f| < +\infty.$$

Then by Theorem 82(c), for each E_j there exists a finite disjoint union of measurable rectangles $F_j = \bigcup_{k=1}^{N_j} R_k^j$, where the R_j are have interval sides, such that $m(E_j \triangle F_j) < \frac{\varepsilon}{2|a_j| n}$. Note $\chi_{F_j} = \sum_{k=1}^{N_j} \chi_{R_k^j}$ since the R_k^j are disjoint. Now define

$$\phi = \sum_{j=1}^n a_j \chi_{F_j} = \sum_{j=1}^n \sum_{k=1}^{N_j} a_j \chi_{R_k^j} = \sum_{j,k} a_j \chi_{R_j^k}.$$

Then ϕ is a simple function, and

$$\int |\phi_0 - \phi| = \int \Big| \sum_{j=1}^n a_j \chi_{E_j} - \sum_{j=1}^n a_j \chi_{F_j} \Big| \leqslant \sum_{j=1}^n |a_j| \int |\chi_{E_j} - \chi_{F_j}|.$$

$$\overset{(*)}{=} \sum_{j=1}^n |a_j| \int \chi_{E_j \triangle F_j} = \sum_{j=1}^n |a_j| m(E_j \triangle F_j) < \sum_{j=1}^n \frac{|a_j| \varepsilon}{2|a_j| n} = \frac{\varepsilon}{2},$$

where the equality at $(*)$ is because for any subsets $A, B \subset X$, we have $|\chi_A - \chi_B| = \chi_{A \triangle B}$. (This can be seen by showing equality on each of $A \smallsetminus B$, $B \smallsetminus A$, $A \cap B$, and $(A \cup B)^c$. On $A \smallsetminus B$, we have $|\chi_A - \chi_B| = |1 - 0| = 1 = \chi_{A \triangle B}$. On $B \smallsetminus A$, we have $|\chi_A - \chi_B| =$

$|0-1| = 1 = \chi_{A \triangle B}$. On $A \cap B$, we have we have $|\chi_A - \chi_B| = |1-1| = 0 = \chi_{A\triangle B}$. And on $(A \cup B)^c$, we have $|\chi_A - \chi_B| = |0-0| = 0 = \chi_{A\triangle B}$.) We can now write

$$\int |f - \phi| \leqslant \int |f - \phi_0| + \int |\phi_0 - \phi| < \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

so ϕ is the desired simple function.

For the second point of the statement, we use the fact we have just proved. Since there exists a simple function $\phi = \sum_{k=1}^N a_k \chi_{R_k}$ such that each R_k is a rectangle with interval sides, $a_k \neq 0$, and $\int |\phi - f| < \varepsilon/2$. Now, for each index $1 \leqslant k \leqslant N$, let α_k be a continuous map satisfying $|\chi_{R_k} - \alpha_k| < \varepsilon/(2|a_k|N)$. (Recall from basic analysis of a single variable functions that such a map α_k exists, as we can connect the intervals on which χ_{R_k} is constant by steeper and steeper slopes.) Then the function g defined by $g = \sum_{k=1}^N a_k \alpha_k$. Then g is continuous as a finite linear combination of continuous functions, and we can write

$$\int |f - g| \leqslant \int |f - \phi| + \int |\phi - g| < \varepsilon/2 + \int \Big| \sum_{k=1}^N a_k \chi_{R_k} - \sum_{k=1}^N a_k \alpha_k \Big|$$
$$\leqslant \frac{\varepsilon}{2} + \sum_1^N |a_k| \int |\chi_{R_k} - \alpha_k| \leqslant \frac{\varepsilon}{2} + \sum_{k=1}^N \frac{|a_k|\varepsilon}{2|a_k|N} = \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

so g is the desired continuous function. □

Theorem 2.84: 2.42.

The Lebesgue measure is translation-invariant. More precisely, for $a \in \mathbb{R}^n$ define $\tau_a \colon \mathbb{R}^n \to \mathbb{R}^n$ by $\tau_a(x) = x + a$.

(a) If $E \in \mathcal{L}^n$, then $\tau_a(E) \in \mathcal{L}^n$ and $m(\tau_a(E)) = m(E)$.

(b) If $f \colon \mathbb{R}^n \to \mathbb{C}$ is Lebesgue measurable, then so is $f \circ \tau_\alpha$. Moreover, if either $f \geqslant 0$ or $f \in L^1(m)$, then $\int (f \circ \tau_a)\, dm = \int f\, dm$.

Proof. Since τ_a and its inverse τ_{-a} are continuous, they preserve the class of Borel sets. The formula $m(\tau_a(E)) = m(E)$ follows easily from the one-dimensional result (Theorem 48) if E is a rectangle, and it then follows for general Borel sets since m is determined by its action on rectangles (the uniqueness in Theorem 33). In particular, the collection of Borel sets E such that $m(E) = 0$ is invariant under τ_a. Assertion (a) now follows immediately.

If f is Lebesgue measurable and B is a Borel set in $\mathbb{C}$, we have $f^{-1}(B) = E \cup N$ where E is Borel and $m(N) = 0$. But $\tau_a^{-1}(E)$ is Borel and $m(\tau_a^{-1}(N)) = 0$, so $(f \circ \tau_a)^{-1}(B) \in \mathcal{L}^n$ and f is Lebesgue measurable. The equality $\int (f \circ \tau_a) d\mu = \int f\, d\mu$ reduces to the equality $m(\tau_{-a}(E)) = m(E)$ when $f = \chi_E$. It is then true for simple functions by linearity, and hence for nonnegative measurable functions by the definition of the integral. Taking positive and negative parts of real and imaginary parts then yields the result for $f \in L^1(m)$. □

Let us now compare Lebesgue measure on $\mathbb{R}^n$ to the more naïve theory of n-dimensional volume usually found in advanced calculus books. In this discussion, a cube in $\mathbb{R}^n$ is a

Cartesian product of n closed intervals whose side lengths are all equal.

Given $k \in \mathbb{Z}$, let Q_k be the collection of cubes of side length $1/2^k$ with vertices in the lattice $\left(\frac{1}{2^k}\mathbb{Z}\right)^n$. (That is, $\prod_1^n [a_j, b_j] \in Q_k$ if and only if $2^k a_j$ and $2^k b_j$ are integers and $b_j - a_j = 2^{-k}$ for all j.) Note that any two cubes in Q_k have disjoint interiors, and that the cubes in Q_{k+1} are obtained from the cubes in Q_k by bisecting the sides.

If $E \subset \mathbb{R}^n$, we define the ***inner approximation*** and ***outer approximation*** to E by the grid of cubes Q_k to be

$$\underline{A}(E,k) = \bigcup\{Q \in Q_k \mid Q \subset E\} \quad \text{and} \quad \overline{A}(E,k) = \bigcup\{Q \in Q_k \mid Q \cap E \neq \varnothing\},$$

respectively. The measure of $\underline{A}(E,k)$ (in either the naïve geometric sense or the Lebesgue sense) is just 2^{-nk} times the number of cubes in Q_k that lie in $\underline{A}(E,k)$, and we denote it by $m(\underline{A}(E,k))$; likewise for $m(\overline{A}(E,k))$. Also, the sets $\underline{A}(E,k)$ increase with k while the sets $\overline{A}(E,k)$ decrease, because each cube in Q_k is a union of cubes in Q_{k+1}. Hence the limits

$$\underline{\kappa}(E) = \lim_{k\to\infty} m(\underline{A}(E,k)), \quad \overline{\kappa}(E) = \lim_{k\to\infty} m(\overline{A}(E,k))$$

exist. They are called the ***inner content*** and ***outer content*** of E, and if they are equal, their common value $\kappa(E)$ is the ***Jordan content*** of E.

Remark 85. *We make two remarks. First, Jordan content is usually defined using general rectangles whose sides are intervals rather than our dyadic cubes, but the result is the same. Second, although all the definitions above make sense for arbitrary $E \subset \mathbb{R}^n$, the theory of Jordan content is meaningful only if E is bounded, for otherwise $\overline{\kappa}(E)$ always equals ∞.*

Let

$$\underline{A}(E) = \bigcup_1^\infty \underline{A}(E,k) \qquad \text{and} \qquad \overline{A}(E) = \bigcap_1^\infty \overline{A}(E,k).$$

Then $\underline{A}(E) \subset E \subset \overline{A}(E), \underline{A}(E)$ and $\overline{A}(E)$ are Borel sets, and $\underline{\kappa}(E) = m(\underline{A}(E))$ and $\overline{\kappa}(E) = m(\overline{A}(E))$. Thus the Jordan content of E exists if and only if $m(\overline{A}(E) \setminus \underline{A}(E)) = 0$, which implies that E is Lebesgue measurable and $m(E) = \kappa(E)$.

To clarify further the relationship between Lebesgue measure and the approximation process leading to Jordan content, we establish the following lemma. (The second part of the lemma will be used later.)

Lemma 2.86: 2.43.

If $U \subset \mathbb{R}^n$ is open, then $U = \underline{A}(U)$. Moreover, U is a countable union of cubes with disjoint interiors.

Proof. If $x \in U$, let $\delta = \inf\{|y - x| \mid y \notin U\}$, which is positive since U is open. If Q is a cube in Q_k that contains x, then every $y \in Q$ is at a distance at most $2^{-k}\sqrt{n}$ from x (the worst case being when $|x_j - y_j| = 2^{-k}$ for all j), so we will have $Q \subset U$ provided k is large enough so that $2^{-k}\sqrt{n} < \delta$. But then $x \in \underline{A}(U,k) \subset \underline{A}(U)$.

This shows that $\underline{A}(U) = U$, and the second assertion follows by writing $\underline{A}(U) = \underline{A}(U,0) \cup \bigcup_1^\infty [\underline{A}(U,k) \setminus \underline{A}(U,k-1)] \cdot \underline{A}(U,0)$ is a (countable) union of cubes in Q_0, and for $k \geqslant 1$, the closure of $A(U,k) \setminus \underline{A}(U,k-1)$ is a (countable) union of cubes in Ω_k. These cubes all have disjoint interiors, and the result follows. □

Lemma 86 immediately implies that the Lebesgue measure of any open set is equal to its inner content. On the other hand, suppose that $F \subset \mathbb{R}^n$ is compact. We can find a large cube, say $Q_0 = \{x \mid \max|x_j| \leqslant 2^M\}$, whose interior $\operatorname{int}(Q_0)$ contains F. If $Q \in Q_k$ and $Q \subset Q_0$ then either $Q \cap F \neq \varnothing$ or $Q \subset (Q_0 \setminus F)$, so $m(\overline{A}(F,k)) + m(\underline{A}(Q_0 \setminus F, k)) = m(Q_0)$. Letting $k \to \infty$, we see that $\overline{\kappa}(F) + \kappa(Q_0 \setminus F) = m(Q_0)$. But $Q_0 \setminus F$ is the union of the open set int $(Q_0) \setminus F$ and the boundary of Q_0, which has content zero, so that $\underline{\kappa}(Q_0 \setminus F) = \underline{\kappa}(\operatorname{lnt}(Q_0) \setminus F) = m(Q_0 \setminus F)$. It follows that the Lebesgue measure of any compact set is equal to its outer content.

Combining these results with Theorem 82(a), we can see exactly how Lebesgue measure compares to Jordan content. The Jordan content of E is defined by approximating E from the inside and the outside by finite unions of cubes. The Lebesgue measure of E, on the other hand, is given by a two-step approximation process: First one approximates E from the outside by open sets and from the inside by compact sets, and then approximates the open sets from the inside and the compact sets from the outside by finite unions of cubes. The Lebesgue measurable sets are precisely those for which these outer-inner and inner-outer approximations give the same answer in the limit. (See Folland Exercise 1.19).

We now investigate the behavior of the Lebesgue integral under linear transformations. We identify a linear map $T\colon \mathbb{R}^n \to \mathbb{R}^n$ with the matrix $(T_{ij}) = (e_i \cdot Te_j)$, where $\{e_j\}$ is the standard ordered basis for $\mathbb{R}^n$. We denote the determinant of this matrix by $\det T$ and recall that $\det(T \circ S) = (\det T)(\det S)$. Furthermore, we employ the standard notation $\mathrm{GL}_n(\mathbb{R})$ (the "general linear" group) for the group of invertible linear transformations of $\mathbb{R}^n$. We shall need the fact from elementary linear algebra that every $T \in \mathrm{GL}_n(\mathbb{R})$ can be written as the product of finitely many transformations of the three elementary linear maps given by

$$\begin{aligned} T_1(x_1,\ldots,x_j,\ldots,x_n) &= (x_1,\ldots,cx_j,\ldots,x_n) && (c \neq 0), \\ T_2(x_1,\ldots,x_j,\ldots,x_n) &= (x_1,\ldots,x_j + cx_k,\ldots,x_n), && (k \neq j), \\ T_3(x_1,\ldots,x_j,\ldots,x_k,\ldots,x_n) &= (x_1,\ldots,x_k,\ldots,x_j,\ldots,x_n). \end{aligned}$$

That every invertible transformation is a product of transformations of these three types is simply the fact that every nonsingular matrix can be row-reduced to the identity matrix.

Theorem 2.87: 2.44.

Suppose $T \in \mathrm{GL}_n(\mathbb{R})$.

(a) If f is a Lebesgue measurable function on $\mathbb{R}^n$, so is $f \circ T$. If $f \geqslant 0$ or $f \in L^1(m)$, then

$$\int f(x)\,dx = |\det T| \int f \circ T(x)\,dx. \tag{2.87.1}$$

(b) If $E \in \mathcal{L}^n$, then $T(E) \in \mathcal{L}^n$ and $m(T(E)) = |\det T| m(E)$.

First suppose that f is Borel measurable. Then $f \circ T$ is Borel measurable since T is continuous. If Equation (2.87.1) is true for the transformations T and S, it is also true for $T \circ S$, since

$$\int f(x)\,dx = |\det T| \int f \circ T(x)\,dx = |\det T||\det S| \int (f \circ T) \circ S(x)\,dx$$
$$= |\det(T \circ S)| \int f \circ (T \circ S)(x)\,dx.$$

Hence is suffices to prove Equation (2.87.1) when T is of the types T_1, T_2, T_3 described above. But this is a simple consequence of the Fubini-Tonelli theorem. For T_3 we interchange the order of integration in the variables x_j and x_k, and for T_1 and T_2 we integrate first with respect to x_j and use the one-dimensional formulas

$$\int f(t)\,dt = |c| \int f(ct)\,dt, \qquad \int f(t+a)\,dt = \int f(t)\,dt,$$

which follow from Theorem 48. Since it is easily verified that $\det T_1 = c, \det T_2 = 1$, and $\det T_3 = -1$, Equation (2.87.1) is proved. Moreover, if E is a Borel set, so is $T(E)$ (since T^{-1} is continuous), and by taking $f = \chi_{T(E)}$, we obtain $m(T(E)) = |\det T| m(E)$. In particular, the class of Borel null sets is invariant under T and T^{-1}, and hence so is $\mathcal{L}^n$. The result for Lebesgue measurable functions and sets now follows as in the proof of Theorem 84.

Corollary 2.88: 2.46.

The Lebesgue measure is invariant under rotations.

Proof. Rotations are linear maps satisfying $TT^* = I$ where T^* is the transpose of T. Since $\det T = \det T^*$, this condition implies that $|\det T| = 1$. □

Next we shall generalize Theorem 87 to differentiable maps. Let $G = (g_1, \ldots, g_n)$ be a map from an open set $\Omega \subset \mathbb{R}^n$ into $\mathbb{R}^n$ whose components g_j are of class C^1, i.e., have continuous first-order partial derivatives. We denote by $D_x G$ the linear map defined by the matrix $((\partial g_i/\partial x_j)(x))$ of partial derivatives at x. (Observe that if G is linear, then $D_x G = G$ for all x.) G is called a ***C^1-diffeomorphism*** if G is injective and $D_x G$ is invertible for all $x \in \Omega$. In this case, the inverse function theorem guarantees that $G^{-1}\colon G(\Omega) \to \Omega$ is also a C^1 diffeomorphism and that $D_x(G^{-1}) = [D_{G^{-1}(x)} G]^{-1}$ for all $x \in G(\Omega)$.

Theorem 2.89: 2.47.

Suppose that Ω is an open set in $\mathbb{R}^n$ and $G\colon \Omega \to \mathbb{R}^n$ is a C^1 diffeomorphism.

(a) If f is a Lebesgue measurable function on $G(\Omega)$, then $f \circ G$ is Lebesgue measurable on $G(\Omega)$, and

$$\int_{G(\Omega)} f(x)\,dx = \int_{\Omega} f \circ G(x) |\det D_x G|\,dx.$$

(b) If $E \subset \Omega$ and $E \in \mathcal{L}^n$, then $G(E) \in \mathcal{L}^n$ and $m(G(E)) = \int_E |\det D_x G|\,dx$.

Proof. It suffices to consider Borel measurable functions and sets. Since G and G^{-1} are both continuous, there are no measurability problems in this case, and the general case follows as in the proof of Theorem 84. A bit of notation: For $x \in \mathbb{R}^n$ and $T = (T_{ij}) \in \mathrm{GL}_n(\mathbb{R})$, we set

$$\|x\| = \max_{1 \leqslant j \leqslant n} |x_j|, \quad \|T\| = \max_{1 \leqslant i \leqslant n} \sum\nolimits_{j=1}^n |T_{ij}|$$

We then have $\|Tx\| \leqslant \|T\|\|x\|$, and $\{x \mid \|x - a\| \leqslant h\}$ is the cube of side length $2h$ centered at a.

Let Q be a cube in Ω, say $Q = \{x \mid \|x - a\| \leqslant h\}$. By the mean value theorem, $g_j(x) - g_j(a) = \sum_j (x_j - a_j)(\partial g/\partial x_j)(y)$ for some y on the line segment joining x and a, so that for $x \in Q$, $\|G(x) - G(a)\| \leqslant h(\sup_{y \in Q}\|D_y G\|)$. In other words, $G(Q)$ is contained in a cube of side length $\sup_{y\in Q}\|D_y G\|$ times that of Q, so that by Theorem 87, $m(G(Q)) \leqslant (\sup_{y\in Q}\|D_y G\|)^n m(Q)$. If $T \in \mathrm{GL}_n(\mathbb{R})$, we can apply this formula with G replaced by $T^{-1} \circ G$ together with Theorem 87 to obtain

$$m(G(Q)) = |\det T| m(T^{-1}(G(Q))) \leqslant |\det T| (\sup_{y\in Q}\|T^{-1} D_y G\|)^n m(Q).$$

Since $D_y G$ is continuous in y, for any $\varepsilon > 0$ we can choose $\delta > 0$ so that $\|(D_z G)^{-1} D_y G\|^n \leqslant 1 + \varepsilon$ if $y, z \in Q$ and $\|y - z\| \leqslant \delta$. Let us now subdivide Q into subcubes $Q_1, \dots Q_N$ whose interiors are disjoint, whose side lengths are at most δ, and whose centers are $x_1, \dots x_N$. Applying Section 2.7 with Q replaced by Q_j and with $T = D_{x_j} G$, we obtain

$$\begin{aligned} m(G(Q)) &\leqslant \sum\nolimits_1^N m(G(Q_j)) \\ &\leqslant \sum\nolimits_{j=1}^N |\det D_{x_j} G| (\sup_{y\in Q_j}\|(D_{x_j} G)^{-1} D_y G\|)^n m(Q_j) \\ &\leqslant (1+\varepsilon) \sum\nolimits_1^N |\det D_{x_j} G| m(Q_j) \end{aligned}$$

This last sum is the integral of $\sum_1^N |\det D_{x_j} G| \chi_{Q_j}(x)$, which tends uniformly on Q to $|\det D_x G|$ as $\delta \to 0$ since $D_x G$ is continuous. Thus, letting $\delta \to 0$ and $\varepsilon \to 0$, we find that

$$m(G(Q)) \leqslant \int_Q |\det D_x G|\,dx.$$

We claim that this estimate holds with Q replaced by any Borel set in Ω. Indeed, if $U \subset \Omega$ is open, by Lemma 86 we can write $U = \bigcup_1^\infty Q_j$ where the Q_js are cubes with disjoint interiors. Since the boundaries of the cubes have Lebesgue measure zero, we have

$$m(G(U)) \leqslant \sum_1^\infty m(G(Q_j)) \leqslant \sum_1^\infty \int_{Q_j} |\det D_x G|\,dx = \int_U |\det D_x G|\,dx.$$

Moreover, if $E \subset \Omega$ is any Borel set of finite measure, by Theorem 82 there is a decreasing sequence of open sets $U_j \subset \Omega$ of finite measure such that $E \subset \bigcap_1^\infty U_j$ and $m(\bigcap_1^\infty U_j \setminus E) = 0$. Hence by the DCT,

$$\begin{aligned} m(G(E)) \leqslant m\Big(G\Big(\bigcap_1^\infty U_j\Big)\Big) &= \lim m(G(U_j)) \\ &\leqslant \lim \int_{U_i} |\det D_x G|\,dx = \int_E |\det D_x G|\,dx \end{aligned}$$

Finally, since m is σ-finite, it follows from this that $m(G(E)) \leqslant \int_E |\det D_x G|\,dx$ for any Borel set $E \subset \Omega$. If $f = \sum a_j \chi_{A_j}$ is a nonnegative simple function on $G(\Omega)$, we therefore have

$$\begin{aligned} \int_{G(\Omega)} f(x)\,dx = \sum a_j m(A_j) &\leqslant \sum a_j \int_{G^{-1}(A_j)} |\det D_x G|\,dx \\ &= \int_\Omega f \circ G(x) |\det D_x G|\,dx. \end{aligned}$$

Theorem 18 and the MCT then imply that

$$\int_{G(\Omega)} f(x)\,dx \leqslant \int_\Omega f \circ G(x) |\det D_x G|\,dx$$

for any nonnegative measurable f. But the same reasoning applies with G replaced by G^{-1} and f replaced by $f \circ G$, so that

$$\begin{aligned} &\int_\Omega f \circ G(x) |\det D_x G|\,dx \\ &\quad \leqslant \int_{G(\Omega))} f \circ G \circ G^{-1}(x) |\det D_{G^{-1}(x)} G| |\det D_x G^{-1}|\,dx = \int_{G(\Omega)} f(x)\,dx. \end{aligned}$$

This establishes (a) for $f \geqslant 0$, and the case $f \in L^1$ follows immediately. Since (b) is just the special case of (a) where $f = \chi_{G(E)}$, the proof is complete. □

Exercise 2.90: Folland Exercise 2.53.

Fill in the details of the proof of Lemma 86.

Exercise 2.91: Folland Exercise 2.54.

How much of Theorem 87 remains valid if T is not invertible?

Exercise 2.92: Folland Exercise 2.55.

Let $E = [0,1]\times[0,1]$. Investigate the existence and equality $\int_E f\, dm^2$, $\int_0^1\int_0^1 f(x,y)\,dx\,dy$, and $\int_0^1\int_0^1 f(x,y)\,dy\,dx$ for the following f.

(a) $f(x,y) = (x^2-y^2)(x^2+y^2)^{-2}$.

(b) $f(x,y) = (1-xy)^{-a}(a>0)$.

(c) $f(x,y) = \left(x-\frac{1}{2}\right)^{-3}$ if $0<y<\left|x-\frac{1}{2}\right|$, $f(x,y)=0$ otherwise.

Solution. (a) Note that

$$\frac{d}{dx}\left(\frac{x}{x^2+y^2}\right) = \frac{x^2+y^2-2x^2}{(x^2+y^2)^2} = \frac{y^2-x^2}{(x^2+y^2)^2}$$

so we have

$$\int_0^1\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}dxdy = -\int_0^1 \frac{x}{x^2+y^2}\bigg|_0^1 dy = -\int_0^1 \frac{1}{1+y^2}dy = -\tan^{-1}(1)+\tan^{-1}(0) = -\pi/4$$

$$\int_0^1\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}dxdy = -\int_0^1\int_0^1 \frac{y^2-x^2}{(x^2+y^2)^2}dxdy = \pi/4$$

Since the integrals are not equal, $f \notin L^1(E;dm^2)$ and the first integral above does not exist. Alternatively, we can prove that $f\notin L^1(E;dm^2)$ directly using Tonelli's:

$$\int_E |f| dm^2 = \int_0^1\int_0^x \frac{x^2-y^2}{(x^2+y^2)^2}dxdy - \int_0^1\int_0^y \frac{x^2-y^2}{(x^2+y^2)^2}dydx = 2\int_0^1 \frac{dx}{x} = \infty$$

where we have used the fact that for $x,y\geqslant 0$, $x^2-y^2=(x-y)(x+y)>0$ if and only if $x>y$. (b) Note that $f\in L^+$on E, so Tonelli's applies and shows the three integrals are equal for all $\alpha>0$. For those interested, direct calculation yields

$$\int_0^1\int_0^1 (1-xy)^{-\alpha}dxdy = \begin{cases} \int_0^1 \frac{(1-y)^{1-\alpha}-1}{y(\alpha-1)}dy, & \alpha\neq 1 \\ \int_0^1 \frac{\log(1-y)}{-y}dy, & \alpha = 1\end{cases}$$

both of which need special functions to explicitly calculate (the first needs hypergeometric functions and the second is a polylogrithm). □

Exercise 2.93: Folland Exercise 2.56.

If f is Lebesgue integrable on $(0,a)$ and $g(x)=\int_x^a t^{-1}f(t)\,dt$, then g is integrable on $(0,a)$ and $\int_0^a g(x)\,dx = \int_0^a f(x)\,dx$.

Exercise 2.94: Folland Exercise 2.57.

Show that $\int_0^\infty e^{-sx}x^{-1}\sin x\,dx = \arctan(s^{-1})$ for $s>0$ by integrating $e^{-sxy}\sin x$ with respect to x and y. (It may be useful to recall that $\tan\left(\frac{\pi}{2}-\theta\right) = (\tan\theta)^{-1}$. (See Folland Exercise 2.31(d).)

Exercise 2.95: Folland Exercise 2.58.

Show that $\int e^{-sx}x^{-1}\sin^2 x\,dx = \frac{1}{4}\log(1+4s^{-2})$ for $s>0$ by integrating $e^{-sx}\sin 2xy$ with respect to x and y.

Exercise 2.96: Folland Exercise 2.59.

Let $f(x) = x^{-1}\sin x$.

(a) Show that $\int_0^\infty |f(x)|\,dx = \infty$.

(b) Show that $\lim_{b\to\infty}\int_0^b f(x)\,dx = \frac{1}{2}\pi$ by integrating $e^{-xy}\sin x$ with respect to x and y. (In view of part (a), some care is needed in passing to the limit as $b\to\infty$.)

Exercise 2.97: Folland Exercise 2.60.

$\Gamma(x)\Gamma(y)/\Gamma(x+y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt$ for $x,y>0$, where Γ is defined in Folland Section 2.3. (Write $\Gamma(x)\Gamma(y)$ as a double integral and use the argument of the exponential as a new variable of integration.)

Solution. Recall

$$\Gamma(x) = \int_0^\infty u^{x-1}e^{-u}\,du, \quad \mathrm{Re}(x) > 0.$$

By Folland Exercise 2.51, if $f\in L^1(\mu), g\in L^1(\nu)$, and $h(x,y) = f(x)g(x)$, then $f\in L^1(\mu\times\nu)$ and

$$\int h\,d(\mu\times\nu) = \left(\int f d\mu\right)\left(\int g d\nu\right)$$

Hence we can write

$$\Gamma(x)\Gamma(y) = \int_0^\infty\int_0^\infty u^{x-1}e^{-u}v^{y-1}e^{-v}dudv$$

Now define $G\colon (0,\infty)\times(0,1)\to(0,\infty)\times(0,\infty), G(s,t) = (st, s(1-t))$, noting that the Jacobian determinant of G is

$$\det\begin{pmatrix}\frac{\partial G_1}{\partial s} & \frac{\partial G_1}{\partial t}\\ \frac{\partial G_2}{\partial s} & \frac{\partial G_2}{\partial t}\end{pmatrix} = \det\begin{pmatrix} t & s\\ 1-t & -s\end{pmatrix} = -st - s(1-t) = -s$$

Therefore, by the change of variables theorem we have

$$\begin{aligned}\Gamma(x)\Gamma(y) &= \int_0^1\int_0^\infty (st)^{x-1}e^{-st}[s(1-t)]^{y-1}e^{-s(1-t)}sdsdt\\ &= \int_0^1\int_0^\infty s^{x+y-1}e^{-s}(1-t)^{y-1}dsdt\\ &= \Gamma(x+y)\int_0^1 t^{x-1}(1-t)^{y-1}dt, \quad x,y>0,\end{aligned}$$

where we have applied Folland Exercise 2.51 once again. ☐

Exercise 2.98: Folland Exercise 2.61.

If f is continuous on $[0, \infty)$, for $\alpha > 0$ and $x \geqslant 0$ let

$$I_\alpha f(x) = \frac{1}{\Gamma(\alpha)} \int_0^x (x-t)^{\alpha-1} f(t)\, dt.$$

$I_\alpha f$ is called the ***αth fractional integral of f***.

(a) $I_{\alpha+\beta} f = I_\alpha(I_\beta f)$ for all $\alpha, \beta > 0$. (Use Folland Exercise 2.60.)
(b) If $n \in \mathbb{Z}_{\geqslant 0}$, $I_n f$ is an nth-order antiderivative of f.

Solution. By definition,

$$\begin{aligned} I_\alpha(I_\beta f)(x) &= \frac{1}{\Gamma(\alpha)} \int_0^x (x-t)^{\alpha-1} I_\beta f(t) dt \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x (x-t)^{\alpha-1} \int_0^t (t-s)^{\beta-1} f(s) ds dt \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x \int_0^t (x-t)^{\alpha-1}(t-s)^{\beta-1} f(s) ds dt, \quad \alpha, \beta > 0, x \geqslant 0. \end{aligned}$$

As we cannot integrate f directly, reversing the order of integration is needed. Note that f is continuous on $[0, \infty)$ so it is bounded above by some positive constant on $[0, t]$, and the remainder of the integrand is nonnegative. Thus to apply Fubini's Theorem, it suffices to consider for $\alpha, \beta > 0$,

$$\int_0^x \int_0^t (x-t)^{\alpha-1}(t-s)^{\beta-1} ds dt = \frac{1}{\beta} \int_0^x (x-t)^{\alpha-1} t^\beta dt < \infty,$$

where we conclude the integral is less than infinity as convergence near 0 requires $\beta > -1$ whereas convergence near x requires $\alpha - 1 > -1$. Hence, we apply Fubini to interchange the order of integration (being mindful of the variables in the bounds of integration) as follows:

$$\begin{aligned} I_\alpha(I_\beta f)(x) &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x \int_0^t (x-t)^{\alpha-1}(t-s)^{\beta-1} f(s) ds dt \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x \int_s^x (x-t)^{\alpha-1}(t-s)^{\beta-1} f(s) dt ds, \quad \alpha, \beta > 0, x \geqslant 0. \end{aligned}$$

In order to integrate with respect to t, we will need to perform a substitution. We first recall what was proven in Folland Exercise 2.60 regarding the beta function:

$$\mathrm{B}(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = \int_0^1 v^{x-1}(1-v)^{y-1}\, dv, \quad x, y > 0.$$

Comparing this result with we are proving, we notice we need the limits of integration to change from s and x to 0 and 1. This results in the substitution $u = (t-s)/(x-s)$ where $u(x) = 1, u(s) = 0$, and $dt = (x-s)du$. Hence, we compute

$$I_\alpha(I_\beta f)(x) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x \int_s^x (x-t)^{\alpha-1}(t-s)^{\beta-1} f(s) dt ds$$

$$= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x \int_0^1 (x - u(x-s) - s)^{\alpha-1}(x-s)^{\beta-1}u^{\beta-1}f(s)(x-s)duds$$
$$= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x \int_0^1 (1-u)^{\alpha-1}(x-s)^{\alpha-1}(x-s)^{\beta}u^{\beta-1}f(s)duds$$
$$= \frac{1}{\Gamma(\alpha+\beta)} \int_0^1 (x-s)^{\alpha+\beta-1}f(s)ds = I_{\alpha+\beta}f(x), \quad \alpha, \beta > 0, x \geqslant 0,$$

where we used Proposition 2 in the last line.

(b) Since

$$I_1 f(x) = \int_0^x f(t)dt, \quad x \geqslant 0,$$

it is clear that $I_1 f$ is an antiderivative of f, and recall that the Lebesgue integral is continuous (see Folland Exercise 2.26). The result now follows by induction and applying what was proven in (a): Assume that $I_n f, n \in \mathbb{Z}_{\geqslant 0}$ is an n th-order antiderivative of f noting that the integral is continuous once again, then write $(I_{n+1}f)^{(n+1)} = [(I_n(I_1 f))^{(n)}]' = (I_1 f)' = f$. □

2.8 Integration in Polar Coordinates

The most important nonlinear coordinate systems in $\mathbb{R}^2$ and $\mathbb{R}^3$ are polar coordinates $(x = r\cos\theta, y = r\sin\theta)$ and spherical coordinates $(x = r\sin\phi\cos\theta,\ y = r\sin\phi\sin\theta, z = r\cos\phi)$. Theorem 89, applied to these coordinates, yields the familiar formulas (loosely stated) $dxdy = rdrd\theta$ and $dxdydz = r^2\sin\phi drd\theta d\phi$. Similar coordinate systems exist in higher dimensions, but they become increasingly complicated as the dimension increases. (See Exercise 109.) For most purposes, however, it is sufficient to know that Lebesgue measure is effectively the product of the measure $r^{n-1}dr$ on $(0,\infty)$ and a certain "surface measure" on the unit sphere ($d\theta$ for $n = 2, \sin\phi d\theta d\phi$ for $n = 3$).

Our construction of this surface measure is motivated by a familiar fact from plane geometry. Namely, if S_θ is a sector of a disc of radius r with central angle θ (i.e., the region in the disc contained between the two sides of the angle), the area $m(S_\theta)$ is proportional to θ; in fact, $m(S_\theta) = \frac{1}{2}r^2\theta$. This equation can be solved for θ and hence used to define the angular measure θ in terms of the area $m(S_\theta)$. The same idea works in higher dimensions: We shall define the surface measure of a subset of the unit sphere in terms of the Lebesgue measure of the corresponding sector of the unit ball.

We shall denote the unit sphere $\{x \in \mathbb{R}^n \mid |x| = 1\}$ by S^{n-1}. If $x \in \mathbb{R}^n\backslash\{0\}$, the polar coordinates of x are

$$r = |x| \in (0,\infty), \quad x' = \frac{x}{|x|} \in S^{n-1}$$

The map $\Phi(x) = (r, x')$ is a continuous bijection from $\mathbb{R}^n\backslash\{0\}$ to $(0,\infty) \times S^{n-1}$ whose (continuous) inverse is $\Phi^{-1}(r, x') = rx'$. We denote by m_* the Borel measure on $(0,\infty)\times S^{n-1}$ induced by Φ from Lebesgue measure on $\mathbb{R}^n$, that is, $m_*(E) = m(\Phi^{-1}(E))$. Moreover, we define the measure $\rho = \rho_n$ on $(0, \infty)$ by $\rho(E) = \int_E r^{n-1} dr$.

Theorem 2.99.

There is a unique Borel measure $\sigma = \sigma_{n-1}$ on S^{n-1} such that $m_* = \rho \times \sigma$. If f is Borel measurable on $\mathbb{R}^n$ and $f \geqslant 0$ or $f \in L^1(m)$, then

$$\int_{\mathbb{R}^n} f(x)dx = \int_0^\infty \int_{S^{n-1}} f(rx')r^{n-1} d\sigma(x')dr. \tag{2.99.1}$$

Proof. Equation (2.99.1), when f is a characteristic function of a set, is merely a restatement of the equation $m_* = \rho \times \sigma$, and it follows for general f by the usual linearity and approximation arguments. Hence we need only to construct σ.

If E is a Borel set in S^{n-1}, for $a > 0$ let

$$E_a = \Phi^{-1}((0, a] \times E) = \{rx' \mid 0 < r \leqslant a, x' \in E\}$$

If Equation (2.99.1) is to hold when $f = \chi_{E_1}$, we must have

$$m(E_1) = \int_0^1 \int_E r^{n-1} d\sigma(x')dr = \sigma(E) \int_0^1 r^{n-1} dr = \frac{\sigma(E)}{n}.$$

We therefore define $\sigma(E)$ to be $n \cdot m(E_1)$. Since the map $E \mapsto E_1$ takes Borel sets to Borel sets and commutes with unions, intersections, and complements, it is clear that σ is a Borel measure on S^{n-1}. Also, since E_a is the image of E_1 under the map $x \mapsto ax$, it follows from Theorem 87 that $m(E_a) = a^n m(E_1)$, and hence, if $0 < a < b$,

$$\begin{aligned} m_*((a, b] \times E) = m(E_b \backslash E_a) &= \frac{b^n - a^n}{n} \sigma(E) = \sigma(E) \int_a^b r^{n-1} dr \\ &= \rho \times \sigma((a, b] \times E). \end{aligned}$$

Fix $E \in \mathcal{B}_{S^{n-1}}$ and let $\mathcal{A}_E$ be the collection of finite disjoint unions of sets of the form $(a, b] \times E$. By Proposition 7, $\mathcal{A}_E$ is an algebra on $(0, \infty) \times E$ that generates the σ-algebra $\mathcal{M}_E = \{A \times E \mid A \in \mathcal{B}_{(0,\infty)}\}$. By the preceding calculation we have $m_* = \rho \times \sigma$ on $\mathcal{A}_E$, and hence by the uniqueness assertion of Theorem 33, $m_* = \rho \times \sigma$ on $\mathcal{M}_E$. But $\bigcup\{\mathcal{M}_E \mid E \in \mathcal{B}_{S^{n-1}}\}$ is precisely the set of Borel rectangles in $(0, \infty) \times S^{n-1}$, so another application of the uniqueness theorem shows that $m_* = \rho \times \sigma$ on all Borel sets. □

Of course, Equation (2.99.1) can be extended to Lebesgue measurable functions by considering the completion of the measure σ. Details are left to the reader.

Corollary 2.100.

If f is a measurable function on $\mathbb{R}^n$, nonnegative or integrable, such that $f(x) = g(|x|)$ for some function g on $(0, \infty)$, then

$$\int f(x)dx = \sigma(S^{n-1}) \int_0^\infty g(r)r^{n-1} dr.$$

Corollary 2.101.

Let c and C denote positive constants, and let $B = \{x \in \mathbb{R}^n \mid |x| < c\}$. Suppose that f is a measurable function on $\mathbb{R}^n$.

(a) If $|f(x)| \leqslant C|x|^{-a}$ on B for some $a < n$, then $f \in L^1(B)$. However, if $|f(x)| \geqslant C|x|^{-n}$ on B, then $f \notin L^1(B)$.

(b) If $|f(x)| \leqslant C|x|^{-a}$ on B^c for some $a > n$, then $f \in L^1(B^c)$. However, if $|f(x)| \geqslant C|x|^{-n}$ on B^c, then $f \notin L^1(B^c)$.

Proof. Apply Corollary 100 to $|x|^{-a}\chi_B$ and $|x|^{-a}\chi_{B^c}$. □

We shall compute $\sigma(S^{n-1})$ shortly. Of course, we know that $\sigma(S^1) = 2\pi$; this is just the definition of 2π as the ratio of the circumference of a circle to its radius. Armed with this fact, we can compute a very important integral.

Proposition 2.102: 2.53.

If $a > 0$,

$$\int_{\mathbb{R}^n} \exp(-a|x|^2)dx = \left(\frac{\pi}{a}\right)^{n/2}$$

Proof. Denote the integral on the left by I_n. For $n = 2$, by Corollary 100 we have

$$I_2 = 2\pi \int_0^\infty re^{-ar^2}dr = -\left(\frac{\pi}{a}\right)e^{-ar^2}\Big|_0^\infty = \frac{\pi}{a}$$

Since $\exp(-a|x|^2) = \prod_1^n \exp(-ax_j^2)$, Tonelli's theorem implies that $I_n = (I_1)^n$. In particular, $I_1 = (I_2)^{1/2}$, so $I_n = (I_2)^{n/2} = (\pi/a)^{n/2}$. □

Once we know this result, the device used in its proof can be turned around to compute $\sigma(S^{n-1})$ for all n in terms of the gamma function introduced in §2.3.

Proposition 2.103: 2.54.

$\sigma(S^{n-1}) = \frac{2\pi^{n/2}}{\Gamma(n/2)}$.

Proof. By Corollary 100, Proposition 102, and the substitution $s = r^2$,

$$\pi^{n/2} = \int_{\mathbb{R}^n} e^{-|x|^2}dx = \sigma(S^{n-1})\int_0^\infty r^{n-1}e^{-r^2}dr$$
$$= \frac{\sigma(S^{n-1})}{2}\int_0^\infty s^{(n/2)-1}e^{-s}ds = \frac{\sigma(S^{n-1})}{2}\Gamma\left(\frac{n}{2}\right).$$ □

Corollary 2.104: 2.55.

If $B^n = \{x \in \mathbb{R}^n \mid |x| < 1\}$, then $m(B^n) = \frac{\pi^{n/2}}{\Gamma(\frac{1}{2}n+1)}$.

Proof. $m(B^n) = n^{-1}\sigma(S^{n-1})$ by definition of σ, and $\frac{1}{2}n\Gamma\left(\frac{1}{2}n\right) = \Gamma\left(\frac{1}{2}n+1\right)$ by the functional equation for the gamma function. □

We observed in §2.3 that $\Gamma(n) = (n-1)!$. The following proposition shows that we can also evaluate the gamma function at the half-integers.

Proposition 2.105: 2.56.

$\Gamma\left(n+\frac{1}{2}\right) = \left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)\cdots\left(\frac{1}{2}\right)\sqrt{\pi}$.

Proof. We have $\Gamma\left(n+\frac{1}{2}\right) = \left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)\cdots\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)$ by the functional equation, and by Proposition 102 and the substitution $s = r^2$,

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty s^{-1/2}e^{-s}ds = 2\int_0^\infty e^{-r^2}dr = \int_{-\infty}^\infty e^{-r^2}dr = \sqrt{\pi}$$

An amusing consequence of Proposition 105 and the formula $\Gamma(n) = (n-1)!$ is that the surface measure of the unit sphere and the Lebesgue measure of the unit ball in $\mathbb{R}^n$ are always rational multiples of integer powers of π, and the power of π increases by 1 when n increases by 2. □

Exercise 2.106: Folland Exercise 2.62.

The measure σ on S^{n-1} is invariant under rotations.

Exercise 2.107: Folland Exercise 2.63.

The technique used to prove Proposition 103 can also be used to integrate any polynomial over S^{n-1}. In fact, suppose $f(x) = \prod_1^n x_j^{\alpha_j}(\alpha_j \in \mathbb{N} \cup \{0\})$ is a monomial. Then $\int f d\sigma = 0$ if any α_j is odd, and if all α_j's are even,

$$\int f d\sigma = \frac{2\Gamma(\beta_1)\cdots\Gamma(\beta_n)}{\Gamma(\beta_1+\cdots+\beta_n)}, \text{ where } \beta_j = \frac{\alpha_j+1}{2}.$$

Exercise 2.108: Folland Exercise 2.64.

For which real values of a and b is $|x|^a|\log|x||^b$ integrable over $\left\{x \in \mathbb{R}^n \mid |x| < \frac{1}{2}\right\}$? Over $\{x \in \mathbb{R}^n \mid |x| > 2\}$?

Exercise 2.109: Folland Exercise 2.65.

Define $G\colon \mathbb{R}^n \to \mathbb{R}^n$ by $G(r, \phi_1, \ldots, \phi_{n-2}, \theta) = (x_1, \ldots, x_n)$, where

$$x_1 = r\cos\phi_1, \quad x_2 = r\sin\phi_1\cos\phi_2, \quad x_3 = r\sin\phi_1\sin\phi_2\cos\phi_3, \ldots,$$
$$x_{n-1} = r\sin\phi_1\cdots\sin\phi_{n-2}\cos\theta, \quad x_n = r\sin\phi_1\cdots\sin\phi_{n-2}\sin\theta.$$

(a) G maps $\mathbb{R}^n$ onto $\mathbb{R}^n$, and $|G(r, \phi_1, \ldots, \phi_{n-2}, \theta)| = |r|$.
(b) $\det D_{(r,\phi_1,\ldots,\phi_{n-2},\theta)} G = r^{n-1}\sin^{n-2}\phi_1\sin^{n-3}\phi_2\cdots\sin\phi_{n-2}$.
(c) Let $\Omega = (0,\infty)\times(0,\pi)^{n-2}\times(0,2\pi)$. Then $G \mid \Omega$ is a diffeomorphism and $m(\mathbb{R}^n\backslash G(\Omega)) = 0$.
(d) Let $F(\phi_1, \ldots, \phi_{n-2}, \theta) = G(1, \phi_1, \ldots, \phi_{n-2}, \theta)$ and $\Omega' = (0,\pi)^{n-2}\times(0,2\pi)$. Then $(F \mid \Omega')^{-1}$ defines a coordinate system on S^{n-1} except on a σ-null set, and the measure σ is given in these coordinates by

$$d\sigma(\phi_1, \ldots \phi_{n-2}, \theta) = \sin^{n-2}\phi_1\sin^{n-3}\phi_2\cdots\sin\phi_{n-2}d\phi_1\cdots d\phi_{n-2}d\theta.$$

3 Signed Measures and Differentiation

The principal theme of this chapter is the concept of differentiating a measure ν with respect to another measure μ on the same σ-algebra. We do this first on the abstract level, then obtain a more refined result when μ is Lebesgue measure on $\mathbb{R}^n$. When the latter is specialized to the case $n = 1$, it joins with classical real-variable theory to produce a version of the fundamental theorem of calculus for Lebesgue integrals.

In developing this program it is useful to generalize the notion of measure so as to allow measures to assume negative values. In applications such "signed measures" can represent things such as electric charge that can be either positive or negative.

3.1 Signed Measures

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

Definition 1. *A* ***signed measure*** *on $(X, \mathcal{M})$ is a function $\nu\colon \mathcal{M} \to [-\infty, \infty]$ such that*

- *$\nu(\varnothing) = 0$;*
- *ν assumes at most one of the values $\pm\infty$;*
- *if $\{E_j\}$ is a sequence of disjoint sets in $\mathcal{M}$, then $\nu(\bigcup_1^\infty E_j) = \sum_1^\infty \nu(E_j)$, where the latter sum converges absolutely if $\nu(\bigcup_1^\infty E_j)$ is finite.*

Thus every measure is a signed measure; for emphasis we shall sometimes refer to a measure as a ***positive measure***.

Example 2. *First, if μ_1, μ_2 are measures on $\mathcal{M}$ and at least one of them is finite, then $\nu = \mu_1 - \mu_2$ is a signed measure. Second,*

Example 3. *if μ is a measure on $\mathcal{M}$ and $f\colon X \to [-\infty, \infty]$ is a measurable function such that at least one of $\int f^+ d\mu$ and $\int f^- d\mu$ is finite (in which case we shall call f* ***extended μ-integrable****), then the set function ν defined by $\nu(E) = \int_E f\, d\mu$ is a signed measure.*

In fact, we shall see shortly that these are really the *only* examples: Every signed measure can be represented in either of these two forms.

Proposition 3.4: 3.1.

Let ν be a signed measure on $(X, \mathcal{M})$.

(i) If $\{E_j\}$ is an increasing sequence in $\mathcal{M}$, then $\nu(\bigcup_1^\infty E_j) = \lim_{j\to\infty} \nu(E_j)$.

(ii) If $\{E_j\}$ is a decreasing sequence in $\mathcal{M}$ and $\nu(E_1)$ is finite, then $\nu(\bigcap_1^\infty E_j) = \lim_{j\to\infty} \nu(E_j)$.

The proof is essentially the same as for positive measures (Theorem 14) and is left to the reader (Folland Exercise 3.1).

If ν is a signed measure on $(X, \mathcal{M})$, a set $E \in \mathcal{M}$ is called ***positive for ν*** (resp. ***negative for ν***, ***null for ν***) for ν if $\nu(F) \geqslant 0$ (resp. $\nu(F) \leqslant 0, \nu(F) = 0$) for all $F \in \mathcal{M}$ such that $F \subset E$. (Thus, in the example $\nu(E) = \int_E f\, d\mu$ described above, E is positive, negative, or null precisely when $f \geqslant 0, f \leqslant 0$, or $f = 0$ μ-a.e. on E.)

Lemma 3.5: 3.2.

Any measurable subset of a positive set is positive, and the union of any countable family of positive sets is positive.

Proof. The first assertion follows from the definition of positivity. If $P_1, P_2, \ldots$ are positive sets, let $Q_n = P_n \setminus \bigcup_1^{n-1} P_j$. Then $Q_n \subset P_n$, so Q_n is positive. Hence if $E \subset \bigcup_1^\infty P_j$, then $\nu(E) = \sum_1^\infty \nu(E \cap Q_j) \geqslant 0$, as desired. □

Theorem 3.6: 3.3: The Hahn Decomposition Theorem.

If ν is a signed measure on $(X, \mathcal{M})$, there exist a positive set P and a negative set N for ν such that $P \cup N = X$ and $P \cap N = \varnothing$. If P', N' is another such pair, then $P \triangle P' (= N \triangle N')$ is null for ν.

Proof. Without loss of generality, we assume that ν does not assume the value $-\infty$. (Otherwise, consider $-\nu$.) Let m be the supremum of $\nu(E)$ as E ranges over all positive sets; thus there is a sequence $\{P_j\}$ of positive sets such that $\nu(P_j) \to m$. Let $P = \bigcup_1^\infty P_j$. By Lemma 5 and Proposition 4, P is positive and $\nu(P) = m$; in particular, $m < \infty$. We claim that $N = X \setminus P$ is negative. To this end, we assume that N is nonnegative and derive a contradiction.

First, notice that N cannot contain any nonnull positive sets. Indeed, if $E \subset N$ is positive and $\nu(E) > 0$, then $E \cup P$ is positive and $\nu(E \cup P) = \nu(E) + \nu(P) > m$, which is impossible.

Second, if $A \subset N$ and $\nu(A) > 0$, there exists $B \subset A$ with $\nu(B) > \nu(A)$. Indeed, since A cannot be positive, there exists $C \subset A$ with $\nu(C) < 0$; thus if $B = A \setminus C$ we have $\nu(B) = \nu(A) - \nu(C) > \nu(A)$.

If N is not negative, then, we can specify a sequence of subsets $\{A_j\}$ of N and a sequence $\{n_j\}$ of positive integers as follows: n_1 is the smallest integer for which there exists a set $B \subset N$ with $\nu(B) > n_1^{-1}$, and A_1 is such a set. Proceeding inductively, n_j is the smallest integer for which there exists a set $B \subset A_{j-1}$ with $\nu(B) > \nu(A_{j-1}) + n_j^{-1}$, and A_j is such a set.

Let $A = \bigcap_1^\infty A_j$. Then $\infty > \nu(A) = \lim \nu(A_j) > \sum_1^\infty n_j^{-1}$, so $n_j \to \infty$ as $j \to \infty$. But once again, there exists $B \subset A$ with $\nu(B) > \nu(A) + n^{-1}$ for some integer n. For j sufficiently large we have $n < n_j$, and $B \subset A_{j-1}$, which contradicts the construction of n_j and A_j. Thus the assumption that N is not negative is untenable.

Finally, if P', N' is another pair of sets as in the statement of the theorem, we have $P \setminus P' \subset P$ and $P \setminus P' \subset N'$, so that $P \setminus P'$ is both positive and negative, hence null; likewise for $P' \setminus P$. □

The decomposition $X = P \cup N$ if X as the disjoint union of a positive set and a negative set is called a ***Hahn decomposition for ν***. It is usually not unique (ν-null sets can be transferred from P to N or from N to P), but it leads to a canonical representation of ν as the difference of two positive measures.

To state this result we need a new concept: We say that two signed measures μ and ν on $(X, \mathcal{M})$ are ***mutually singular***, or that ν is ***singular with respect to μ***, or vice versa, if there exist $E, F \in \mathcal{M}$ such that $E \cap F = \varnothing, E \cup F = X, E$ is null for μ, and F is null for ν. Informally speaking, mutual singularity means that μ and ν "live on disjoint sets." We express this relationship symbolically with the perpendicularity sign, namely $\mu \perp \nu$.

Theorem 3.7: 3.4: The Jordan Decomposition Theorem.

If ν is a signed measure, there exist unique positive measures ν^+ and ν^- such that $\nu = \nu^+ - \nu^-$ and $\nu^+ \perp \nu^-$.

Proof. Let $X = P \cup N$ be a Hahn decomposition for ν, and define $\nu^+(E) = \nu(E \cap P)$ and $\nu^-(E) = -\nu(E \cap N)$. Then clearly $\nu = \nu^+ - \nu^-$ and $\nu^+ \perp \nu^-$. If also $\nu = \mu^+ - \mu^-$ and $\mu^+ \perp \mu^-$, let $E, F \in \mathcal{M}$ be such that $E \cap F = \varnothing, E \cup F = X$, and $\mu^+(F) = \mu^-(E) = 0$. Then $X = E \cup F$ is another Hahn decomposition for ν, so $P \triangle E$ is ν-null. Therefore, for any $A \in \mathcal{M}, \mu^+(A) = \mu^+(A \cap E) = \nu(A \cap E) = \nu(A \cap P) = \nu^+(A)$, and likewise $\nu^- = \mu^-$. □

Definition 8. *The measures ν^+ and ν^- are called the* ***positive variation of ν*** *and* ***negative variation of ν****, respectively, and $\nu = \nu^+ - \nu^-$ is called the* ***Jordan decomposition of ν****, by analogy with the representation of a function of bounded variation on $\mathbb{R}$ as the difference of two increasing functions (see Folland Section 3.5). Furthermore, we define the* ***total variation of ν*** *to be the measure $|\nu|$ defined by*

$$|\nu| := \nu^+ + \nu^-.$$

Exercise 3.9: Folland Exercise 3.2.

If ν is a signed measure,

$$E \text{ is } \nu\text{-null} \iff |\nu|(E) = 0.$$

Also, if ν and μ are signed measures, then

$$\nu \perp \mu \iff |\nu| \perp \mu \iff \nu^+ \perp \mu \text{ and } \nu^- \perp \mu.$$

Solution. Suppose E is ν-null and $X = P \cup N$ is a Hahn decomposition of X for ν. Then

$$|\nu|(E) = |\nu|(E \cap X) = \nu^+(E \cap P) + \nu^-(E \cap N).$$

If $|\nu|(E) > 0$, either $\nu^+(E \cap P) > 0$ or $\nu^-(E \cap N) > 0$. Without loss of generality, assume $\nu^+(E \cap P) > 0$. Then

$$\nu(E \cap P) = \nu^+(E \cap P) - \nu^-(E \cap P \cap N) = \nu^+(E \cap P) > 0.$$

However, $E \cap P \subset E$ with E is ν-null, so we must have $\nu(E \cap P) = 0$, a contradiction. Suppose $|\nu|(E) = 0$. Then $|\nu|(A) = 0$ for all measurable $A \subset E$ since $|\nu|$ is positive. Hence,

$$0 = |\nu|(A) = \nu^+(A) + \nu^-(A) \iff \nu^+(A) = 0 = \nu^-(A),$$

so that $\nu(A) = \nu^+(A) - \nu^-(A) = 0$ for all measurable $A \subset E$. Thus, E is ν-null. The second assertion is by definition:

$$\begin{aligned} \nu \perp \mu &\iff \text{there exists } \{P, N\} \text{ such that } P \text{ is } \mu\text{-null, } N \text{ is } \nu\text{-null} \\ &\iff |\nu| \perp \mu \iff \nu^+ \perp \mu \text{ and } \nu^- \perp \mu. \end{aligned}$$

□

We observe that if ν omits the value ∞ then $\nu^+(X) = \nu(P) < \infty$, so that ν^+ is a finite positive measure and ν is bounded above by $\nu^+(X)$; similarly if ν omits the value $-\infty$. In particular, if the range of ν is contained in $\mathbb{R}$, then ν is bounded. We observe also that ν is of the form $\nu(E) = \int_E f\, d\mu$, where $\mu = |\nu|$ and $f = \chi_P - \chi_N, X = P \cup N$ being a Hahn decomposition for ν.

Integration with respect to a signed measure ν is defined in the obvious way: We set

$$L^1(\nu) = L^1(\nu^+) \cap L^1(\nu^-),$$

$$\int f\, d\nu = \int f\, d\nu^+ - \int f\, d\nu^- \quad (f \in L^1(\nu)).$$

Definition 10. *A* ***finite signed measure*** *(resp.* ***σ-finite signed measure****) is a signed measure ν such that $|\nu|$ is finite (resp. σ-finite).*

Exercise 3.11: Folland Exercise 3.1.

Prove Proposition 4.

Exercise 3.12: Folland Exercise 3.3.

Let ν be a signed measure on $(X, \mathcal{M})$.

(a) $L^1(\nu) = L^1(|\nu|)$.
(b) If $f \in L^1(\nu)$, then $\left|\int f\, d\nu\right| \leqslant \int |f| d|\nu|$.
(c) If $E \in \mathcal{M}$, $|\nu|(E) = \sup\left\{\left|\int_E f\, d\nu\right| \,\middle|\, |f| \leqslant 1\right\}$.

Solution.

(a) Let $X = P \cup N$ be a Hahn decomposition of X for ν, and let $\nu = \nu^+ - \nu^-$ be the corresponding Jordan decomposition of ν. Since $|\nu| = \nu^+ + \nu^- \geqslant \nu^+ - \nu^- = \nu$, for any nonnegative simple function $\phi = \sum_{j=1}^n a_j \chi_{E_j}$ we can write
$$\int \phi\, d|\nu| = \sum_{j=1}^n a_j |\nu|(E_j) = \sum_{j=1}^n a_j \nu^+(E_j) + \sum_{j=1}^n a_j \nu^-(E_j)$$
$$\geqslant \sum_{j=1}^n a_j \nu^+(E_j) - \sum_{j=1}^n a_j \nu^-(E_j) = \sum_{j=1}^n a_j \nu(E_j) = \int \phi\, d\nu.$$
Taking the supremum of both sides over all simple functions ϕ satisfying $0 \leqslant \phi \leqslant |f|$, we obtain $\int |f|\, d|\nu| \geqslant \int |f|\, d\nu$. Now $f \in L^1(|\nu|)$ implies $\int |f|\, d\nu \leqslant \int |f|\, d|\nu| < \infty$, which in turn implies $f \in L^1(\nu)$, so $L^1(|\nu|) \subset L^1(\nu)$. On the other hand, suppose $f \in L^1(\nu)$, so that $\int |f|\, d\nu^+, \int |f|\, d\nu^- < \infty$. Then $\int |f|\, d|\nu| = \int |f|\, d\nu^+ + \int |f|\, d\nu^- < \infty$, so $f \in L^1(|\nu|)$. Hence $L^1(\nu) = L^1(|\nu|)$. [Note that the equality $\int |f|\, d|\nu| = \int |f|\, d\nu^+ + \int |f|\, d\nu^-$ holds for simple functions as above and so by the monotone convergence theorem it holds for $|f|$.]

(b) Let $f \in L^1(\nu)$. Then
$$\left|\int f\, d\nu\right| = \left|\int f\, d\nu^+ - \int f\, d\nu^-\right| \leqslant \left|\int f\, d\nu^+\right| + \left|\int f\, d\nu^-\right|$$
$$\leqslant \int |f|\, d\nu^+ + \int |f|\, d\nu^- = \int |f|\, d|\nu|,$$
as desired.

(c) Let $E \in \mathcal{M}$. If $|\nu|(E) = \infty$ then the equality $|\nu|(E) = \sup\left\{\left|\int_E f\, d\nu\right| \,\middle|\, |f| \leqslant 1\right\}$ holds as $\sup\left\{\left|\int_E f\, d\nu\right| \,\middle|\, |f| \leqslant 1\right\} = \infty$ (see inequality below which forces this). Thus we may assume $|\nu|(E) < \infty$. Let f be a measurable function with $|f| \leqslant 1$. Then $f\chi_E \in L^1(\nu)$, since $\int |f\chi_E|\, d\nu \leqslant \int_E 1\, d\nu = \nu(E) \leqslant |\nu|(E) < \infty$, so we can apply part

(b) as follows:

$$|\nu|(E) = \int \chi_E \, d|\nu| \geqslant \int |f\chi_E| \, d|\nu| \overset{\text{(b)}}{\geqslant} \left| \int f\chi_E \, d\nu \right| = \left| \int_E f \, d\nu \right|.$$

Taking the supremum of both sides over all such f, we obtain $|\nu(E)| \geqslant \sup\{|\int_E f \, d\nu| \mid |f| \leqslant 1\}$.

Conversely (this does not assume $|\nu|(E)$ finite),

$$\begin{aligned} |\nu|(E) &= \int \chi_E \, d|\nu| = \int \chi_E \, d\nu^+ + \int \chi_E \, d\nu^- = \int \chi_{E\cap P} \, d\nu - \int \chi_{E\cap N} \, d\nu \\ &= \int_E (\chi_P - \chi_N) \, d\nu \leqslant \left| \int_E (\chi_P - \chi_N) \, d\nu \right| \leqslant \sup\left\{ \left| \int_E f \, d\nu \right| \,\middle|\, |f| \leqslant 1 \right\} \end{aligned}$$

where the last inequality is because $\chi_P - \chi_N$ is a measurable function whose absolute value is at most 1. This proves the reverse inequality, so we conclude

$$|\nu|(E) = \sup\left\{ \left| \int_E f \, d\nu \right| \,\middle|\, |f| \leqslant 1 \right\}. \qquad \square$$

Exercise 3.13: Folland Exercise 3.4.

If ν is a signed measure and λ, μ are positive measures such that $\nu = \lambda - \mu$, then $\lambda \geqslant \nu^+$ and $\mu \geqslant \nu^-$.

Solution. Consider a Hahn decomposition $X = P \cup N$ and notice for all $E \in \mathcal{M}$,

$$\begin{aligned} \nu^+(E) &= \nu(E \cap P) = \lambda(E \cap P) - \mu(E \cap P) \leqslant \lambda(E \cap P) \leqslant \lambda(E), \\ \nu^-(E) &= -\nu(E \cap N) = -(\lambda(E \cap N) - \mu(E \cap N)) \leqslant \mu(E \cap N) \leqslant \mu(E). \end{aligned} \qquad \square$$

Exercise 3.14.

If μ and ν are signed measures and $\lambda = \mu + \nu$, is it true that $\lambda^+ = \mu^+ + \nu^+$?

Solution. No, not in general unless $(\mu^+ + \nu^+) \perp (\mu^- + \nu^-)$. For example, consider the finite signed measures $\mu = \delta_x - \delta_y$ and $\nu = \delta_y - \delta_z$, where δ_x is the point mass at x. Then $\lambda = \delta_x - \delta_z$ and

$$\mu^+ = \delta_x, \quad \nu^+ = \delta_y, \quad \lambda^+ = \delta_x \neq \delta_x + \delta_y = \mu^+ + \nu^+. \qquad \square$$

Exercise 3.15: Folland Exercise 3.5.

If ν_1, ν_2 are signed measures that both omit the value $+\infty$ or $-\infty$, then $|\nu_1 + \nu_2| \leqslant |\nu_1| + |\nu_2|$. (Use Folland Exercise 3.4.)

Solution. Consider the Jordan decomposition $\lambda = \nu_1 + \nu_2 = \lambda^+ - \lambda^-, \nu_1 + \nu_2 = (\nu_1^+ + \nu_2^+) - (\nu_1^- + \nu_2^-)$, noting that $\nu_1^+ + \nu_2^+$ and $\nu_1^- + \nu_2^-$are positive measures. By Folland Exercise

3.4,

$$\nu_1^+ + \nu_2^+ \geqslant \lambda^+ \qquad \text{and} \qquad \nu_1^- + \nu_2^- \geqslant \lambda^-,$$

so

$$|\nu_1 + \nu_2| = \lambda^+ + \lambda^- \leqslant |\nu_1| + |\nu_2|. \qquad \square$$

Exercise 3.16: Folland Exercise 3.6.

Suppose $\nu(E) = \int f \, d\mu$ where μ is a positive measure and f is an extended μ-integrable function. Describe the Hahn decompositions of ν and the positive, negative, and total variations of ν in terms of f and μ.

Exercise 3.17: Folland Exercise 3.7.

Suppose that ν is a signed measure on $(X, \mathcal{M})$ and $E \in \mathcal{M}$.

(a) $\nu^+(E) = \sup\{\nu(F) \mid E \in \mathcal{M}, F \subset E\}$ and $\nu^-(E) = -\inf\{\nu(F) \mid F \in \mathcal{M}, F \subset E\}$.

(b) $|\nu|(E) = \sup\{\sum_1^n |\nu(E_j)| \mid n \in \mathbb{Z}_{\geqslant 0}, E_1, \ldots, E_n \text{ are disjoint, and } \bigcup_1^n E_j = E\}$.

3.2 The Lebesgue Decomposition and the Radon-Nikodym Derivative

Suppose that ν is a signed measure and μ is a positive measure on $(X, \mathcal{M})$. We say that ν is ***absolutely continuous with respect to μ*** and write $\nu \ll \mu$ if $\nu(E) = 0$ for every $E \in \mathcal{M}$ for which $\mu(E) = 0$. It is easily verified that $\nu \ll \mu$ if and only if $|\nu| \ll \mu$ if and only if $\nu^+ \ll \mu$ and $\nu^- \ll \mu$ (Folland Exercise 3.8).

Absolute continuity is in a sense the antithesis of mutual singularity. More precisely, if $\nu \perp \mu$ and $\nu \ll \mu$, then $\nu = 0$, for if E and F are disjoint sets such that $E \cup F = X$ and $\mu(E) = |\nu|(F) = 0$, then the fact that $\nu \ll \mu$ implies that $|\nu|(E) = 0$, whence $|\nu| = 0$ and $\nu = 0$. One can extend the notion of absolute continuity to the case where μ is a signed measure (namely, $\nu \ll \mu$ if and only if $\nu \ll |\mu|$), but we shall have no need of this more general definition.

The term "absolute continuity" is derived from real-variable theory; see Folland Section 3.5. For finite signed measures it is equivalent to another condition that is obviously a form of continuity.

Theorem 3.18: 3.5.

Let ν be a finite signed measure and μ a positive measure on (X, M). Then $\nu \ll \mu$ if and only if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|\nu(E)| < \varepsilon$ whenever $\mu(E) < \delta$.

Proof. Since $\nu \ll \mu$ if and only if $|\nu| \ll \mu$ and $|\nu(E)| \leqslant |\nu|(E)$, it suffices to assume that $\nu = |\nu|$ is positive. Clearly the ε-δ condition implies that $\nu \ll \mu$. On the other hand, if the ε-δ condition is not satisfied, there exists $\varepsilon > 0$ such that for all $n \in \mathbb{Z}_{\geqslant 0}$ we can find $E_n \in \mathcal{M}$ with $\mu(E_n) < 2^{-n}$ and $\nu(E_n) \geqslant \varepsilon$. Let $F_k = \bigcup_k^\infty E_n$ and $F = \bigcap_1^\infty F_k$. Then $\mu(F_k) < \sum_k^\infty 2^{-n} = 2^{1-k}$, so $\mu(F) = 0$; but $\nu(F_k) \geqslant \varepsilon$ for all k and hence, since ν is finite, $\nu(F) = \lim \nu(F_k) \geqslant \varepsilon$. Thus it is false that $\nu \ll \mu$. □

If μ is a measure and f is an extended μ-integrable function, the signed measure ν defined by $\nu(E) = \int_E f\, d\mu$ is clearly absolutely continuous with respect to μ; it is finite if and only if $f \in L^1(\mu)$. For any complex-valued $f \in L^1(\mu)$, the preceding theorem can be applied to $\operatorname{Re} f$ and $\operatorname{Im} f$, and we obtain the following useful result:

Corollary 3.19: 3.6.

If $f \in L^1(\mu)$, for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\left|\int_E f\, d\mu\right| < \varepsilon$ whenever $\mu(E) < \delta$.

We shall use the following notation to express the relationship $\nu(E) = \int_E f\, d\mu$:

$$d\nu = f\, d\mu.$$

Sometimes, by a slight abuse of language, we shall refer to ***the signed measure $f\, d\mu$***.

Given a measure, μ, one can construct a new (signed/complex) measure, ν, by defining

$$\int_E d\nu = \nu(E) = \int_E f d\mu,$$

for μ-measurable f. In other words, we can use a measure that we know about (the reference measure μ) to define a new measure (ν) that acts by integration of a measurable function with respect to the reference measure.

What about the opposite direction? That is, if you want to learn about a measure that you do not know about, can you now write it as 'almost' a measure you know about in the sense of the above integration of a measurable function with respect to the known reference measure? If we can do this, it would make using this measure much easier!

The Lebesgue-Radon-Nikodym Theorem answers how to do this by using the Lebesgue decomposition of the σ-finite signed measure $\nu = \mu_s + \mu_{ac}$ where $\mu_s \perp \mu$ and $\mu_{ac} \ll \mu$ for μ a σ-finite positive measure (the reference measure we know about), with all measures understood on the same measurable space $(X, \mathcal{M})$.

In general, we only have the above decomposition and there will be a singular part, μ_s, which will need to be contended with. However, since $\nu = \mu_{ac} + \mu_s$ with $\mu_s \perp \mu, \mu_{ac} \ll \mu$, if we further assume that $\nu \ll \mu$, then $\mu_s \ll \mu$. But this implies that $\mu_s = 0$ (singular and absolutely continuous), so the representation simply becomes $d\nu = f d\mu$. If this is the case, we then call f the ***Radon-Nikodym derivative*** of ν with respect to μ and denote it by $d\nu/d\mu$.

We now come to the main theorem of this section, which gives a complete picture of the structure of a signed measure relative to a given positive measure. First, a technical lemma.

Lemma 3.20: 3.7.

Suppose that ν and μ are finite measures on $(X, \mathcal{M})$. Either $\nu \perp \mu$, or there exist $\varepsilon > 0$ and $E \in \mathcal{M}$ such that $\mu(E) > 0$ and $\nu \geqslant \varepsilon\mu$ on E (that is, E is a positive set for $\nu - \varepsilon\mu$).

Proof. Let $X = P_n \cup N_n$ be a Hahn decomposition for $\nu - n^{-1}\mu$, and let $P = \bigcup_1^\infty P_n$ and $N = \bigcap_1^\infty N_n = P^c$. Then N is a negative set for $\nu - n^{-1}\mu$ for all n, i.e., $0 \leqslant \nu(N) \leqslant n^{-1}\mu(N)$ for all n, so $\nu(N) = 0$. If $\mu(P) = 0$, then $\nu \perp \mu$. If $\mu(P) > 0$, then $\mu(P_n) > 0$ for some n, and P_n is a positive set for $\nu - n^{-1}\mu$. □

Theorem 3.21: 3.8: The Lebesgue-Radon-Nikodym Theorem.

Let ν be a σ-finite signed measure and μ a σ-finite positive measure on $(X, \mathcal{M})$. There exist unique σ-finite signed measures ν_s, ν_{ac} on $(X, \mathcal{M})$ such that

$$\nu_s \perp \mu, \quad \nu_{ac} \ll \mu, \quad \text{and} \quad \nu = \nu_s + \nu_{ac}.$$

Moreover, there is an extended μ-integrable function $\frac{d\nu_{ac}}{d\mu} \colon X \to \mathbb{R}$, called the ***Radon-Nikodym derivative*** of ν_{ac} with respect to μ, such that for all $E \in \mathcal{M}$,

$$\nu_{ac}(E) = \int_E \frac{d\nu_{ac}}{d\mu}\, d\mu,$$

and any two such functions are equal μ-a.e.

Proof. For the proof we use Folland's notation $\nu_s = \rho$ and $\nu_{ac} = \lambda$.

Case I: Suppose that ν and μ are finite positive measures. Let

$$\mathcal{F} = \left\{ f \colon X \to [0, \infty] \;\middle|\; \int_E f\, d\mu \leqslant \nu(E) \text{ for all } E \in \mathcal{M} \right\}.$$

$\mathcal{F}$ is nonempty since $0 \in \mathcal{F}$. Also, if $f, g \in \mathcal{F}$, then $h = \max(f, g) \in \mathcal{F}$, for if $A = \{x \mid f(x) > g(x)\}$, for any $E \in \mathcal{M}$ we have

$$\int_E h\, d\mu = \int_{E \cap A} f\, d\mu + \int_{E \setminus A} g\, d\mu \leqslant \nu(E \cap A) + \nu(E \setminus A) = \nu(E).$$

Let $a = \sup\{\int f\, d\mu \mid f \in \mathcal{F}\}$, noting that $a \leqslant \nu(X) < \infty$, and choose a sequence $\{f_n\} \subset \mathcal{F}$ such that $\int f_n\, d\mu \to a$. Let $g_n = \max(f_1, \ldots, f_n)$ and $f = \sup_n f_n$. Then $g_n \in \mathcal{F}, g_n$ increases pointwise to f, and $\int g_n\, d\mu \geqslant \int f_n\, d\mu$. It follows that $\lim \int g_n\, d\mu = a$ and hence, by the MCT, that $f \in \mathcal{F}$ and $\int f\, d\mu = a$. (In particular, $f < \infty$ a.e., so we may take f to be real-valued everywhere.)

We claim that the measure $d\lambda = d\nu - f\, d\mu$ (which is positive since $f \in \mathcal{H}$) is singular with respect to μ. If not, by Lemma 20 there exist $E \in \mathcal{M}$ and $\varepsilon > 0$ such that $\mu(E) > 0$ and $\lambda \geqslant \varepsilon\mu$ on E. But then $\varepsilon\chi_E\, d\mu \leqslant d\lambda = d\nu - f\, d\mu$, that is, $(f + e\chi_E)d\mu \leqslant d\nu$, so $f + \varepsilon\chi_E \in \mathcal{F}$ and $\int(f + \varepsilon\chi_E)d\mu = a + \varepsilon\mu(E) > a$, contradicting the definition of a.

Thus the existence of λ, f, and $d\rho = f\,d\mu$ is proved. As for uniqueness, if also $d\nu = d\lambda' + f'd\mu$, we have $d\lambda - d\lambda' = (f' - f)d\mu$. But $\lambda - \lambda' \perp \mu$ (see Folland Exercise 3.9), while $(f' - f)d\mu \ll d\mu$; hence $d\lambda - d\lambda' = (f' - f)d\mu = 0$, so that $\lambda = \lambda'$ and (by Proposition 39) $f = f'\mu$-a.e. Thus we are done in the case when μ and ν are finite measures.

Case II: Suppose that μ and ν are σ-finite measures. Then X is a countable disjoint union of μ-finite sets and a countable disjoint union of ν-finite sets; by taking intersections of these we obtain a disjoint sequence $\{A_j\} \subset \mathcal{M}$ such that $\mu(A_j)$ and $\nu(A_j)$ are finite for all j and $X = \bigcup_1^\infty A_j$. Define $\mu_j(E) = \mu(E \cap A_j)$ and $\nu_j(E) = \nu(E \cap A_j)$. By the reasoning above, for each j we have $d\nu_j = d\lambda_j + f_j\,d\mu_j$ where $\lambda_j \perp \mu_j$. Since $\mu_j(A_j^c) = \nu_j(A_j^c) = 0$, we have $\lambda_j(A_j^c) = \nu_j(A_j^c) - \int_{A_j^c} f\,d\mu_j = 0$, and we may assume that $f_j = 0$ on A_j^c. Let $\lambda = \sum_1^\infty \lambda_j$ and $f = \sum_1^\infty f_j$. Then $d\nu = d\lambda + f\,d\mu, \lambda \perp \mu$ (see Folland Exercise 3.9), and $d\lambda$ and $f\,d\mu$ are σ-finite, as desired. Uniqueness follows as before.

The General Case: If ν is a signed measure, we apply the preceding argument to ν^+ and ν^- and subtract the results. □

The decomposition $\nu = \lambda + \rho$ where $\lambda \perp \mu$ and $\rho \ll \mu$ is called the ***Lebesgue decomposition of ν with respect to μ***. In the case where $\nu \ll \mu$, Theorem 21 says that $d\nu = f\,d\mu$ for some f. This result is usually known as the ***Radon-Nikodym theorem***, and f is called the ***Radon-Nikodym derivative of ν with respect to μ***. We denote it by $d\nu/d\mu$:

$$d\nu = \frac{d\nu}{d\mu}d\mu.$$

(Strictly speaking, $d\nu/d\mu$ should be construed as the as the class of functions μ-a.e. equal to f.) The formulas suggested by the differential notation $d\mu/d\nu$ are generally correct. For example, it is obvious that $d(\nu_1 + \nu_2)/d\mu = (d\nu_1/d\mu) + (d\nu_2/d\mu)$, and we also have the chain rule:

Proposition 3.22: 3.9.

Suppose that ν is a σ-finite signed measure and μ, λ are σ-finite measures on $(X, \mathcal{M})$ such that $\nu \ll \mu$ and $\mu \ll \lambda$.

(a) If $g \in L^1(\nu)$, then $g \cdot \frac{d\nu}{d\mu} \in L^1(\mu)$ and

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

(b) We have $\nu \ll \lambda$, and

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda} \quad \lambda\text{-a.e.}$$

Proof. By considering ν^+ and ν^-separately, we may assume that $\nu \geqslant 0$. The equation

$\int g\,d\nu = \int g(d\nu/d\mu)d\mu$ is true when $g = \chi_E$ by definition of $d\nu/d\mu$. It is therefore true for simple functions by linearity, then for nonnegative measurable functions by the MCT, and finally for functions in $L^1(\nu)$ by linearity again. Replacing ν, μ by μ, λ and setting $g = \chi_E(d\nu/d\mu)$, we obtain

$$\nu(E) = \int_E \frac{d\nu}{d\mu}d\mu = \int_E \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}d\lambda$$

for all $E \in \mathcal{M}$, whence $(d\nu/d\lambda) = (d\nu/d\mu)(d\mu/d\lambda)\lambda$-a.e. by Proposition 39. □

Corollary 3.23: 3.10.

If $\mu \ll \lambda$ and $\lambda \ll \mu$, then $(d\lambda/d\mu)(d\mu/d\lambda) = 1$ a.e. (with respect to either λ or μ).

Example 24. *Non-example: Let μ be Lebesgue measure and ν the point mass at 0 on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$. Clearly $\nu \perp \mu$. The nonexistent Radon-Nikodym derivative $d\nu/d\mu$ is popularly known as the Dirac δ-function.*

We conclude this section with a simple but important observation, whose proof is straightforward:

Proposition 3.25: 3.11.

If $\mu_1, \ldots, \mu_n$ are measures on $(X, \mathcal{M})$, there is a measure μ such that $\mu_j \ll \mu$ for all j—namely, $\mu = \sum_1^n \mu_j$.

Exercise 3.26: Folland Exercise 3.9.

Suppose $\{\nu_j\}$ is a sequence of positive measures. If $\nu_j \perp \mu$ for all j, then $\sum_1^\infty \nu_j \perp \mu$; and if $\nu_j \ll \mu$ for all j, then $\sum_1^\infty \nu_j \ll \mu$.

Exercise 3.27: Folland Exercise 3.10.

Theorem 18 may fail when ν is not finite. (Consider $d\nu(x) = dx/x$ and $d\mu(x) = dx$ on $(0, 1)$, or ν is a counting measure and $\mu(E) = \sum_{n \in E} 2^{-n}$ on $\mathbb{Z}_{\geqslant 0}$.)

Example 28. *If $\mu \perp \nu$ and $\nu \perp \lambda$, is it true that $\mu \perp \lambda$? Of course not, since if $\mu = \lambda$ then this would mean $\mu \perp \mu$, and if μ then this fails.*

Example 29. *Even if $\mu \neq \lambda$ in Example 28, we still cannot say $\mu \perp \lambda$. Indeed, we can just take $\lambda \neq 0$ supported on any subset of the support of μ. Then $\nu \perp \lambda$, but μ and λ are not mutually singular.*

Remark 30. *There is a connection to calculus: Thinking about this new notion of an abstract derivative seems to be difficult at first, so let's convince ourselves we have seen it before without knowing so.*

- *In single-variable calculus, the Radon-Nikodym derivative appears when performing a change of variables since we 'pay the price' of a new function times the differential when doing substitution: dx become $g(u)\,du$. This g is indeed the Radon-Nikodym derivative telling us how the Lebesgue measures rates of change compare in each situation.*
- *If we consider multi-variable calculus change of variables, we see that the Radon-Nikodym derivative is the Jacobian determinant! Hence, this derivative is an abstract generalization of the change of variables from calculus.*

Remark 31 (Connection to Probability Theory)**.** *The Radon-Nikodym derivative gives the **density**, f, with respect to a reference measure μ, associated to the random variable X in the measurable space $(\mathcal{X}, \mathcal{X})$, defined by*

$$\mathbb{P}(X \in A) := \int_A f\,d\mu.$$

In the continuous univariate setting, the reference measure for the probability density function is taken as the Lebesgue measure. For discrete random variables, the probability mass function is the density with respect to the counting measure over the sample space (usually $\mathbb{Z}, \mathbb{Z}_{\geqslant 0}$, or some subset).

Example 32.

In $\mu = m$ (the Lebesgue measure) and ν is measure assigning twice the Lebesgue measure, then

$$\frac{d\nu}{dm} = 2$$

since

$$\nu(E) = \int_E 2\,dm.$$

Example 33. *If $\mu = m + \delta_0, \nu = m$, then $\nu \ll \mu$ and*

$$\frac{d\nu}{d\mu} = \chi_{(0,1]}$$

since we need to remove zero from being measured by μ:

$$1 = \nu([0,1]) = \int [0,1]\chi_{(0,1]} d\mu \neq \int_{[0,1]} d\mu = 1 + 1 = 2.$$

Exercise 3.34: Folland Exercise 3.11.

Let μ be a positive measure. A collection of functions $\{f_\alpha\}_{\alpha \in A} \subset L^1(\mu)$ is called ***uniformly integrable*** if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\left|\int_E f_\alpha\,d\mu\right| < \varepsilon$ for all $\alpha \in A$ whenever $\mu(E) < \delta$.

(a) Any finite subset of $L^1(\mu)$ is uniformly integrable.
(b) If $\{f_n\}$ is a sequence in $L^1(\mu)$ that converges in the L^1 metric to $f \in L^1(\mu)$, then

$\{f_n\}$ is uniformly integrable.

Solution.

(a) Let $\{f_1, \ldots, f_n\}$ be a finite subset of $L^1(\mu)$ and fix $\delta > 0$. Given $j \in \{1, \ldots, n\}$, by absolute continuity of the integral we can choose $\varepsilon_j > 0$ such that $\left|\int_E f_j \, d\mu\right| < \varepsilon_j$ whenever $\mu(E) < \delta$. Then choose $\varepsilon = \min\{\varepsilon_1, \ldots, \varepsilon_n\}$.

(b) Fix $\varepsilon > 0$ and suppose $f_n \to f$ in L^1. Then for all sufficiently large n,
$$\int |f_n - f| < \varepsilon.$$
Then there exists a finite subset $\{f_{n_1}, \ldots, f_{n_k}\}$ such that for all $j \in \{1, \ldots, k\}$,
$$\int |f_{n_j} - f| \geqslant \varepsilon.$$
Applying part (a) to the finite subset $\{|f_{n_j} - f|\}_{j=1}^k$, there exists $\delta > 0$ such that $\int_E |f_{n_j} - f| < \varepsilon_1$ whenever $\mu(E) < \delta$. Then
$$\int_E |f_{n_j} - f| < \varepsilon$$
for each $j \in \{1, \ldots, k\}$. Since these were the only exceptions, we conclude any L^1-convergent sequence in L^1 is uniformly integrable. □

Exercise 3.35: Folland Exercise 3.12.

For $j = 1, 2$, let μ_j, ν_j be σ-finite measures on $(X_j, \mathcal{M}_j)$ such that $\nu_j \ll \mu_j$. Then $\nu_1 \times \nu_2 \ll \mu_1 \times \mu_2$ and
$$\frac{d(\nu_1 \times \nu_2)}{d(\mu_1 \times \mu_2)}(x_1, x_2) = \frac{d\nu_1}{d\mu_1}(x_1) \frac{d\nu_2}{d\mu_2}(x_2). \tag{3.35.1}$$

Solution. As μ_j, ν_j are σ-finite for $j \in \{1, 2\}$, we know that $\mu_1 \times \mu_2, \nu_1 \times \nu_2$ are σ-finite.

We first show $\nu_1 \times \nu_2 \ll \mu_1 \times \mu_2$. Suppose $\mu_1 \times \mu_2(E) = 0$. Then $\chi_E \in L^+(\mu_1 \times \mu_2)$, so by Tonelli's Theorem,
$$0 = \mu_1 \times \mu_2(E) = \int \chi_E d(\mu_1 \times \mu_2) = \int \mu_1(E^{x_2}) d\mu_2.$$
Hence $\mu_1(E^{x_2}) = 0, \mu_2$-a.e. Since $\nu_j \ll \mu_j, j = 1, 2$, this implies $\nu_1(E^{x_2}) = 0$ ν_2-a.e. Therefore,
$$\nu_1 \times \nu_2(E) = \int \chi_E d(\nu_1 \times \nu_2) = \int \nu_2(E^{x_2}) d\nu_2 = 0,$$
showing $\nu_1 \times \nu_2 \ll \mu_1 \times \mu_2$.

To show Equation (3.35.1) holds, first note for $j \in \{1, 2\}$ that $d\nu_j / d\mu_j \geqslant 0$ μ_j-a.e., since if $d\nu_j/d\mu_j < 0$ on some F with $\mu_j(F) > 0$, then

$$\nu_j(F) = \int_F \frac{d\nu_j}{d\mu_j} d\mu_j < 0$$

contradicting ν_j is a positive measure. Then by Folland Exercise 2.51(a), $\frac{d\nu_1}{d\mu_1}(x_1) \cdot \frac{d\nu_2}{d\mu_2}(x_2) \in L^+(\mathcal{M}_1 \times \mathcal{M}_2)$, so by Tonelli's Theorem (applying the Radon-Nikodym Theorem twice),

$$\begin{aligned} \nu_1 \times \nu_2(E) &= \int_E d(\nu_1 \times \nu_2) = \int \left(\int \chi_E d\nu_1 \right) d\nu_2 = \int \left(\int \chi_E \frac{d\nu_1}{\mu_1} d\mu_1 \right) \frac{d\nu_2}{d\mu_2} d\mu_2 \\ &= \int \chi_E \frac{d\nu_1}{\mu_1}(x_1) \frac{d\nu_2}{d\mu_2}(x_2) d(\mu_1 \times \mu_2) = \int_E \frac{d\nu_1}{\mu_1}(x_1) \frac{d\nu_2}{d\mu_2}(x_2) d(\mu_1 \times \mu_2) \end{aligned}$$

whenever $E \in \mathcal{M}_1 \times \mathcal{M}_2$ Therefore,

$$\frac{d(\nu_1 \times \nu_2)}{d(\mu_1 \times \mu_2)}(x_1, x_2) = \frac{d\nu_1}{d\mu_1}(x_1) \frac{d\nu_2}{d\mu_2}(x_2), \quad (\mu_1 \times \mu_2)\text{-a.e.}$$

by Propositions 22 and 39 as the result holds for all $E \in \mathcal{M}_1 \times \mathcal{M}_2$. □

Exercise 3.36: Folland Exercise 3.13.

Let $X = [0,1], \mathcal{M} = \mathcal{B}_{[0,1]}, m =$ Lebesgue measure, and $\mu =$ counting measure on $\mathcal{M}$.

(a) $m \ll \mu$ but $dm \neq f\, d\mu$ for any f.

(b) μ has no Lebesgue decomposition with respect to m.

Solution.

(a) If $\mu(E) = 0$ then $E = \varnothing$, so $m(E) = m(\varnothing) = 0$. On the other hand, if $dm = f\, d\mu$ for some extended μ-integrable function $f\colon [0,1] \to \mathbb{R}$, then

$$0 = m(\{x\}) = \int_{\{x\}} f(t)\, d\mu(t) = f(x).$$

Thus $f \equiv 0$. But then

$$1 = m([0,1]) = \int_{[0,1]} 0\, d\mu = 0,$$

a contradiction.

(b) Suppose there exist signed measures μ_s, μ_{ac} such that $\mu_s \perp m$, $\mu_{ac} \ll m$, and $\mu = \mu_s + \mu_{ac}$. Since $\mu_s \perp m$, there exists $E \in \mathcal{L}([0,1])$ such that for all $F \subset E$, $m(F) = 0$ and for all $F \subset E^c$, $\mu(F) = 0$. But $\mu_s \ll m$, so $\mu_s(F) = 0$ too. Since μ is the counting measure, any $x \in F$ has $m(\{x\}) = 0$, so $\mu_s(\{x\}) = 1$, which means $F = \varnothing$.

But then $\mu = \mu_{ac}$, contradicting $\mu(\{x\}) = 1$, since $\mu(\{x\}) = \mu_{ac}(\{x\}) = 0$ (because $\mu_{ac} \ll m$ and $m(\{x\}) = 0 \implies \mu_{ac}(\{x\}) = 0$). □

Exercise 3.37: Folland Exercise 3.16.

Suppose that μ, ν are measures on $(X, \mathcal{M})$ with $\nu \ll \mu$, and let $\lambda = \mu + \nu$. If $f = d\nu/d\lambda$, then $0 \leqslant f < 1$ μ-a.e. and $d\nu/d\mu = f/(1-f)$.

Solution. $f = d\nu/d\lambda$, then $0 \leqslant f < 1\mu$-a.e. and $d\nu/d\mu = f/(1-f)$. μ, ν positive measures, so $\mu + \nu = 0 \Rightarrow \nu = \mu = 0$, so $r \ll \lambda, \mu \ll \lambda$. Also $\lambda \ll \mu$, because $\mu = 0 \overset{\text{because}}{\underset{}{\overset{\lambda \ll \mu}{\Longrightarrow}}} \nu = 0$, which in turn implies $\lambda = \mu + \nu = 0$. Hence $d\nu/d\mu, d\lambda/d\mu, d\nu/d\lambda$, and $d\lambda/d\nu$ exist.

Since $\lambda = \mu + \nu$, by additivity and the chain rule we have

$$f = \frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda} = \frac{d\nu}{d\mu}\Big(1 - \underbrace{\frac{d\nu}{d\lambda}}_{=f}\Big),$$

so $d\nu/d\mu = 1/(1-f)$. Then $0 \leqslant f < 1$ a.e., since if $f \geqslant 1$ on a positive measure set then $d\nu/d\mu$ is undefined on a positive measure set, contradicting Section 3.2, and if $f < 1$ on a positive measure set E, then

$$0 < \nu(E) = \int_E \frac{d\mu}{d\mu}\, d\mu < 0,$$

a contradiction. □

Exercise 3.38: Folland Exercise 3.17.

Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space, $\mathcal{N}$ a σ-finite σ-subalgebra of $\mathcal{M}$, and $\nu = \mu|_{\mathcal{N}}$. If $f \in L^1(\mu)$, there exists $g \in L^1(\nu)$ (thus g is $\mathcal{N}$-measurable) such that $\int_E f\, d\mu = \int_E g\, d\nu$ for all $E \in \mathcal{N}$; if g' is another such function then $g = g'$ ν-a.e. (In probability theory, g is called the ***conditional expectation*** of f on $\mathcal{N}$.)

Solution. $\nu = \mu|_{\mathcal{N}}$, so $\mu \ll \nu$. On the other hand, define $\lambda \colon \mathcal{N} \to [-\infty, \infty]$ by

$$\lambda(E) := \int_E f\, d\mu.$$

Then λ is finite since $f \in L^1(\mu)$, so λ is σ-finite. And $\lambda \ll \nu$, since

$$\nu(E) = 0 \implies \mu(E) = 0 \implies \lambda(E) = \int_E f\, d\mu \leqslant \overset{0}{\cancel{\mu(E)}} \sup_{x \in E} |f(x)| = 0.$$

Thus the Radon-Nikodym derivative $g := d\lambda/d\nu$ exists ν-a.e., so for all $E \in \mathcal{N}$,

$$\int_E f\, d\mu = \lambda(E) = \int_E g\, d\nu,$$

as desired. And if g' is another such function, then by the Radon-Nikodym Theorem $g' = g$ ν-a.e. □

3.3 Differentiation Theory on Euclidean Space

The Radon-Nikodym theorem provides an abstract notion of the "derivative" of a signed measure ν with respect to a measure μ. In this section we analyze more deeply the special case where $(X, \mathcal{M}, \mu)$ is the Lebesgue measure space $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$. Here one can define a pointwise derivative of ν with respect to m in the following way. Let $B_r(x)$ be the open ball of radius r about x in $\mathbb{R}^n$; then one can consider the limit

$$F(x) = \lim_{r \to 0} \frac{\nu(B_r(x))}{m(B_r(x))}$$

when it exists. One can also replace the balls $B_r(x)$ by other sets which, in a suitable sense, shrink to x in a regular way; we shall examine this point later.)

Remark 39. *If $\nu \ll m$, so that $d\nu = f\,dm$, then $\nu(B_r(x))/m(B_r(x))$ is simply the average value of f on $B_r(x)$.*

By Note 39, one would hope that $F = f$ m-a.e. This turns out to be the case, provided that $\nu(B_r(x)) < \infty$ for all r and x. From the point of view of the function f, this may be regarded as a generalization of the fundamental theorem of calculus: The derivative of the indefinite integral of f (namely, ν) is f.

For the remainder of this section, terms such as "integrable" and "almost everywhere" refer to the Lebesgue measure unless otherwise specified. We begin our analysis with a technical lemma that is of interest in its own right.

Lemma 3.40: 3.15.

Let $\mathcal{C}$ be a collection of open balls in $\mathbb{R}^n$, and let $U = \bigcup_{B \in \mathcal{C}} B$. If $c < m(U)$, there exist disjoint $B_1, \ldots, B_k \in \mathcal{C}$ such that $\sum_1^k m(B_j) > 3^{-n}c$.

Proof. If $c < m(U)$, by Theorem 82 there is a compact $K \subset U$ with $m(K) > c$, and finitely many of the balls in $\mathcal{C}$—say, $A_1, \ldots, A_m$—cover K. Let B_1 be the largest of the A_js (that is, choose B_1 to have maximal radius), let B_2 be the largest of the A_js that are disjoint from B_1, let B_3 be the largest of the A_js that are disjoint from B_1 and B_2, and so on, until the list of A_js is exhausted. According to this construction, if A_i is not one of the B_js, there is some j such that $A_i \cap B_j \neq \varnothing$, and if j is the smallest integer with this property, the radius of A_i is at most that of B_j. Hence $A_i \subset B_j^*$, where B_j^* is the ball concentric with B_j whose radius is three times that of B_j. But then $K \subset \bigcup_1^k B_j^*$, so

$$c < m(K) \leqslant \sum_1^k m(B_j^*) = 3^n \sum_1^k m(B_j). \qquad \square$$

Definition 41. *A measurable function $f : \mathbb{R}^n \to \mathbb{C}$ is called* ***locally integrable (with respect to the Lebesgue measure)*** *if $\int_K |f(x)|\,dx < \infty$ for every bounded measurable set $K \subset \mathbb{R}^n$.*

We denote the space of locally integrable functions by L^1_{loc}. If $f \in L^1_{\text{loc}}, x \in \mathbb{R}^n$, and $r > 0$, we define $A_r f(x)$ to be the average value of f on $B_r(x)$:

$$A_r f(x) = \frac{1}{m(B_r(x))} \int_{B_r(x)} f(y)\, dy.$$

Lemma 3.42: 3.16.

If $f \in L^1_{\text{loc}}, A_r f(x)$ is jointly continuous in r and $x (r > 0, x \in \mathbb{R}^n)$.

Proof. From the results in Folland Section 2.7, we know that $m(B_r(x)) = cr^n$ where $c = m(B_1(0))$, and $m(S(r,x)) = 0$ where $S(r,x) = \{y \mid |y - x| = r\}$. Moreover, as $r \to r_0$ and $x \to x_0, \chi_{B_r(x)} \to \chi_{B(r_0,x_0)}$ pointwise on $\mathbb{R}^n \setminus S(r_0,x_0)$. Hence $\chi_{B_r(x)} \to \chi_{B(r_0,x_0)}$ a.e., and $|\chi_{B_r(x)}| \leqslant \chi_{B(r_0+1,x_0)}$ if $r < r_0 + \frac{1}{2}$ and $|x - x_0| < \frac{1}{2}$. By the DCT, it follows that $\int_{B_r(x)} f(y)\, dy$ is continuous in r and x, and hence so is $A_r f(x) = c^{-1} r^{-n} \int_{B_r(x)} f(y)\, dy$. □

Definition 43. *If $f \in L^1_{\text{loc}}$, we define its* ***Hardy-Littlewood maximal function*** *Hf by*

$$Hf(x) = \sup_{r>0} A_r|f|(x).$$

Hf is measurable, for $(Hf)^{-1}((a,\infty)) = \bigcup_{r>0} (A_r|f|)^{-1}((a,\infty))$ is open for any $a \in \mathbb{R}$, by Lemma 42.

Theorem 3.44: 3.17: The Maximal Theorem.

There exists a constant $C > 0$ such that for all $f \in L^1$ and all $\alpha > 0$,

$$m(\{Hf > \alpha\}) \leqslant \frac{C}{\alpha} \int |f(x)|\, dx.$$

Proof. Let $E_\alpha = \{Hf > \alpha\}$. For each $x \in E_\alpha$, we can choose $r_x > 0$ such that $A_{r_x}|f|(x) > \alpha$. The balls $B_{r_x}(x)$ cover E_α, so by Lemma 40, if $c < m(E_\alpha)$ there exist $x_1, \dots, x_k \in E_\alpha$ such that the balls $B_j := B_{r_{x_j}}(x_j)$ are disjoint and $\sum_1^k m(B_j) > 3^{-n} e$. But then

$$c < 3^n \sum_1^k m(B_j) \leqslant \frac{3^n}{\alpha} \sum_1^k \int_{B_j} |f(y)|\, dy \leqslant \frac{3^n}{\alpha} \int_{\mathbb{R}^n} |f(y)|\, dy.$$

Letting $c \to m(E_\alpha)$, we obtain the desired result. □

With this tool in hand, we now present three successively sharper versions of the fundamental differentiation theorem. In the proofs we shall use the notion of ***limit superior for real-valued functions of a real variable***,

$$\limsup_{r \to R} \phi(r) = \lim_{\varepsilon \to 0} \sup_{0<|r-R|<\varepsilon} \phi(r) = \inf_{\varepsilon > 0} \sup_{0<|r-R|<\varepsilon} \phi(r),$$

and the easily verified fact that

$$\lim_{r \to R} \phi(r) = c \iff \limsup_{r \to R} |\phi(r) - c| = 0.$$

(This is shown in disguise in Folland Exercise 2.23.)

Theorem 3.45: 3.18.

If $f \in L^1_{\text{loc}}$, then $\lim_{r\to 0} A_r f(x) = f(x)$ for a.e. $x \in \mathbb{R}^n$.

Proof. It suffices to show that for $N \in \mathbb{Z}_{\geqslant 0}$, $A_r f(x) \to f(x)$ for a.e. x with $|x| \leqslant N$. But for $|x| \leqslant N$ and $r \leqslant 1$ the values $A_r f(x)$ depend only on the values $f(y)$ for $|y| \leqslant N+1$, so by replacing f with $f\chi_{B_{N+1}(0)}$ we may assume that $f \in L^1$.

Given $\varepsilon > 0$, by Theorem 83 we can find a continuous integrable function g such that $\int |g(y) - f(y)|\,dy < \varepsilon$. Continuity of g implies that for every $x \in \mathbb{R}^n$ and $\delta > 0$ there exists $r > 0$ such that $|g(y) - g(x)| < \delta$ whenever $|y - x| < r$, and hence

$$|A_r g(x) - g(x)| = \frac{1}{m(B_r(x))}\left|\int_{B_r(x)} [g(y) - g(x)]\,dy\right| < \delta.$$

Therefore $A_r g(x) \to g(x)$ as $r \to 0$ for every x, so

$$\begin{aligned}\limsup_{r\to 0}|A_r f(x) - f(x)| &= \limsup_{r\to 0}|A_r(f-g)(x) + (A_r g - g)(x) + (g-f)(x)| \\ &\leqslant H(f-g)(x) + 0 + |f-g|(x).\end{aligned}$$

Hence, if

$$E_\alpha = \{\limsup_{r\to 0}|A_r f - f| > \alpha\} \quad \text{and} \quad F_\alpha = \{|f-g| > \alpha\},$$

then

$$E_\alpha \subset F_{\alpha/2} \cup \{H(f-g) > \alpha/2\}.$$

But $(\alpha/2)m(F_{\alpha/2}) \leqslant \int_{F_{\alpha/2}} |f(x) - g(x)|\,dx < \varepsilon$, so by the maximal theorem,

$$m(E_\alpha) \leqslant \frac{2\varepsilon}{\alpha} + \frac{2C\varepsilon}{\alpha}.$$

Since ε is arbitrary, $m(E_\alpha) = 0$ for all $\alpha > 0$. But $\lim_{r\to 0} A_r f(x) = f(x)$ for all $x \notin \bigcup_1^\infty E_{1/n}$, so we are done. □

This result can be rephrased as follows: If $f \in L^1_{\text{loc}}$,

$$\lim_{r\to 0} \frac{1}{m(B_r(x))}\int_{B_r(x)} [f(y) - f(x)]\,dy = 0 \text{ for a.e. } x. \tag{3.45.1}$$

Actually, something stronger is true: Equation (3.45.1) remains valid if one replaces the integrand by its absolute value. That is, let us define the ***Lebesgue set*** L_f of f to be

$$L_f = \left\{x \in X \;\middle|\; \lim_{r\searrow 0} \frac{1}{m(B_r(x))}\int_{B_r(x)} |f(y) - f(x)|\,dy = 0\right\}.$$

Then the following theorem holds.

Theorem 3.46: 3.20.

If $f \in L^1_{\text{loc}}$, then $m((L_f)^c) = 0$.

Proof. For each $c \in \mathbb{C}$ we can apply Theorem 45 to $g_c(x) = |f(x) - c|$ to conclude that, except on a Lebesgue null set E_c, we have

$$\lim_{r\to 0} \frac{1}{m(B_r(x))} \int_{B_r(x)} |f(y) - c|\, dy = |f(x) - c|.$$

Let D be a countable dense subset of $\mathbb{C}$, and let $E = \bigcup_{c\in D} E_c$. Then $m(E) = 0$, and if $x \notin E$, for any $\varepsilon > 0$ we can choose $c \in D$ with $|f(x) - c| < \varepsilon$, so that $|f(y) - f(x)| < |f(y) - c| + \varepsilon$, and it follows that

$$\limsup_{r\to 0} \frac{1}{m(B_r(x))} \int_{B_r(x)} |f(y) - f(x)|\, dy \leqslant |f(x) - c| + \varepsilon < 2\varepsilon.$$

Since ε is arbitrary, the desired result follows. □

Finally, we consider families of sets more general than balls.

Definition 47. *A family $\{E_r\}_{r>0}$ of Borel subsets of $\mathbb{R}^n$ is said to* ***shrink nicely*** *to $x \in \mathbb{R}^n$ if*

- *$E_r \subset B_r(x)$ for each r;*
- *there exists a constant $\alpha > 0$ such that $m(E_r) > \alpha m(B_r(x))$.*

Remark 48. *The sets E_r in Definition 47 need not contain x itself. For example, if U is any Borel subset of $B_1(0)$ such that $m(U) > 0$, and $E_r = \{x + ry \mid y \in U\}$, then $\{E_r\}$ shrinks nicely to x.*

Here, then, is the final version of the differentiation theorem.

Theorem 3.49: 3.21: The Lebesgue Differentiation Theorem (LDT).

Suppose $f \in L^1_{\text{loc}}$. For every x in the Lebesgue set of f—in particular, for m-a.e. $x \in \mathbb{R}^n$—we have

$$\lim_{r\to 0} \frac{1}{m(E_r)} \int_{E_r} |f(y) - f(x)|\, dy = 0 \quad \text{and} \quad \lim_{r\to 0} \frac{1}{m(E_r)} \int_{E_r} f(y)\, dy = f(x)$$

for every family $\{E_r\}_{r>0}$ that shrinks nicely to x.

Proof. For some $\alpha > 0$ we have

$$\begin{aligned}\frac{1}{m(E_r)} \int_{E_r} |f(y) - f(x)|\, dy &\leqslant \frac{1}{m(E_r)} \int_{B_r(x)} |f(y) - f(x)|\, dy \\ &\leqslant \frac{1}{\alpha m(B_r(x))} \int_{B_r(x)} |f(y) - f(x)|\, dy.\end{aligned}$$

The first equality therefore follows from Theorem 46, and one sees immediately that it implies the second one by writing the latter in the form Equation (3.45.1). □

We now return to the study of measures.

Definition 50. *A Borel measure ν on $\mathbb{R}^n$ will be called* ***regular*** *if*

(i) $\nu(K) < \infty$ *for every compact* K*;*
(ii) $\nu(E) = \inf\{\nu(U) \mid U$ *open,* $E \subset U\}$ *for every* $E \in \mathcal{B}_{\mathbb{R}^n}$*.*

Remark 51. *Condition (ii) is actually implied by condition (i). For* $n = 1$ *this follows from Theorems 43 and 45, and the proof of this for arbitrary* n *can be found in Folland Section 7.2. For the time being, we assume (ii) explicitly.*

We observe that by (i), every regular measure is σ-finite. A signed Borel measure ν will be called ***regular*** if $|\nu|$ is regular.

Proposition 3.52.

If $f \in L^+(\mathbb{R}^n)$, the measure

$$f\,dm \text{ is regular} \qquad \iff \qquad f \in L^1_{\text{loc}}.$$

Proof. Indeed, the condition $f \in L^1_{\text{loc}}$ is clearly equivalent to (i) in Definition 50. If this holds, (ii) in Definition 50 may be verified directly as follows. Suppose that E is a bounded Borel set. Given $\delta > 0$, by Theorem 82 there is a bounded open $U \supset E$ such that $m(U) < m(E) + \delta$ and hence $m(U \setminus E) < \delta$. But then, given $\varepsilon > 0$, by Corollary 19 there is an open $U \supset E$ such that $\int_{U \setminus E} f\,dm < \varepsilon$ and hence $\int_U f\,dm < \int_E f\,dm + \varepsilon$. The case of unbounded E follows easily by writing $E = \bigcup_1^\infty E_j$ where E_j is bounded and finding an open $U_j \supset E_j$ such that $\int_{U_j \setminus E_j} f\,dm < \varepsilon 2^{-j}$. □

Theorem 3.53: 3.22.

Let ν be a regular signed or complex Borel measure on $\mathbb{R}^n$, and let $d\nu = d\nu_s + f\,dm$ be its Lebesgue-Radon-Nikodym representation. Then for m-almost every $x \in \mathbb{R}^n$.

$$\lim_{r \to 0} \frac{\nu(E_r)}{m(E_r)} = f(x)$$

for every family $\{E_r\}_{r>0}$ that shrinks nicely to x.

Proof. It is easily verified that $d|\nu| = d|\nu_s| + |f|dm$, so the regularity of ν implies the regularity of both ν_s and $f\,dm$ (Folland Exercise 3.26). In particular, $f \in L^1_{\text{loc}}$ by Proposition 52, so in view of Theorem 49, it suffices to show that if ν_s is regular and $\nu_s \perp m$, then for m-almost every x, $\nu_s(E_r)/m(E_r) \to 0$ as $r \to 0$ when E_r shrinks nicely to x. It also suffices to take $E_r = B_r(x)$ and to assume that ν_s is positive, since for some $\alpha > 0$ we have

$$\left|\frac{\nu_s(E_r)}{m(E_r)}\right| \leqslant \frac{|\nu_s|(E_r)}{m(E_r)} \leqslant \frac{|\nu_s|(B_r(x))}{m(E_r)} \leqslant \frac{|\nu_s|(B_r(x))}{\alpha m(B_r(x))}.$$

Assuming $\nu_s \geqslant 0$, then, let A be a Borel set such that $\nu_s(A) = m(A^c) = 0$, and let

$$F_k = \left\{x \in A \;\middle|\; \limsup_{r \to 0} \frac{\nu_s(B_r(x))}{m(B_r(x))} > \frac{1}{k}\right\}.$$

We shall show that $m(F_k) = 0$ for all k, and this will complete the proof.

The argument is similar to the proof of the maximal theorem. By regularity of ν_s, given $\varepsilon > 0$ there is an open $U_\varepsilon \supset A$ such that $\nu_s(U_\varepsilon) < \varepsilon$. Each $x \in F_k$ is the center of a ball $B_x \subset U_\varepsilon$ such that $\nu_s(B_x) > k^{-1}m(B_x)$. By Lemma 40, if $V_\varepsilon = \bigcup_{x\in F_k} B_x$ and $c < m(V_\varepsilon)$ there exist $x_1, \dots, x_J$ such that $B_{x_1}, \dots, B_{x_J}$ are disjoint and

$$c < 3^n \sum_1^J m(B_{x_j}) \leqslant 3^n k \sum_1^J \nu_s(B_{x_j}) \leqslant 3^n k\nu_s(V_\varepsilon) \leqslant 3^n k\nu\nu_s(U_\varepsilon) \leqslant 3^n k\varepsilon.$$

We conclude that $m(V_\varepsilon) \leqslant 3^n k\varepsilon$, and since $F_k \subset V_\varepsilon$ and ε is arbitrary, $m(F_k) = 0$. □

Exercise 3.54: Folland Exercise 3.22.

If $f \in L^1(\mathbb{R}^n), f \neq 0$, there exist $C, R > 0$ such that $Hf(x) \geqslant C|x|^{-n}$ for $|x| > R$. Hence $m(\{Hf > \alpha\}) \geqslant C'/\alpha$ when α is small, so the estimate in the maximal theorem is essentially sharp. [a]

[a] *Hint: estimate* $(A_{2|x|}|f|)(x)$.

Solution. Since $f \neq 0$, there exists $R > 1$ such that

$$\int_{B_R(0)} |f(y)|\, dy > \varepsilon > 0$$

for some ε. Then for all $|x| \geqslant R$,

$$\begin{aligned} Hf(x) \geqslant A_{2|x|}|f|(x) &\geqslant \frac{1}{m(B_{2|x|}(x))} \int_{B_{2|x|}(0)} |f(y)|\, dy \\ &\geqslant \frac{1}{m(B_{2|x|}(x))} \underbrace{\int_{B_R(0)} |f(y)|\, dy}_{>\varepsilon} \\ &> \frac{\varepsilon}{(2|x|)^n m(B_1(0))} = \underbrace{\frac{\varepsilon}{2^n m(B_1(0))}}_{=:C} |x|^{-n}. \end{aligned}$$

Thus $Hf(x) \geqslant C|x|^{-n}$ for $|x| > R$.

This shows the estimate in the maximal theorem is essentially sharp, because for sufficiently small positive α we have

$$m\{Hf > \alpha\} \geqslant m\{x \in \mathbb{R}^n \mid C|x|^{-n} > \alpha\} = m\left\{x \in \mathbb{R}^n \,\middle|\, |x| < \left(\frac{C}{\alpha}\right)^{1/n}\right\} = \frac{C}{\alpha} m(B_1(0)).$$

□

Exercise 3.55: Folland Exercise 3.23.

A useful variant of the Hardy-Littlewood maximal function is

$$H^* f(x) = \sup\left\{\frac{1}{m(B)} \int_B |f(y)|\, dy \,\middle|\, B \text{ is a ball and } x \in B\right\}.$$

Show that $Hf \leqslant H^*f \leqslant 2^n Hf$.

Solution. Fix $x \in \mathbb{R}^n$, let S be the collection of open balls containing x, let let T be the collection of open balls *centered at* x, and for all Lebesgue measurable subsets E of $\mathbb{R}^n$ define

$$A_E|f| := \frac{1}{m(E)} \int_E |f(y)|\, dy.$$

Then $T \subset S$, so then

$$Hf(x) = \sup_{E \in T} A_E|f| \leqslant \sup_{E \in S} A_E|f| = H^*f(x).$$

For the other inequality, let B_r be any ball containing x, say of radius r. Then $B \subset B_{2r}(x)$, so

$$\frac{1}{m(B_r)} \int_{B_r} |f(y)|\, dy \leqslant \frac{m(B_{2r}(x))}{m(B_r)} \frac{1}{m(B_{2r}(x))} \int_{B_{2r}(x)} |f(y)| dy \leqslant 2^n Hf(x)$$

Since B was any ball containing x, taking the supremum over all such balls shows that $H^*f(x) \leqslant 2^n Hf(x)$. □

Exercise 3.56: Folland Exercise 3.24.

If $f \in L^1_{\text{loc}}$ and f is continuous at x, then x is in the Lebesgue set of f.

Solution. Let $\varepsilon > 0$. Since f is continuous, we can choose $\delta > 0$ such that $\|f(x) - f(y)\| < \varepsilon$ whenever $\|x - y\| < \delta$. Then for all $y \in B_\delta(x)$,

$$\frac{1}{m(B_r(x))} \int_{B_r(x)} \|f(y) - f(x)\|\, dy < \frac{\varepsilon m(B_r(x))}{m(B_r(x))} = \varepsilon.$$

Since ε was arbitrary, $\frac{1}{m(B_r(x))} \int_{B_r(x)} \|f(y) - f(x)\|\, dy \to 0$ as $r \searrow 0$. Hence x is in the Lebesgue set of f. □

Exercise 3.57: Folland Exercise 3.25.

If E is a Borel set in $\mathbb{R}^n$, the ***density*** $D_E(x)$ of E at x is defined as

$$D_E(x) = \lim_{r \to 0} \frac{m(E \cap B_r(x))}{m(B_r(x))}$$

whenever the limit exists.

(a) Show that $D_E(x) = 1$ for a.e. $x \in E$ and $D_E(x) = 0$ for a.e. $x \in E^c$.
(b) Find examples of E and x such that $D_E(x)$ is a given number $\alpha \in (0, 1)$, or such that $D_E(x)$ does not exist.

Solution. We only solve (a) and leave (b) as an exercise. Since $E \in \mathcal{B}_{\mathbb{R}^n}$, χ_E is measurable, so $\chi_E \in L^1_{\text{loc}}$. (Indeed, if K is any compact set of $\mathbb{R}^n$ then $\int_K \chi_E\, dm = m(K \cap E) \leqslant$

$m(K) < \infty$). Then by the LDT, we have for a.e. $x \in \mathbb{R}^n$ that

$$\lim_{r \searrow 0} A_r \chi_E(x) = \chi_E(x) = \begin{cases} 0 & \text{if } x \in E^c, \\ 1 & \text{if } x \in E, \end{cases}$$

so since the left-hand side is just the definition of D_E, we are done. □

3.4 Functions of Bounded Variation

All functions in this section are to be assumed Lebesgue measurable unless otherwise stated. The theorems of the preceding section apply in particular on the real line, where, because of the correspondence between regular Borel measures and increasing functions that we established in Folland Section 1.5, they yield results about differentiation and integration of functions. As in Folland Section 1.5, we adopt the notation that if F is an increasing, right continuous function on $\mathbb{R}$, μ_F is the Borel measure determined by the relation $\mu_F((a, b]) = F(b) - F(a)$. Also, throughout this section the term "almost everywhere" will always refer to the Lebesgue measure.

Our first result uses the Lebesgue differentiation theorem to prove the a.e. differentiability of increasing functions.

Theorem 3.58: 3.23.

Let $F\colon \mathbb{R} \to \mathbb{R}$ be increasing, and let $G(x) = F(x+)$.

(a) The set of points at which F is discontinuous is countable.

(b) F and G are differentiable a.e., and $F' = G'$ a.e.

Proof. Since F is increasing, the intervals $(F(x-), F(x+))(x \in \mathbb{R})$ are disjoint, and for $|x| < N$ they lie in the interval $(F(-N), F(N))$. Hence

$$\sum_{|x|<N} [F(x+) - F(x-)] \leqslant F(N) - F(-N) < \infty,$$

which implies that $\{x \in (-N, N) \mid F(x+) \neq F(x-)\}$ is countable. As this is true for all N, (a) is proved.

Next, we observe that G is increasing and right continuous, and $G = F$ except perhaps where F is discontinuous. Moreover,

$$G(x + h) - G(x) = \begin{cases} \mu_G((x, x + h]) & \text{if } h > 0, \\ -\mu_G((x + h, x]) & \text{if } h < 0, \end{cases}$$

and the families $\{(x - r, x]\}$ and $\{(x, x + r]\}$ shrink nicely to x as $r = |h| \to 0$. Thus, an application of Theorem 53 to the measure μ_G (which is regular by Theorem 45) shows that $G'(x)$ exists for a.e. x. To complete the proof, it remains to show that if $H = G - F$, then H' exists and equals zero a.e.

Let $\{x_j\}$ be an enumeration of the points at which $H \neq 0$. Then $H(x_j) > 0$, and as above we have $\sum_{\{j \mid |x_j|<N\}} H(x_j) < \infty$ for any N. Let δ_j be the point mass at x_j and

$\mu = \sum_j H(x_j)\delta_j$. Then μ finite on compact sets by the preceding sentence, and hence μ is regular by Theorems 43 and 45; also, $\mu \perp m$ since$m(E) = \mu(E^c) = 0$ where $E = \{x_j\}_1^\infty$. But then

$$\left|\frac{H(x+h) - H(x)}{h}\right| \leqslant \frac{H(x+h) + H(x)}{|h|} \leqslant 4\frac{\mu((x-2|h|, x+2|h|))}{4|h|},$$

which tends to zero as $h \to 0$ for a.e. x, by Theorem 53. Thus $H' = 0$ a.e., and we are done. □

As positive measures on $\mathbb{R}$ are related to increasing functions, complex measures on $\mathbb{R}$ are related to so-called functions of bounded variation. The definition of the latter concept is a bit technical, so some motivation may be appropriate. Intuitively, if $F(t)$ represents the position of a particle moving along the real line at time t, the "total variation" of F over the interval $[a, b]$ is the total distance traveled from time a to time b, as shown on an odometer. If F has a continuous derivative, this is just the integral of the "speed," $\int_a^b |F'(t)|dt$. To define the total variation without any smoothness hypotheses on F requires a different approach; namely, one partitions $[a, b]$ into subintervals $[t_{j-1}, t_j]$ and approximates F on each subinterval by the linear function whose graph joins $(t_{j-1}, F(t_{j-1}))$ to $(t_j, F(t_j))$, and then passes to a limit.

In making this precise, we begin with a slightly different point of view, taking $a = -\infty$ and considering the total variation as a function of b.

Definition 59. *If $f\colon \mathbb{R} \to \mathbb{R}$ and $x \in \mathbb{R}$, we define the* ***total varition function of F*** *by*

$$T_F(x) := \sup\left\{\sum_1^n |F(x_j) - F(x_{j-1})| \;\middle|\; n \in \mathbb{Z}_{\geqslant 0}, -\infty < x_0 < \cdots < x_n = x\right\}.$$

Define the collection of ***functions of bounded variation on*** $\mathbb{R}$ *by*

$$\mathrm{BV} := \left\{\textit{set functions } F\colon \mathbb{R} \to \mathbb{C} \;\middle|\; \lim_{x\to\infty} T_F(x) < \infty\right\}.$$

If $a < b$, we call the quantity $T_F(a) - T_F(b)$ the ***total variation of F on $[a, b]$****, that is,*

$$T_F(b) - T_F(a) = \sup\left\{\sum_{j=1}^n |F(x_j) - F(x_{j-1})| \;\middle|\; n \in \mathbb{Z}_{\geqslant 0}, a = x_0 < \cdots < x_n = b\right\}.$$

It depends only on the values of F on $[a, b]$, so we may define $\mathrm{BV}([a, b])$ to be the set of all functions on $[a, b]$ whose total variation on $[a, b]$ is finite.

If $F \in \mathrm{BV}$, the restriction of F to $[a, b]$ is in $\mathrm{BV}([a, b])$ for all a, b; indeed, its total variation on $[a, b]$ is nothing but $T_F(b) - T_F(a)$. Conversely, if $F \in \mathrm{BV}([a, b])$ and we set $F(x) = F(a)$ for $x < a$ and $F(x) = F(b)$ for $x > b$, then $F \in \mathrm{BV}$. By this device the results that we shall prove for BV can also be applied to $\mathrm{BV}([a, b])$.

Example 60 (3.25)**.**

(a) If $F\colon \mathbb{R} \to \mathbb{R}$ is bounded and increasing, then $F \in \mathrm{BV}$ (in fact, $T_F(x) = F(x) - F(-\infty)$).

(b) If $F, G \in \mathrm{BV}$ and $a, b \in \mathbb{C}$, then $aF + bG \in \mathrm{BV}$.

(c) If F is differentiable on $\mathbb{R}$ and F' is bounded, then $F \in \mathrm{BV}([a,b])$ for $-\infty < a < b < \infty$ (by the mean value theorem).

The verification of these examples is left as an exercise (Folland Exercise 3.27).

Exercise 3.61.

If $F(x) = \sin x$, then $F \in \mathrm{BV}([a,b])$ for $-\infty < a < b < \infty$, but $F \notin \mathrm{BV}$.

Solution. $F \in \mathrm{BV}([a,b])$ by the mean value theorem and Exercise 61, since $F'(x) = \cos x$ is bounded. To see $F \notin \mathrm{BV}$, let $x_n = \pi(2n+1)/2, n \in \mathbb{Z}_{\geqslant 0}$ to see $\sum_{n=1}^{N} |\sin x_{n+1} - \sin x_n| = 2N$. Hence, $T_F(x) \geqslant 2N$ for all $x \in \mathbb{R}, N \in \mathbb{Z}_{\geqslant 0}$, showing that $\lim_{x\to\infty} T_F(x) = \infty$. □

Exercise 3.62: Folland Exercise 3.27.

Verify the assertions in Example 60.

Solution. Suppose F is continuous on $[a,b]$ and F' is bounded on $[a,b]$. Then there exists M such that $|F| \leqslant M$ on $[a,b]$. By the mean value theorem, for all $[x,y] \subset [a,b]$, there exists $c \in (a,b)$ such that

$$M \geqslant F'(c) = \frac{F(x) - F(y)}{x - y}.$$

Thus $F(x) - F(y) \leqslant M|x-y|$. Then any partition of the real line has

$$\sum\nolimits_1^n |F(x_j) - F(x_{j-1})| \leqslant |M| \sum\nolimits_1^n |x_j - x_{j-1}| = |M|(b-a),$$

Taking the supremum of both sides over all partitions of (a,b), we conclude $T_F \leqslant M(b-a) < \infty$, so $F \in \mathrm{BV}$. The rest of the verifications are left as exercises. □

Lemma 3.63: 3.26.

If $F \in \mathrm{BV}$ is real-valued, then $T_F + F$ and $T_F - F$ are increasing.

Proof. If $x < y$ and $\varepsilon > 0$, choose $x_0 < \cdots < x_n = x$ such that

$$\sum\nolimits_1^n |F(x_j) - F(x_{j-1})| \geqslant T_F(x) - \varepsilon.$$

Then $\sum_1^n |F(x_j) - F(x_{j-1})| + |F(y) - F(x)|$ is an approximating sum for $T_F(y)$, and $F(y) = [F(y) - F(x)] + F(x)$, so

$$\begin{aligned} T_F(y) \pm F(y) \geqslant & \left(\sum\nolimits_1^n |F(x_j) + F(x_{j-1})|\right) + |F(y) - F(x)| \pm [F(y) - F(x)] \pm F(x) \\ \geqslant & T_F(x) - \varepsilon \pm F(x). \end{aligned}$$

Since ε is arbitrary, $T_F(y) \pm F(y) \geqslant T_F(x) \pm F(x)$, as desired. □

Theorem 3.64: 3.27.

(a) $F \in \mathrm{BV}$ if and only if $\operatorname{Re} F \in \mathrm{BV}$ and $\operatorname{Im} F \in \mathrm{BV}$.
(b) If $f\colon \mathbb{R} \to \mathbb{R}$, then $F \in \mathrm{BV}$ if and only if F is the difference of two bounded increasing functions; for $F \in \mathrm{BV}$ these functions may be taken to be $\frac{1}{2}(T_F + F)$ and $\frac{1}{2}(T_F - F)$.
(c) If $F \in \mathrm{BV}$, then $F(x+) = \lim_{y \searrow x} F(y)$ and $F(x-) = \lim_{y \nearrow x} F(y)$ exist for all $x \in \mathbb{R}$, as do $F(\pm\infty) = \lim_{y \to \pm\infty} F(y)$.
(d) If $F \in \mathrm{BV}$, the set of points at which F is discontinuous is countable. In particular, any $F \in \mathrm{BV}$ is Lebesgue integrable.
(e) If $F \in \mathrm{BV}$ and $G(x) = F(x+)$, then F' and G' exist and are equal a.e.

Proof. (a) is obvious. For (b), the "if" implication is easy (see Example 60(a,b)). To prove "only if," observe that by Lemma 63, the equation $F = \frac{1}{2}(T_F + F) - \frac{1}{2}(T_F - F)$ expresses F as the difference of two increasing functions. Also, the inequalities

$$T_F(y) \pm F(y) \geqslant T_F(x) \pm F(x) \quad (y > x)$$

imply that

$$|F(y) - F(x)| \leqslant T_F(y) - T_F(x) \leqslant T_F(\infty) - T_F(-\infty) < \infty,$$

so that F, and hence $T_F \pm F$, is bounded. Finally, (c), (d), and (e) follow from (a), (b), and Theorem 58. □

The representation $F = \frac{1}{2}(T_F + F) - \frac{1}{2}(T_F - F)$ of a real-valued $F \in \mathrm{BV}$ is called the ***Jordan decomposition of F***, and $\frac{1}{2}(T_F + F)$ and $\frac{1}{2}(T_F - F)$ are called the ***positive variation of F*** and the ***negative variation of F, respectively***. Since $x^+ = \max(x, 0) = \frac{1}{2}(|x| + x)$ and $x^- = \max(-x, 0) = \frac{1}{2}(|x| - x)$ for $x \in \mathbb{R}$, we have

$$\frac{1}{2}(T_F \pm F)(x) = \sup\left\{ \sum^n [F(x_j) - F(x_{j-1})]^\pm \;\middle|\; x_0 < \cdots < x_n = x \right\} \pm \frac{1}{2} F(-\infty),$$

so Theorem 64(a,b) leads to the connection between BC and the space of complex Borel measures on $\mathbb{R}$. To that end, we need the following definition:

Definition 65. *Define the collection of* ***normalized functions of bounded variation on*** $\mathbb{R}$ *by*

$$\mathrm{NBV} = \left\{ F\colon \mathbb{R} \to \mathbb{C} \;\middle|\; F \in \mathrm{BV},\ F \text{ is right continuous, and } \lim_{x \to -\infty} F(x) = 0 \right\}.$$

Remark 66. *If $F \in \mathrm{BV}$, then the function G defined by $G(x) = F(x+) - F(-\infty)$ is in NBV and $G' = F'$ a.e. (That $G \in \mathrm{BV}$ follows easily from Theorem 64(a,b): if F is real and $F = F_1 - F_2$ where F_1, F_2 are increasing, then $G(x) = F_1(x+) - [F_2(x+) + F(-\infty)]$, which is again the difference of two increasing functions.)*

Lemma 3.67: 3.28.

If $F \in \mathrm{BV}$, then $T_F(-\infty) = 0$. If F is also right continuous, then so is T_F.

Proof. If $\varepsilon > 0$ and $x \in \mathbb{R}$, choose $x_0 < \cdots < x_n = x$ so that

$$\sum\nolimits_1^n |F(x_j) - F(x_{j-1})| \geqslant T_F(x) - \varepsilon.$$

From ?? we see that $T_F(x) - T_F(x_0) \geqslant T_F(x) - \varepsilon$, and hence $T_F(y) \leqslant \varepsilon$ for $y \leqslant x_0$. Thus $T_F(-\infty) = 0$.

Now suppose that F is right continuous. Given $x \in \mathbb{R}$ and $\varepsilon > 0$, let $\alpha = T_F(x+) - T_F(x)$, and choose $\delta > 0$ so that $|F(x+h) - F(x)| < \varepsilon$ and $T_F(x+h) - T_F(x+) < \varepsilon$ whenever $0 < h < \delta$. For any such h, by ?? there exist $x_0 < \cdots < x_n = x + h$ such that

$$\sum\nolimits_1^n |F(x_j) - F(x_{j-1})| \geqslant \frac{3}{4}[T_F(x+h) - T_F(x)] \geqslant \frac{3}{4}\alpha,$$

and hence

$$\sum\nolimits_2^n |F(x_j) - F(x_{j-1})| \geqslant \frac{3}{4}\alpha - |F(x_1) - F(x_0)| \geqslant \frac{3}{4}\alpha - \varepsilon.$$

Likewise, there exist $x = t_0 < \cdots < t_m = x_1$ such that $\sum_1^n |F(t_j) - F(t_{j-1})| \geqslant \frac{3}{4}\alpha$, and hence

$$\begin{aligned} \alpha + \varepsilon &> T_F(x+h) - T_F(x) \\ &\geqslant \sum\nolimits_1^m |F(t_j) - F(t_{j-1})| + \sum\nolimits_2^n |F(x_j) - F(x_{j-1})| \\ &\geqslant \frac{3}{2}\alpha - \varepsilon \end{aligned}$$

Thus $\alpha < 4\varepsilon$, and since ε is arbitrary, $\alpha = 0$. □

Theorem 3.68: 3.29.

There is a bijective correspondence between real- (resp. complex-)valued functions in NBV and signed- (resp. complex-)Borel measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ given by

$$\begin{aligned} \mathrm{NBV} &\longleftrightarrow \left\{ \substack{\text{complex} \\ \text{measures on} \\ (\mathbb{R}, \mathcal{B}_{\mathbb{R}})} \right\}, \\ F_1(x) + iF_2(x) = F(x) &\longmapsto \mu_F \coloneqq (\mu_{F_1}^+ - \mu_{F_1}^-) + i(\mu_{F_2}^+ - \mu_{F_2}^-), \\ \mu((-\infty, x]) \eqqcolon F_\mu(x) &\longleftarrow \mu. \end{aligned}$$

Moreover, $|\mu_F| = \mu_{T_F}$.

Proof. If μ is a complex measure, we have $\mu = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-)$where the $\mu_j^\pm$are finite measures. If $F_j^\pm(x) = \mu_j^\pm((-\infty, x])$, then $F_j^\pm$is increasing and right continuous, $F_j^\pm(-\infty) = 0$, and $F_j^\pm(\infty) = \mu_j^\pm(\mathbb{R}) < \infty$. By Theorem 64(a,b) the function $F = F_1^+ - F_1^- + i(F_2^+ - F_2^-)$is in NBV. Conversely, by Theorem 64 and Lemma 67, any $F \in \mathrm{NBV}$ can be written in this form with the $F_j^\pm$ increasing and in NBV. Each $F_j^\pm$gives rise to a measure

$\mu_j^{\pm}$according to Theorem 43, so $F(x) = \mu_F((-\infty, x])$ where $\mu_F = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-)$. The proof that $|\mu_F| = \mu_{T_F}$ is outlined in Folland Exercise 3.28. □

The next obvious question is: Which functions in NBV correspond to measures μ such that $\mu \perp m$ or $\mu \ll m$? One answer is the following:

Proposition 3.69: 3.30.

If $F \in \text{NBV}$, then $F' \in L^1(m)$, and

$$\mu_F \perp m \quad \iff \quad F' = 0 \text{ a.e.},$$

and

$$\mu_F \ll m \quad \iff \quad F(x) = \int_{-\infty}^{x} F'(t)\,dt.$$

Proof. We have merely to observe that $F'(x) = \lim_{r\to 0} \mu_F(E_r)/m(E_r)$ where $E_r = (x, x+r)$ or $(x - r, x]$ and apply Theorem 53. (The measure μ_F is automatically regular by Theorem 45.) □

The condition $\mu_F \ll m$ can also be expressed directly in terms of F, as follows.

Definition 70. *A function $f\colon \mathbb{R} \to \mathbb{R}$ is called* ***absolutely continuous****, denoted $f \in \text{AC}$, if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for any finite set of disjoint intervals $(a_1, b_1), \ldots, (a_N, b_N)$,*

$$\sum_{j=1}^{N} (b_j - a_j) < \delta \Longrightarrow \sum_{j=1}^{N} |F(b_j) - F(a_j)| < \varepsilon. \tag{3.70.1}$$

More generally, F is said to be ***absolutely continuous*** *on $[a, b]$, denoted $f \in \text{AC}([a, b])$, if this condition is satisfied whenever the intervals (a_j, b_j) all lie in $[a, b]$.*

Remark 71. *Clearly, if F is absolutely continuous, then F is uniformly continuous (take $N = 1$ in Equation (3.70.1)). On the other hand, if F is everywhere differentiable and F' is bounded, then F is absolutely continuous, for $|F(b_j) - F(a_j)| \leqslant (\max|F'|)(b_j - a_j)$ by the mean value theorem.*

Example 72. *Consider*

$$f_k(x) = \begin{cases} 0, & x = 0, \\ x^k \sin(1/x), & x \neq 0, \end{cases} \qquad k = 0, 1, 2,$$

on $[a, b]$ with $a \leqslant 0 < b$ or $a < 0 \leqslant b$. Then, $f_0, f_1 \notin \text{BV}([a, b])$, hence, $f_0, f_1 \notin \text{AC}([a, b])$, but $f_2 \in \text{BV}([a, b])$. (Look at the graphs to analyze the difference in behaviors here.)

Proposition 3.73: 3.32.

If $F \in \text{NBV}$, then

$$F \in \text{AC} \quad \iff \quad \mu_F \ll m.$$

Proof. If $\mu_F \ll m$, the absolute continuity of F follows by applying Theorem 18 to the sets $E = \bigcup_1^N (a_j, b_j)$. To prove the converse, suppose that E is a Borel set such that $m(E) = 0$. If ε and δ are as in the definition of absolute continuity of F, by Theorem 45 we can find open sets $U_1 \supset U_2 \supset \cdots \supset E$ such that $m(U_1) < \delta$ (and thus $\mu(U_j) < \delta$ for all j) and $\mu_F(U_j) \to \mu_F(E)$. Each U_j is a disjoint union of open intervals (a_j^k, b_j^k), and

$$\sum_{k=1}^N |\mu_F((a_j^k, b_j^k))| \leqslant \sum_{k=1}^N |F(b_j^k) - F(a_j^k)| < \varepsilon$$

for all N. Letting $N \to \infty$, we obtain $|\mu_F(U_j)| < \varepsilon$ and hence $|\mu_F(E)| \leqslant \varepsilon$. Since ε is arbitrary, $\mu_F(E) = 0$, which-shows that $\mu_F \ll m$. □

Corollary 3.74: 3.33.

There is a bijective correspondence between $L^1(m)$ and $\mathrm{AC} \cap \mathrm{NBV}$, given by

$$\begin{aligned} L^1(m) &\longleftrightarrow \mathrm{AC} \cap \mathrm{NBV}, \\ f(x) &\longmapsto F(x) \coloneqq \int_{-\infty}^x f(t)\, dt, \\ F'(x) &\longleftarrow F(x). \end{aligned}$$

Proof. This follows immediately from Propositions 69 and 73 □

If we consider functions on bounded intervals, this result can be refined a bit.

Lemma 3.75: 3.34.

$\mathrm{AC}([a, b]) \subset \mathrm{BV}([a, b])$.

Proof. Let $F : \mathbb{R} \to \mathbb{C}$ be absolutely continuous on $[a, b]$. Let δ be as in the definition of absolute continuity, corresponding to $\varepsilon \neq 1$, and let N be the greatest integer less than $\delta^{-1}(b - a) + 1$. If $a = x_0 < \cdots < x_n = b$, by inserting more subdivision points if necessary, we can collect the intervals (x_{j-1}, x_j) into at most N groups of consecutive intervals such that the sum of the lengths in each group is less than δ. The sum $\sum |F(x_j) - F(x_{j-1})|$ over each group is at most 1, and hence the total variation of F on $[a, b]$ is at most N. □

Theorem 3.76: 3.35: The Fundamental Theorem of Calculus for Lebesgue Integrals.

If $-\infty < a < b < \infty$ and $F : [a, b] \to \mathbb{C}$, the following are equivalent:

(a) $F \in \mathrm{AC}([a, b])$.
(b) $F(x) - F(a) = \int_a^x f(t)\, dt$ for some $f \in L^1([a, b], m)$.
(c) F is differentiable a.e. on $[a, b]$, $F' \in L^1([a, b], m)$, and $F(x) - F(a) = \int_a^x F'(t)\, dt$.

Proof. To prove that (a) implies (c), we may assume by subtracting a constant from F that $F(a) = 0$. If we set $F(x) = 0$ for $x < a$ and $F(x) = F(b)$ for $x > b$, then $F \in \mathrm{NBV}$ by Lemma 75, so (c) follows from Corollary 74. That (c) implies (b) is trivial. Finally, (b) implies (a) by setting $f(t) = 0$ for $t \notin [a, b]$ and applying Corollary 74. □

The short form of the above theorem is the following:

Corollary 3.77.

If $[a, b]$ is a compact interval and $F\colon [a, b] \to \mathbb{C}$, then

$$F \in \mathrm{AC}([a,b]) \quad \iff \quad F' \text{ exists a.e. on } [a,b] \text{ and } \int_a^x F'(t)\,dt = F(x) - F(a).$$

The following decomposition of Borel measures on $\mathbb{R}^n$ is sometimes important. A complex Borel measure μ on $\mathbb{R}^n$ is called ***discrete*** if there is a countable set $\{x_j\} \subset \mathbb{R}^n$ and complex numbers c_j such that $\sum |c_j| < \infty$ and $\mu = \sum c_j \delta_{x_j}$, where δ_x is the point mass at x. On the other hand, μ is called continuous if $\mu(\{x\}) = 0$ for all $x \in \mathbb{R}^n$. Any complex measure μ can be written uniquely as $\mu = \mu_d + \mu_c$ where μ_d is discrete and μ_c is continuous. Indeed, let $E = \{x \mid \mu(\{x\}) \neq 0\}$. For any countable subset F of E the series $\sum_{x\in F} \mu(\{x\})$ converges absolutely to $\mu(F)$), so $\{x \in E \mid |\mu(\{x\})| > k^{-1}\}$ is finite for all k, and it follows that E itself is countable. Hence $\mu_d(A) = \mu(A \cap E)$ is discrete and $\mu_c(A) = \mu(A \setminus E)$ is continuous.

Obviously, if μ is discrete, then $\mu \perp m$; and if $\mu \ll m$, then μ is continuous. Thus, by Theorem 53, any (regular) complex Borel measure on $\mathbb{R}^n$ can be written uniquely as

$$\mu = \mu_d + \mu_{ac} + \mu_{sc}$$

where μ_d is discrete, μ_{ac} is absolutely continuous with respect to m, and μ_{sc} is a "singular continuous" measure, that is, μ_{sc} is continuous but $\mu_{sc} \perp m$.

The existence of nonzero singular continuous measures in $\mathbb{R}^n$ is evident enough when $n > 1$; the surface measure on the unit sphere discussed in Folland Section 2.7 is one example. Their existence when $n = 1$ is not quite so obvious; they correspond via Theorem 68 to nonconstant functions $F \in \mathrm{NBV}$ such that F is continuous but $F' = 0$ a.e. One such function is the Cantor function constructed in Folland Section 1.5 (extended to $\mathbb{R}$ by setting $F(x) = 0$ for $x < 0$ and $F(x) = 1$ for $x > 1$). More surprisingly, there exist strictly increasing continuous functions F such that $F' = 0$ a.e.; see Folland Exercise 3.40.

Notation 78. *If $F \in \mathrm{NBV}$, it is customary to denote the integral of a function g with respect to the measure μ_F by $\int g\,dF$ or $\int g(x)dF(x)$; that is,*

$$\int g\,dF \coloneqq \int g\,d\mu_F.$$

Integrals of these form are called ***Lebesgue-Stieltjes integrals****.*

We conclude by presenting an integration-by-parts formula for Lebesgue-Stieltjes integrals; other variants of this result can be found in Folland Exercise 3.34,Folland Exercise 3.35.

Theorem 3.79: 3.36.

If F and G are in NBV and at least one of them is continuous, then for all $-\infty < a < b < \infty$,

$$\int_{(a,b]} F\,dG + \int_{(a,b]} G\,dF = F(b)G(b) - F(a)G(a).$$

Proof. F and G are linear combinations of increasing functions in NBV by Theorem 64(a,b), so a simple calculation shows that it suffices to assume F and G increasing. Suppose for the sake of definiteness that G is continuous, and let $\Omega = \{(x,y) \mid a < x \leqslant y \leqslant b\}$. We use Fubini's theorem to compute $\mu_F \times \mu_G(\Omega)$ in two ways:

$$\mu_F \times \mu_G(\Omega) = \int_{(a,b]} \int_{(a,y]} dF(x)dG(x) = \int_{(a,b]} [F(y) - F(a)]dG(y)$$
$$= \int_{(a,b]} FdG - F(a)[G(b) - G(a)]$$

and since $G(x) = G(x-)$,

$$\mu_F \times \mu_G(\Omega) = \int_{(a,b]} \int_{[x,b]} dG(y)dF(x) = \int_{(a,b]} [G(b) - G(x)]dF(x)$$
$$= G(b)[F(b) - F(a)] - \int_{(a,b]} GdF$$

Subtracting these two equations, we obtain the desired result. □

Exercise 3.80: Folland Exercise 3.28.

If $F \in \text{NBV}$, let $G(x) = |\mu_F|((-\infty, x])$. Prove that $|\mu_F| = \mu_{T_F}$ by showing that $G = T_F$ via the following steps.

(a) From the definition of $T_F, T_F \leqslant G$.

(b) $|\mu_F(E)| \leqslant \mu_{T_F(E)}$ when E is an interval, and hence when E is a Borel set. c. $|\mu_F| \leqslant \mu_{T_F}$, and hence $G \leqslant T_F$. (Use Folland Exercise 3.21.)

Exercise 3.81: Folland Exercise 3.29.

If $F \in \text{NBV}$ is real-valued, then $\mu_F^+ = \mu_P$ and $\mu_F^- = \mu_N$ where P and N are the positive and negative variations of F. (Use Lemma 67.)

Exercise 3.82: Folland Exercise 3.30.

Construct an increasing function on $\mathbb{R}$ whose set of discontinuities is $\mathbb{R}$.

Exercise 3.83: Folland Exercise 3.31.

Let $F(x) = x^2 \sin(x^{-1})$ and $G(x) = x^2 \sin(x^{-2})$ for $x \neq 0$, and $F(0) = G(0) = 0$.

(a) F and G are differentiable everywhere (including $x = 0$).

(b) $F \in \mathrm{BV}([-1, 1])$, but $G \notin \mathrm{BV}([-1, 1])$.

Exercise 3.84: Folland Exercise 3.32.

If $F_1, F_2, \ldots, F \in \mathrm{NBV}$ and $F_j \to F$ pointwise, then $T_F \leqslant \liminf T_{F_j}$.

Exercise 3.85: Folland Exercise 3.33.

If $F\colon \mathbb{R} \to \mathbb{C}$ is increasing, then

$$F(b) - F(a) \geqslant \int_a^b F'(t)\, dt.$$

Solution. By a previous Folland Exercise F increasing on $\mathbb{R} \implies F$ measurable and bounded on $[a, b]$. By Theorem 58, F' exists m-a.e. By Theorem 58, the increasing right continuous function $G(x) \coloneqq F(x+)$ is also differentiable a.e. and bounded on $[a, b]$, and $F' = G'$ a.e., so it suffices to show $F(b) - F(a) \geqslant \int_a^b G'(t)\, dt$.

As G is an increasing right continuous function $\mathbb{R} \to \mathbb{R}$, by Theorem 43 there exists a unique Borel measure μ_G on $\mathbb{R}$ such that $\mu_G((x, y]) = G(x) - G(y)$ for all $x, y \in \mathbb{R}$. In particular, this shows $\mu_G \ll m$.

As $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mu_G)$ and $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, m)$ are σ-finite, by the Radon-Nikodym theorem there exists a unique m-measurable function $g \coloneqq d\mu_G/dm$. By a corollary to the LDT, $g(x) = \lim_{y\to x} \frac{\mu_G((x,y])}{m((x,y])} - \lim_{y\to x} \frac{G(x)-G(y)}{x-y} = G'(x)$. Hence

$$F(b) - F(a) \geqslant G(b) - G(a) \geqslant \mu_G((a, b]) = \int_{(a,b]} g\, dm = \int_a^b G'(t)\, dt = \int_a^b F'(t)\, dt. \qquad \square$$

Exercise 3.86: Folland Exercise 3.34.

Suppose $F, G \in \mathrm{NBV}$ and $-\infty < a < b < \infty$.

(a) By adapting the proof of Theorem 79, show that

$$\int_{[a,b]} \frac{F(x) + F(x-)}{2} dG(x) + \int_{[a,b]} \frac{G(x) + G(x-)}{2} dF(x)$$
$$= F(b)G(b) - F(a-)G(a-).$$

(b) If there are no points in $[a,b]$ where F and G are both discontinuous, then
$$\int_{[a,b]} FdG + \int_{[a,b]} GdF = F(b)G(b) - F(a-)G(a-).$$

Exercise 3.87: Folland Exercise 3.35.

If $F, G \in \mathrm{AC}([a,b])$, then so is FG, and
$$\int_a^b (FG' + GF')(x)dx = F(b)G(b) - F(a)G(a).$$

Exercise 3.88: Folland Exercise 3.36.

Let G be a continuous increasing function on $[a,b]$.

(a) If $E \subset [c,d]$ is a Borel set, then $m(E) = \mu_G(G^{-1}(E))$. (First consider the case where E is an interval.)

(b) If f is a Borel measurable and integrable function on $[c,d]$, then
$$\int_{G(a)}^{G(b)} f(y)\,dy = \int_a^b f(G(x))\,dG(x).$$
In particular, if G is absolutely continuous, then
$$\int_{G(a)}^{G(b)} f(y)\,dy = \int_a^b f(G(x))G'(x)\,dx.$$

(c) The validity of (b) may fail if G is merely right continuous rather than continuous.

Exercise 3.89: Folland Exercise 3.37.

Suppose $f\colon \mathbb{R} \to \mathbb{R}$. There is a constant M such that $|F(x) - F(y)| \leqslant M|x-y|$ for all $x, y \in \mathbb{R}$ (that is, F is Lipschitz continuous) if and only if F is absolutely continuous and $|F'| \leqslant M$ a.e.

Solution.

($\Rightarrow$) Suppose $|F(x) - F(y)| \leqslant M|x-y|$ for all $x, y \in \mathbb{R}$ and $\varepsilon > 0$. Then
$$\sum\nolimits_1^n |F(x_j) - F(x_{j-1})| \leqslant M\Big(\sum\nolimits_1^n |x;j - x_{j-1}|\Big),$$
so we can choose $\delta := \varepsilon/M$. Then $\sum_1^n |x_j - x_{j-1}| < \delta \implies \sum_1^n |F(x_j) - F(x_{j-1})| < \varepsilon$, so F is absolutely continuous. Also,
$$F'(x) = \lim_{x\to y} \frac{|f(x) - f(y)|}{|x-y|} \leqslant \lim_{x\to y} \frac{M|x-y|}{|x-y|} = M.$$

($\Leftarrow$) Suppose F is absolutely continuous and $|F'| \leqslant M$ a.e. Then by the mean value theorem, for all $(x, y) \subset [a, b]$, there exists $c \in (x, y)$ such that

$$M \geqslant |F'(c)| = \frac{|F(x) - F(y)|}{|x - y|},$$

so $|F(x) - F(y)| \leqslant M|x - y|$. Hence F is Lipschitz. □

Exercise 3.90: Folland Exercise 3.38.

If $f: [a, b] \to \mathbb{R}$, consider the graph of f as a subset of $\mathbb{R}$, namely, $\{t + if(t) \mid t \in [a, b]\}$. The length L of this graph is by definition the supremum of the lengths of all inscribed polygons. (An “inscribed polygon” is the union of the line segments joining $t_{j-1} + if(t_{j-1})$ to $t_j + if(t_j), 1 \leqslant j \leqslant n$, where $a = t_0 < \cdots < t_n = b$.)

(a) Let $F(t) = t + if(t)$; then L is the total variation of F on $[a, b]$.
(b) If f is absolutely continuous, $L = \int_a^b [1 + f'(t)^2]^{1/2} dt$.

Exercise 3.91: Folland Exercise 3.39.

If $\{F_j\}$ is a sequence of nonnegative increasing functions on $[a, b]$ such that $F(x) = \sum_1^\infty F_j(x) < \infty$ for all $x \in [a, b]$, then $F'(x) = \sum_1^\infty F_j'(x)$ for a.e. $x \in [a, b]$. (It suffices to assume $F_j \in \text{NBV}$. Consider the measures μ_{F_j}.)

Solution. Without loss of generality F_j is right continuous, since otherwise consider $G_j(x) = F_j(x+)$. Also we may assume $F_j(x) \to 0$ as $x \to -\infty$ since we only care about F_j on $[a, b]$, so we may assume $F_j \in \text{NBV}$. Then μ_{F_j} satisfies

$$\mu_{F_j}((-\infty, x]) = F_j(x) - \underbrace{F_j(-\infty)}_{=0} = F_j \in \text{NBV},$$

which then implies $F(x) \in \text{NBV}$, so

$$\begin{aligned} F(x) &= \sum\nolimits_1^\infty F_j(x) = \sum\nolimits_1^\infty \mu_{F_j}((-\infty, x]) \\ &= \sum\nolimits_1^\infty \int_{-\infty}^x F_j' \\ &= \int_{-\infty}^x \sum\nolimits_1^\infty F_j' && \text{(by MCT for series since } F_j \in L^+(m)) \end{aligned}$$

On the other hand, since F is also right continuous and increasing,

$$\mu_F((a, b]) = F(b) - F(a) \text{ has } b - a = 0 \implies F(x) - \cancelto{0}{F(-\infty)} = \int_{-\infty}^x F'(t)\, dt$$

so by uniqueness of the Radon-Nikodym derivative we conclude $\sum_1^\infty F_j' = F'$. □

Exercise 3.92: Folland Exercise 3.40.

Let F denote the Cantor function on $[0,1]$ (see Σ1.5), and set $F(x) = 0$ for $x < 0$ and $F(x) = 1$ for $x > 1$. Let $\{[a_n, b_n]\}$ be an enumeration of the closed subintervals of $[0,1]$ with rational endpoints, and let $F_n(x) = F((x - a_n)/(b_n - a_n))$. Then $G = \sum_1^\infty 2^{-n} F_n$ is continuous and strictly increasing on $[0,1]$, and $G' = 0$ a.e. (Use Folland Exercise 3.39.)

Exercise 3.93: Folland Exercise 3.41.

Let $A \subset [0,1]$ be a Borel set such that $0 < m(A \cap I) < m(I)$ for every subinterval I of $[0,1]$ (Folland Exercise 1.33).

(a) Let $F(x) = m([0,x] \cap A)$. Then F is absolutely continuous and strictly increasing on $[0,1]$, but $F' = 0$ on a set of positive measure.

(b) Let $G(x) = m([0,x] \cap A) - m([0,x] \setminus A)$. Then G is absolutely continuous on $[0,1]$, but G is not monotone on any subinterval of $[0,1]$.

Exercise 3.94: Folland Exercise 3.42.

A function $F\colon (a,b) \to \mathbb{R}$ $(-\infty \leqslant a < b \leqslant \infty)$ is called ***convex*** if

$$F(\lambda s + (1-\lambda)t) \leqslant \lambda F(s) + (1-\lambda)F(t)$$

for all $s, t \in (a,b)$ and $\lambda \in (0,1)$. (Geometrically, this says that the graph of F over the interval from s to t lies underneath the line segment joining $(s, F(s))$ to $(t, F(t))$.)

(a) F is convex if and only if for all $s, t, s', t' \in (a,b)$ such that $s \leqslant s' < t'$ and $s < t \leqslant t'$.

$$\frac{F(t) - F(s)}{t - s} \leqslant \frac{F(t') - F(s')}{t' - s'}.$$

(b) F is convex if and only if F is absolutely continuous on every compact subinterval of (a,b) and F' is increasing (on the set where it is defined).

(c) If F is convex and $t_0 \in (a,b)$, there exists $\beta \in \mathbb{R}$ such that $F(t) - F(t_0) \geqslant \beta(t - t_0)$ for all $t \in (a,b)$.

(d) (Jensen's Inequality) If $(X, \mathcal{M}, \mu)$ is a measure space with $\mu(X) = 1$, $g\colon X \to (a,b)$ is in $L^1(\mu)$, and F is convex on (a,b), then

$$F\left(\int g\, d\mu\right) \leqslant \int F \circ g\, d\mu.$$

(Let $t_0 = \int g d\mu$ and $t = g(x)$ in (c), and integrate.)

4 Point-Set Topology

4.1 Topological Spaces

The concepts of limit, convergence, and continuity are central to all of analysis, and it is useful to have a general framework for studying them that includes the classical manifestations as special cases. One such framework, which has the advantage of not requiring many ideas beyond those occurring in analysis on Euclidean space, is that of metric spaces. However, metric spaces are not sufficiently general to describe even some very classical modes of convergence, for example, pointwise convergence of functions on $\mathbb{R}$. A more flexible theory can be built by taking the open sets, rather than a metric, as the primitive data, and it is this theory that we shall explore in the present chapter.

Let X be a nonempty set.

Definition 1. *A* ***topology*** *on X is a family $\mathfrak{T}$ of subsets of X that contains $\varnothing$ and X and is closed under arbitrary unions and finite intersections (i.e., if $\{U_\alpha\}_{\alpha\in A} \subset \mathfrak{T}$ then $\bigcup_{\alpha\in A} U_\alpha \in \mathfrak{T}$, and if $U_1, \ldots, U_n \in \mathfrak{T}$ then $\bigcap_1^n U_j \in \mathfrak{T}$). The pair $(X, \mathfrak{T})$ is called a* ***topological space***. *If $\mathfrak{T}$ is understood, we shall simply refer to the topological space X.*

Example 2. *Let us examine a few examples:*

(1) If X is any nonempty set, $\mathcal{P}(X)$ and $\{\varnothing, X\}$ are topologies on X. They are called the discrete topology and the trivial (or indiscrete) topology, respectively.

(2) If X is an infinite set, $\{U \subset X \mid U = \varnothing \text{ or } U^c \text{ is finite}\}$ is a topology on X, called the cofinite topology.

(3) If X is a metric space, the collection of all open sets with respect to the metric is a topology on X.

(4) If $(X, \mathfrak{T})$ is a topological space and $Y \subset X$, then $\mathfrak{T}_Y = \{U \cap Y \mid U \in \mathfrak{T}\}$ is a topology on Y, called the relative topology induced by $\mathfrak{T}$.

We now present the basic terminology concerning topological spaces. Most of these concepts are already familiar in the context of metric spaces. Until further notice, $(X, \mathfrak{T})$ will be a fixed topological space.

The members of $\mathfrak{T}$ are called ***open sets***, and their complements are called ***closed sets***. If $Y \subset X$, the open (resp. closed) subsets of Y in the relative topology are called ***relatively open*** (resp. ***relatively closed***). We observe that, by DeMorgan's laws, the family of closed sets is closed under arbitrary intersections and finite unions.

If $A \subset X$, the union of all open sets contained in A is called the ***interior*** of A, and the intersection of all closed sets containing A is called the ***closure*** of A. We denote the interior and closure of A by A° and $\overline{A}$, respectively. Observe that A° is the largest open set contained in A and $\overline{A}$ is the smallest closed set containing A, and we have $(A^o)^c = \overline{A^c}$ and $(\overline{A})^c = (A^c)^o$. The difference $\overline{A} \backslash A^o = \overline{A} \cap \overline{A^c}$ is called the ***boundary*** of A and is denoted by ∂A. If $\overline{A} = X$, A is called ***dense*** in X. On the other hand, if $(\overline{A})^\circ = \varnothing$, A is called ***nowhere dense***. (This name comes from the fact that if $(\overline{A})^\circ$ were some nonempty subset E of X, then in the subspace E the set $\overline{A}$ is dense, hence A dense "somewhere". If E were empty, then E is dense "nowhere".)

If $x \in X$ (resp. $E \subset X$), a ***neighborhood*** of x (resp. of E) is a set $A \subset X$ such that $x \in A^\circ$ (or $E \subset A^\circ$). Thus, a set A is open if and only if it is a neighborhood of itself. (Some authors require neighborhoods to be open sets; Folland does not, and to be precise we will sometimes opt for the more common term "open neighborhood" to mean a neighborhood that is open, where neighborhood here is as in Folland's definition above.) A point x is called an ***accumulation point*** of A if $A \cap (U \backslash \{x\}) \neq \varnothing$ for every neighborhood U of x. (Other terms sometimes used for the same concept are "cluster point" and "limit point." We shall use "cluster point" to mean something a bit different below.)

Proposition 4.3: 4.1.

If $A \subset X$, let $\mathrm{Acc}(A)$ be the set of accumulation points of A. Then $\overline{A} = A \cup \mathrm{Acc}(A)$, and A is closed if and only if $\mathrm{Acc}(A) \subset A$.

Proof. If $x \notin \overline{A}$, then A^c is a neighborhood of x that does not intersect A, so $x \notin \mathrm{Acc}(A)$; thus $A \cup \mathrm{Acc}(A) \subset \overline{A}$. If $x \notin A \cup \mathrm{Acc}(A)$, there is an open U containing x such that $U \cap A = \varnothing$, so that $\overline{A} \subset U^c$ and $x \notin \overline{A}$. Thus $\overline{A} \subset A \cup \mathrm{Acc}(A)$. Finally, A is closed if and only if $A = \overline{A}$, and this happens if and only if $\mathrm{Acc}(A) \subset A$. □

If $\mathfrak{T}_1$ and $\mathfrak{T}_2$ are topologies on X such that $\mathfrak{T}_1 \subset \mathfrak{T}_2$, we say that $\mathfrak{T}_1$ is ***weaker*** (or ***coarser***) than $\mathfrak{T}_2$, or that $\mathfrak{T}_2$ is ***stronger*** (or ***finer***) than $\mathfrak{T}_1$.

Clearly the trivial topology is the weakest topology on X, while the discrete topology is the strongest. If $\mathcal{E} \subset \mathcal{E}(X)$, there is a unique weakest topology $\mathfrak{T}(\mathcal{E})$ on X that contains $\mathcal{E}$, namely the intersection of all topologies on X containing $\mathcal{E}$. It is called the ***topology generated by*** $\boldsymbol{\mathcal{E}}$, and $\mathcal{E}$ is sometimes called a ***subbase for*** $\boldsymbol{\mathfrak{T}(\mathcal{E})}$.

Definition 4. *If $\mathfrak{T}$ is a topology on X, a* **neighborhood base** *for $\mathfrak{T}$ at $x \in X$ is a family $\mathcal{N} \subset \mathfrak{T}$ such that*

- *$x \in V$ for all $V \in \mathcal{N}$;*
- *if $x \in U \in \mathfrak{T}$, there exists $V \in \mathcal{N}$ such that $x \in V \subset U$.*

A **base** *for $\mathfrak{T}$ is a family $\mathcal{B} \subset \mathfrak{T}$ that contains a neighborhood base for $\mathfrak{T}$ at each $x \in X$. (So, a neighborhood base really is a base for any neighborhood of U.)*

. For example, if X is a metric space, the collection of open balls centered at x is a neighborhood base for the metric topology at x, and the collection of all open balls in X is a base.

Proposition 4.5: 4.2.

If $(X, \mathcal{T})$ is a topological space and $\mathcal{E} \subset \mathcal{T}$, then $\mathcal{E}$ is a base for $\mathcal{T}$ if and only if every nonempty set $U \in \mathcal{T}$ is the union of elements of $\mathcal{E}$.

Proof. Suppose $\mathcal{E}$ is a base for $\mathcal{T}$. Then for $x \in U$, there exists $V_x \in \mathcal{E}$ such that $x \in V_x \subset U$. Then $\bigcup_{x \in U} V_x = U$. Conversely, suppose $U = \bigcup_{V \in \mathcal{E}} V$. Then $\{v \in \mathcal{E} \mid x \in V\}$ is a neighborhood base for x, so $\mathcal{E}$ is a base. □

Proposition 4.6: 4.3.

If $\mathcal{E} \subset \mathcal{E}(X)$, in order for $\mathcal{E}$ to be a base for a topology on X it is necessary and sufficient that the following two conditions be satisfied:

(a) each $x \in X$ is contained in some $V \in \mathcal{E}$;
(b) if $U, V \in \mathcal{E}$ and $x \in U \cap V$, there exists $W \in \mathcal{E}$ with $x \in W \subset (U \cap V)$.

Proof. The necessity is clear, since if U, V are open, then so is $U \cap V$. To prove the sufficiency, let

$$\mathcal{T} = \{U \subset X \mid \text{for every } x \in U, \text{ there exists } V \in \mathcal{E} \text{ with } x \in V \subset U\}.$$

Then $X \in \mathcal{T}$ by condition (a) and $\varnothing \in \mathcal{T}$ trivially, and $\mathcal{T}$ is obviously closed under unions. If $U_1, U_2 \in \mathcal{T}$ and $x \in U_1 \cap U_2$, there exist $V_1, V_2 \in \mathcal{E}$ with $x \in V_1 \subset U_1$ and $x \in V_2 \subset U_2$, and by condition (b) there exists $W \in \mathcal{E}$ with $x \in W \subset (V_1 \cap V_2)$. Thus $U_1 \cap U_2 \in \mathcal{T}$, so by induction $\mathcal{T}$ is closed under finite intersections. Therefore $\mathcal{T}$ is a topology, and $\mathcal{E}$ is clearly a base for $\mathcal{T}$. □

Proposition 4.7: 4.4.

If $\mathcal{E} \subset \mathcal{E}(X)$, the topology $\mathcal{T}(\mathcal{E})$ generated by $\mathcal{E}$ consists of $\varnothing, X$, and all unions of finite intersections of members of $\mathcal{E}$.

Proof. The family of finite intersections of sets in $\mathcal{E}$, together with X, satisfies the conditions of Proposition 6, so by Proposition 5 the family of all unions of such sets, together with $\varnothing$, is a topology. It is obviously contained in $\mathcal{T}(\mathcal{E})$, hence equal to $\mathcal{T}(\mathcal{E})$. □

Note how the simplicity of this proposition contrasts with the corresponding result for σ-algebras (Proposition 60). What makes life easier here is that only finite intersections are involved.

The concept of topological space is general enough to include a great profusion of interesting examples, but—by the same token—too general to yield many interesting theorems. To build a reasonable theory one must usually restrict the class of spaces under consideration. The remainder of this section is devoted to a discussion of two types of restrictions that are commonly made, the so-called countability and separation axioms.

Definition 8. *A topological space $(X, \mathfrak{T})$ satisfies the* ***first axiom of countability****, or is* ***first countable****, if there is a countable neighborhood base for $\mathfrak{T}$ at every point of X.*

Definition 9. *The space $(X, \mathfrak{T})$ satisfies the* ***second axiom of countability****, or is* ***second countable****, if $\mathfrak{T}$ has a countable base.*

Definition 10. *The space $(X, \mathfrak{T})$ is* ***separable*** *if X has a countable dense subset.*

It is useful to observe that if X is first countable, then for every $x \in X$ there is a neighborhood base $\{U_j\}_1^\infty$ at x such that $U_j \supset U_{j+1}$ for all j. Indeed, if $\{V_j\}_1^\infty$ is any countable neighborhood base at x, we can take $U_j = \bigcap_1^j V_i$.

Every metric space is first countable (the balls of rational radius about x are a neighborhood base at x), and a metric space is second countable if and only if it is separable (Folland Exercise 4.5). The latter fact can be partly generalized:

Proposition 4.11: 4.5.

Every second countable space is separable, but not conversely (see Folland Exercise 4.6).

Proof. Let X be a second countable space and $\{U_n\}_{n \in \mathbb{Z}_{\geqslant 1}}$ be a countable base for the topology. For each U_n, pick any element $x_n \in U_n$, discarding any empty U_n. Then we need only show that $A = \{x_n \mid x_n \in U_n\}$ is dense in X:

Take any nonempty open set E. Then $U_n \subset E$ for some n, hence, $x_n \in E$ for some n. But $x_n \in A$, so $E \cap A \neq \varnothing$ for all E, so A is dense. □

Note that the above proof relies on the Axiom of Countable Choice, and in fact, the previous proposition can be shown to be equivalent to the Axiom of Countable Choice.

Definition 12. *A sequence $\{x_j\}$ in a topological space X* ***converges*** *to $x \in X$ (in symbols: $x_j \to x$) if for every neighborhood U of x there exists $J \in \mathbb{Z}_{\geqslant 0}$ such that $x_j \in U$ for all $j > J$.*

First countable spaces have the pleasant property that such things as closure and continuity can be characterized in terms of sequential convergence—which is not the case in more general spaces, as we shall see. For example, see the following proposition.

Proposition 4.13: 4.6.

If X is first countable and $A \subset X$, then $x \in \overline{A}$ if and only if there is a sequence $\{x_j\}$ in A that converges to x.

Equivalent characterization of denseness is that any open set intersects it nontrivially.

Proof. Let $\{U_j\}$ be a countable neighborhood base at x with $U_j \supset U_{j+1}$ for all j. If $x \in \overline{A}$, then $U_j \cap A \neq \varnothing$ for all j. Pick $x_j \in U_j \cap A$; since $U_k \subset U_j$ for $k > j$ and every neighborhood of x contains some U_j, it is clear that $x_j \to x$. On the other hand, if $x \notin \overline{A}$ and $\{x_j\}$ is any sequence in A, then $(\overline{A})^c$ is a neighborhood of x containing no x_j, so $x_j \nrightarrow x$. □

Lastly, we discuss the ***separation axioms***. These are properties of a topological space, labeled $T_0, \ldots, T_4$, that guarantee the existence of open sets that separate points or closed sets from each other. If X has the property T_j, we say that X is a T_j space or that the topology on X is T_j.

Axiom	Definition
T_0 (Kolmogorov)	If $x \neq y$, there exists an open set containing x but not y *or* an open set containing y but not x.
T_1 (Fréchet)	If $x \neq y$, there is an open set containing y but not x.
T_2 (Hausdorff)	If $x \neq y$, there are disjoint open sets U, V with $x \in U$ and $y \in V$.
T_3 (Regular)	X is T_1 and for any closed $A \subset X$ and any $x \in A^c$ there are disjoint open sets U, V with $x \in U$ and $A \subset V$.
T_4 (Normal)	X is T_1 and for any disjoint closed A, B in X there are disjoint open sets U, V with $A \subset U$ and $B \subset V$.

There is also an additional useful separation condition, intermediate between T_3 and T_4, that we will discuss in Folland Section 4.2.

Warning 4.14.

Note that some authors do not require regular and normal spaces to be T_1.

The following characterization of T_1 spaces is useful. It shows in particular that every normal space is regular and that every regular space is Hausdorff.

Proposition 4.15: 4.7.

X is a T_1 space if and only if $\{x\}$ is closed for every $x \in X$.

Proof. If X is T_1 and $x \in X$, for each $y \neq x$ there is an open U_y containing y but not x; thus $\{x\}^c = \bigcup_{y \neq x} U_y$ is open and $\{x\}$ is closed. Conversely, if $\{x\}$ is closed, then $\{x\}^c$ is an open set containing every $y \neq x$. □

The vast majority of topological space that arise in practice are Hausdorff, or become Hausdorff after simple modifications. (This last phrase refers to spaces such as the space of integrable functions on a measure space, which becomes a Hausdorff space with the L^1 metric when we identify two functions that are equal a.e.) However, two classes of usually non-Hausdorff topologies are of sufficient importance to warrant special mention: the quotient topology on a space of equivalence classes, discussed in Folland Exercise 4.28,Folland Exercise 4.29, and the Zariski topology on an algebraic variety. Without attempting to give the definition of an algebraic variety, we now describe the Zariski topology on a vector space.

Example 16 (Zariski Topology on a Vector Space). *Let k be a field, and let $k[X_1, \ldots, X_n]$ be the ring of polynomials in n variables over x. Each $P \in k[X_1, \ldots, X_n]$ determines a polynomial map $p\colon k^n \to k$ by substituting elements of k for the formal indeterminates $X_1, \ldots, X_n$. The correspondence $P \to p$ is one-to-one precisely when k is infinite. The collection of all sets $p^{-1}(\{0\})$ in k^n, as p ranges over all polynomial maps, is closed under finite unions, since $p^{-1}(\{0\}) \cup q^{-1}(\{0\}) = (pq)^{-1}(\{0\})$, and it contains k^n itself (take $p = 0$). Hence, by Propositions 5 and 6, the collection of all sets of the form $\bigcap_{\alpha \in A} p_\alpha^{-1}(\{0\})$ (p_α being a polynomial map for each α) is the collection of closed sets for a topology on k^n, called the* ***Zariski topology****. The Zariski topology is T_1 by Proposition 15, for if $a = (a_1, \ldots, a_n) \in k^n$ then $\{a\} = \bigcap_1^n p_j^{-1}(\{0\})$ where $p_j(X_1, \ldots, X_n) = X_j - a_j$. If k is finite the Zariski topology is discrete, but if k is infinite the Zariski topology is not Hausdorff; in fact, any two nonempty open sets have nonempty intersection. This is just a restatement of the fact that $k[X_1, \ldots, X_n]$ is an integral domain, that is, if P and Q are nonzero polynomials, then PQ is nonzero. (For $n = 1$, the Zariski topology is the cofinite topology.)*

Exercise 4.17: Folland Exercise 4.1.

If $\operatorname{card}(X) \geqslant 2$, then there exists a topology on X that is T_0 but not T_1.

Exercise 4.18: Folland Exercise 4.2.

If X is an infinite set, the cofinite topology on X is T_1 but not T_2, and is first countable if and only if X is countable.

Exercise 4.19: Folland Exercise 4.3.

Every metric space is normal. (If A, B are closed sets in the metric space (X, ρ), consider the sets of points x where $\rho(x, A) < \rho(x, B)$ or $\rho(x, A) > \rho(x, B)$.)

Solution. Let $X = (X, \rho)$ be a metric space and $x \neq y$ in X.

- X is T_1: If $x \neq y$ then there exists an open subset contaiinng y but not x, namely $U_x = \{z \in X \mid \rho(x, z) > 0\}$. This is open in X because for all $z \in U_x$, $\rho(x, z) =: r > 0$, the open ball $B_r(x) \subset U_x$: indeed, if $z' \in B_r(x)$, then $\rho(z', x) + \rho(z, z') \geqslant \rho(x, z) = r$, so $\rho(z', x) \geqslant \underbrace{\rho(x, z)}_{=r} - \underbrace{\rho(z, z')}_{<r} > 0$ by the triangle inequality, so $z' \in U_x$.
- Now suppose A, B are disjoint closed neighborhoods of x and y, respectively. Then let
$$U_x = \{z \in X \mid \rho(z, A) < \rho(z, B)\} \text{ and } V_y = \{z \in X \mid \rho(z, A) > \rho(z, B)\}.$$
Then U_x, V_y are disjoint because if $z \in U_x$ then $\rho(z, A) < \rho(z, B)$, hence $\rho(z, B)$ is not greater then $\rho(z, A)$; and U_x, V_y are operation because if $z \in U_x$ then $\rho(z, A) -$

$\rho(z, B) =: r > 0$, so again by the triangle inequality we conclude $B_r(z) \subset U_x$. Showing V_y is open is similar. □

Exercise 4.20: Folland Exercise 4.4.

Let $X = \mathbb{R}$, and let $\mathcal{T}$ be the family of all subsets of $\mathbb{R}$ of the form $U \cup (V \cap \mathbb{R})$ where U, V are open in the usual sense. Then $\mathcal{T}$ is a topology that is Hausdorff but not regular. (In view of Folland Exercise 4.3, this shows that a topology stronger than a normal topology need not be normal or even regular.)

Exercise 4.21: Folland Exercise 4.5.

Every separable metric space is second countable.

Solution. Let (X, ρ) be separable metric spaces, say with countable dense subset Q. Then take the countable base to be $\mathscr{B} = \{B_r(q)\}_{r \in \mathbb{Q}_{>0}, q \in Q}$. This is a base: If $U \neq \varnothing$ is open in X, then for all $q \in U$ there exists $r_q \in \mathbb{Q}_{>0}$ such that $B_{r_q}(q) \in U$; we claim $U = \bigcup_{q \in U \cap Q} B_{r_q}(q)$. Indeed, if $x \in U$ then for all $r_x \in \mathbb{Q}_{>0}$ (where r_0 is the guaranteed positive rational numbers such that $B_{r_0} \subset U$), there exists $q \in Q$ such that $q \in B_{r/2}(x)$or equivalently, $x \in B_{r/2}(q)$, which is true in the union. □

The previous proposition proves the forward direction. On the other hand, let A be countable and dense in the separable metric space X. Consider the collection $\mathcal{B}$ of balls $B(x, 1/n), x \in A, n \in \mathbb{Z}_{\geqslant 1}$. Take any open set E and consider $y \in E$. Then $B(y, 1/m) \subset E$ for some m. As A is dense, $A \cap B(y, (2m)^{-1}) \neq \varnothing$, so we can choose an $x \in A$ with $d(x, y) < (2m)^{-1}$. Thus,

$$y \in B(x, (2m)^{-1}) \subset B(y, m^{-1}) \subset E,$$

so E is the union of elements of $\mathcal{B}$ that it contains. Thus $\mathcal{B}$ is a base for X (as E was arbitrary) and $\mathcal{B}$ is countable since A and $\mathbb{Z}_{\geqslant 1}$ are.

Exercise 4.22: Folland Exercise 4.6.

Let $\mathcal{E} = \{(a, b] \mid -\infty < a < b < \infty\}$.

(a) $\mathcal{E}$ is a base for a topology $\mathcal{T}$ on $\mathbb{R}$ in which the members of $\mathcal{E}$ are both open and closed.

(b) $\mathcal{T}$ is first countable but not second countable. (If $x \in \mathbb{R}$, every neighborhood base at x contains a set whose supremum is x.)

(c) $\mathbb{Q}$ is dense in $\mathbb{Q}$ with respect to $\mathcal{T}$. (Thus the converse of Proposition 11 is false.)

Exercise 4.23: Folland Exercise 4.7.

If X is a topological space, a point $x \in X$ is called a ***cluster point*** of the sequence $\{x_j\}$ if for every neighborhood U of $x, x_j \in U$ for infinitely many j. If X is first countable, x is a cluster point of $\{x_j\}$ if and only if some subsequence of $\{x_j\}$ converges to x.

Solution. Let $\{x_j\}_{j=1}^{\infty}$ be a sequence in a first countable space X, and fix $x \in X$. Since X is first countable, there exists a countable neighborhood base $\{V_j\}_{j=1}^{\infty}$ at x. Let $U_j \coloneqq \bigcap_{\ell=1}^{j} V_\ell$. Then $\{U_j\}$ is an countable neighborhood base for x, and moreover $U_1 \supset U_2 \supset \cdots$.

Given a sequence $\{x_n\}_{n=1}^{\infty}$, if a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ converges to x, then for any neighborhood U of x, there exists $N \in \mathbb{Z}_{\geqslant 1}$ such that $x_{n_k} \in U$ whenever $k \geqslant N$. Hence x is a cluster point of $\{x_n\}_{n=1}^{\infty}$.

Conversely, if x is a cluster point of $\{x_n\}_{n=1}^{\infty}$ in a first countable space X, there exists a countable nested neighborhood base $\{U_n\}_{n=1}^{\infty}$ at x. Inductively choose $n_k > n_{k-1}$ such that $x_{n_k} \in U_k$. For any neighborhood U of x, $U_N \subseteq U$ for some N, ensuring $x_{n_k} \in U$ for all $k \geqslant N$. Thus, the subsequence $(x_{n_k})_{k=1}^{\infty}$ converges to x. □

Exercise 4.24: Folland Exercise 4.8.

If X is an infinite set with the cofinite topology and $\{x_j\}$ is a sequence of distinct points in X, then $x_j \to x$ for every $x \in X$.

Solution. Let U be an open neighborhood of $x \in X$. Then U^c must finite, hence U is infinite. Since $\{x_j\}_{j=1}^{\infty}$ is an infinite collection of distinct points, there must be some $J \in \mathbb{Z}_{\geqslant 1}$ such that $x_j \in U$ whenever $j \geqslant J$. Since U was arbitrary, we conclude $x_j \to x$. □

Exercise 4.25: Folland Exercise 4.9.

If X is a linearly ordered set, the topology $\mathfrak{T}$ generated by the sets $\{x \mid x < a\}$ and $\{x \mid x > a\}$ ranging over each $a \in X$ is called the ***order topology***.

(a) If $a, b \in X$ and $a < b$, there exist $U, V \in \mathfrak{T}$ with $a \in U, b \in V$, and $x < y$ for all $x \in U$ and $y \in V$. The order topology is the weakest topology with this property.
(b) If $Y \subset X$, the order topology on Y is never stronger than, but may be weaker than, the relative topology on Y induced by the order topology on X.
(c) The order topology on $\mathbb{R}$ is the usual topology.

Exercise 4.26: Folland Exercise 4.10.

A topological space X is called ***disconnected*** if there exist nonempty open sets U, V such that $U \cap V = \varnothing$ and $U \cup V = X$; otherwise X is ***connected***. When we speak of connected or disconnected subsets of X, we refer to the relative topology on them.

(a) X is connected if and only if $\varnothing$ and X are the only subsets of X that are both open and closed.

(b) If $\{E_\alpha\}_{\alpha\in A}$ is a collection of connected subsets of X such that $\bigcap_{\alpha\in A} E_\alpha \neq \varnothing$, then $\bigcup_{\alpha\in A} E_\alpha$ is connected.

(c) If $A \subset X$ is connected, then $\overline{A}$ is connected.

(d) Every point $x \in X$ is contained in a unique maximal connected subset of X, and this subset is closed. (It is called the connected component of x.)

Solution. (a) ($\implies$) Suppose there exists a clopen $E \subset X$. The obvious choice is $E = U$ and $E^c = V$; since E is clopen we have E, E^c are nonempty disjoint open sets that union to X, hence X is disconnected. ($\impliedby$) Suppose X is disconnected. We want some clopen $E \subset X$. Since X is disconnected, there exist nonempty disjoint open $U, V \subset X$ such that $U \cup V = X$. The obvious choice here is $E = U$; indeed, U is open and $U^c = V$ is open, hence $V^c = U$ is closed, so U is clopen.

(b) Define $E \coloneqq \bigcup_{\alpha\in A} E_\alpha$. For non-empty open sets $U, V \subseteq E$ covering E, choose $x \in \bigcap_{\alpha\in A} E_\alpha$ with $x \in U$, and $y \in V$ such that $y \in E_\alpha$ for some $\alpha \in A$. Then, $U \cap E_\alpha$ and $V \cap E_\alpha$ are non-empty open sets in E_α covering it. Since E_α is connected, $U \cap V \neq \varnothing$, proving E is connected.

(c) For disjoint open sets $U, V \subseteq \overline{A}$ covering $\overline{A}$, write $U = U' \cap \overline{A}$ and $V = V' \cap \overline{A}$ for open sets $U', V' \subseteq X$. Then $U' \cap A$ and $V' \cap A$ are disjoint open sets in A covering A. If A is connected, $U' \cap A = \varnothing$. This implies $U' \cap \operatorname{Acc}(A) = \varnothing$, leading to $U' \cap \overline{A} = \varnothing$. Thus, $\overline{A}$ is connected.

(d) Let $x \in X$ and define $\mathcal{C} \coloneqq \{A \subseteq X \mid A \text{ is connected and } x \in A\}$. Then, $C \coloneqq \bigcup \mathcal{C}$ is connected by part (b). If $A \subseteq X$ is connected with $C \subseteq A$, then $x \in A$ implies $A \in \mathcal{C}$, leading to $A \subseteq C$. Thus, C is maximal. For any maximal connected set $C' \subseteq X$ containing x, $C' \in \mathcal{C}$ implies $C' \subseteq C$, and maximality of C' gives $C' = C$. Hence, C is unique. Since $\overline{C}$ is connected (by part (c)) and contains C, it follows that $C = \overline{C}$, showing C is closed. □

Exercise 4.27: Folland Exercise 4.11.

If $E_1, \ldots, E_n$ are subsets of a topological space, the closure of $\bigcup_1^n E_j$ is $\bigcup_1^n \overline{E}_j$.

Exercise 4.28: Folland Exercise 4.12.

Let X be a set. A ***Kuratowski closure operator*** on X is a map $A \mapsto A^*$ from $\mathcal{P}(X)$ to itself satisfying (i) $\varnothing^* = \varnothing$, (ii) $A \subset A^*$ for all A, (iii) $(A^*)^* = A^*$ for all A, and (iv) $(A \cup B)^* = A^* \cup B^*$ for all A, B.

(a) If X is a topological space, the map $A \mapsto \overline{A}$ is a Kuratowski closure operator. (Use Folland Exercise 4.11.)

(b) Conversely, given a Kuratowski closure operator, let $\mathcal{F} = \{A \subset X \mid A = A^*\}$ and $\mathcal{T} = \{U \subset X \mid U^c \in \mathcal{F}\}$. Then $\mathcal{T}$ is a topology, and for any set $A \subset X$, A^* is its closure with respect to $\mathcal{T}$.

Exercise 4.29: Folland Exercise 4.13.

If X is a topological space, U is open in X, and A is dense in X, then $\overline{U} = \overline{U \cap A}$.

Solution. We use the fact that in any topological space X, for any subset E of X, a point $x \in X$ has $x \in \overline{E}$ if and only if every open neighborhood intersects E.

$\overline{U \cap A} \subset \overline{U}$ since $U \cap A \subset U$, so suffices to show $\overline{U} \subset \overline{U \cap A}$. If $x \in \overline{U}$ then every open neighborhood of x intersects U. But $x \in \overline{A} = X$ too, so every open neighborhood of x also intersects A. Hence every open neighborhood of x intersects $U \cap A$, so $x \in \overline{U \cap A}$. □

4.2 Continuous Maps

Topological spaces are the natural setting for the concept of continuity, which can be described in either global or local terms as follows.

Definition 30. *Let X and Y be topological spaces and f a map from X to Y.*

Then f is called **continuous** *if $f^{-1}(V)$ is open in X for every open $V \subset Y$. (Since $f^{-1}(A^c) = [f^{-1}(A)]^c$, an equivalent condition is that $f^{-1}(A)$ is closed in X for every closed $A \subset Y$.)*

If $x \in X$, f is called **continuous at x** *if for every neighborhood V of $f(x)$ there is a neighborhood U of x such that $f(U) \subset V$, or equivalently, if $f^{-1}(V)$ is a neighborhood of x for every neighborhood V of $f(x)$.*

Exercise 4.31.

Show the equivalence of the above two definitions of continuity at a point.

Clearly, if $f: X \to Y$ and $g: Y \to Z$ are continuous (or f is continuous at x and g is continuous at $f(x)$), then $g \circ f$ is continuous (at x).

Notation 32. *We shall denote the set of continuous maps from X to Y by $C(X, Y)$.*

Proposition 4.33: 4.8.

The map $f: X \to Y$ is continuous if and only if f is continuous at every $x \in X$.

Proof. If f is continuous and V is a neighborhood of $f(x)$, $f^{-1}(V^\circ)$ is an open set containing x, so f is continuous at x. Conversely, suppose that f is continuous at each $x \in X$. If $V \subset Y$ is open, V is a neighborhood of each of its points, so $f^{-1}(V)$ is a neighborhood of each of its points. Thus $f^{-1}(V)$ is open, so f is continuous. □

Proposition 4.34: 4.9.

If the topology on Y is generated by a family of sets $\mathcal{E}$, then $f\colon X \to Y$ is continuous if and only if $f^{-1}(V)$ is open in X for every $V \in \mathcal{E}$.

Proof. This is clear from Proposition 7 and the fact that the set mapping f^{-1} commutes with unions and intersections. □

Definition 35. *If $f\colon X \to Y$ is bijective and f and f^{-1} are both continuous, f is called a **homeomorphism**, and X and Y are said to be **homeomorphic**.*

In the case $f\colon X \to Y$ is homeomorphism, the set mapping f^{-1} is a bijection from the open sets in Y to the open sets in X, so X and Y may be considered identical as far as their topological properties go.

The following provides an example of a homeomorphism between familiar spaces, which can also show how some properties are easily recognized as not being topological of nature.

Example 36. *The tangent function is a homeomorphism between $(-\pi/2, \pi/2)$ and $\mathbb{R}$ (with the usual topologies) and thus preserves topological structures. Properties like boundedness then are not topological in nature.*

Definition 37. *If $f\colon X \to Y$ is injective but not surjective, and $f\colon X \to f(X)$ is a homeomorphism when $f(X) \subset Y$ is given the relative topology, f is called an **embedding**.*

Definition 38. *If X is any set and $\{f_\alpha\colon X \to Y_\alpha\}_{\alpha\in A}$ is a family of maps from X into some topological spaces Y_α, there is a unique weakest topology $\mathfrak{T}$ on X that makes all the f_α continuous; it is called the **weak topology generated by $\{f_\alpha\}_{\alpha\in A}$**, or the **initial topology generated by $\{f_\alpha\}_{\alpha\in A}$**. Namely, $\mathfrak{T}$ is the topology generated by sets of the form $f_\alpha^{-1}(U_\alpha)$ where $\alpha \in A$ and U_α is open in Y_α.*

Thus the initial topology generated by $\{f_\alpha\}_{\alpha\in A}$ is $\mathfrak{T}(\{f_\alpha^{-1}(U_\alpha) \mid \alpha \in A\})$.

Example 39 (Cartesian Product)**.** *The most important example of this construction is the Cartesian product of topological spaces. If $\{X_\alpha\}_{\alpha\in A}$ is any family of topological spaces, the product topology on $X = \prod_{\alpha\in A} X_\alpha$ is the weak topology generated by the coordinate maps $\pi_\alpha\colon X \to X_\alpha$. When we consider a Cartesian product of topological spaces, we always endow it with the product topology unless we specify otherwise. By Proposition 7, a base for the product topology is given by the sets of the form $\bigcap_1^n \pi_{\alpha_j}^{-1}(U_{\alpha_j})$ where $n \in \mathbb{Z}_{\geqslant 0}$ and U_{α_j} is open in X_{α_j} for $1 \leqslant j \leqslant n$. These sets can also be written as $\prod_{\alpha\in A} U_\alpha$ where $U_\alpha = X_\alpha$ if $\alpha \neq \alpha_1, \ldots, \alpha_n$. Notice, in particular, that if A is infinite, a product of nonempty open sets $\prod_{\alpha\in A} U_\alpha$ is open in $\prod_{\alpha\in A} X_\alpha$ if and only if $U_\alpha = X_\alpha$ for all but finitely many α.*

Proposition 4.40: 4.10.

If X_α is Hausdorff for each $\alpha \in A$, then $X = \prod_{\alpha\in A} X_\alpha$ is Hausdorff.

Proof. If x and y are distinct points of X, we must have $\pi_\alpha(x) \neq \pi_\alpha(y)$ for some α. Let U and V be disjoint neighborhoods of $\pi_\alpha(x)$ and $\pi_\alpha(y)$ in X_α. Then $\pi_\alpha^{-1}(U)$ and $\pi_\alpha^{-1}(V)$ are disjoint neighborhoods of x and y in X. □

Proposition 4.41: 4.11.

If $\{X_\alpha\}_{\alpha\in A}$ and Y are topological spaces and $X = \prod_{\alpha\in A} X_\alpha$, then $f\colon Y \to X$ is continuous if and only if $\pi_\alpha \circ f$ is continuous for each α.

Proof. If $\pi_\alpha \circ f$ is continuous for each α, then $f^{-1}(\pi_\alpha^{-1}(U_\alpha))$ is open in Y for each open U_α in X_α. By Proposition 34, f is continuous. The converse is obvious. □

If the spaces X_α are all equal to some fixed space X, the product $\prod_{\alpha\in A} X_\alpha$ is just the set X^A of mappings from A to X, and the product topology is just the topology of pointwise convergence. More precisely:

Proposition 4.42: 4.12.

If X is a topological space, A is a nonempty set, and $\{f_n\}$ is a sequence in X^A, then $f_n \to f$ in the product topology if and only if $f_n \to f$ pointwise.

Proof. The sets

$$N(U_1, \ldots, U_k) = \bigcap_1^k \pi_{\alpha_j}^{-1}(U_j) = \{g \in X^A \mid g(\alpha_j) \in U_j \text{ for } 1 \leqslant j \leqslant k\},$$

where $k \in \mathbb{Z}_{\geqslant 0}$ and U_j is a neighborhood of $f(\alpha_j)$ in X for each j, form a neighborhood base for the product topology at f. If $f_n \to f$ pointwise, then $f_n(\alpha_j) \in U_j$ for $n \geqslant N_j$ and hence $f_n \in N(U_1, \ldots, U_k)$ for $n \geqslant \max(N_1, \ldots, N_k)$; therefore $f_n \to f$ in the product topology. Conversely, if $f_n \to f$ in the product topology, $\alpha \in A$, and U is a neighborhood of $f(\alpha)$, then $f_n \in N(U) = \pi_\alpha^{-1}(U)$ for large n; hence $f_n(\alpha) \in U$ for large n, and so $f_n(\alpha) \to f(\alpha)$. □

We shall be particularly interested in real- and complex-valued functions on topological spaces. If X is any set, we denote by $B(X, \mathbb{R})$ (resp. $B(X, \mathbb{R})$) the space of all bounded real- (resp. complex-)valued functions on X. If X is a topological space, we also have the spaces $C(X, \mathbb{R})$ and $C(X, \mathbb{R})$ of continuous functions on X, and we define

$$BC(X, F) = B(X, F) \cap C(X, F) \quad (F = \mathbb{R} \text{ or } \mathbb{R}).$$

Notation 43. *In speaking of complex-valued functions we shall usually omit the $\mathbb{C}$ and simply write $B(X), C(X)$, and $BC(X)$.*

Since addition and multiplication are continuous from $\mathbb{C} \times \mathbb{C}$ to $\mathbb{C}$, $C(X)$ and $BC(X)$ are complex vector spaces.

Definition 44. *If $f \in B(X)$, we define the* ***uniform norm*** *of f to be*

$$\|f\|_u = \sup\{|f(x)| \mid x \in X\}.$$

The function $\rho(f,g) = \|f-g\|_u$ is easily seen to be a metric on $B(X)$, and convergence with respect to this metric is simply uniform convergence on $X.B(X)$ is obviously complete in the uniform metric: If $\{f_n\}$ is uniformly Cauchy, then $\{f_n(x)\}$ is Cauchy for each x, and if we set $f(x) = \lim_n f_n(x)$, it is easily verified that $\|f_n - f\|_u \to 0$.

Proposition 4.45: 4.13.

If X is a topological space, $BC(X)$ is a closed subspace of $B(X)$ in the uniform metric; in particular, $BC(X)$ is complete.

Proof. Suppose $\{f_n\} \subset BC(X)$ and $\|f_n - f\|_u \to 0$. Given $\varepsilon > 0$, choose N so large that $\|f_n - f\|_u < \varepsilon/3$ for $n > N$. Given $n > N$ and $x \in X$, since f_n is continuous at x there is a neighborhood U of x such that $|f_n(y) - f_n(x)| < \varepsilon/3$ for $y \in U$. But then

$$|f(y) - f(x)| \leqslant |f(y) - f_n(y)| + |f_n(y) - f_n(x)| + |f_n(x) - f(x)| < \varepsilon,$$

so f is continuous at x. By Proposition 33, f is continuous. □

For a given topological space X it may happen that $C(X)$ consists only of constant functions. This is obviously the case, for example, if X has the trivial topology, but it can happen even when X is regular. Normal spaces, however, always have plenty of continuous functions, as the following fundamental theorems show.

Lemma 4.46: 4.14.

Suppose that A and B are disjoint closed subsets of the normal space X, and let $\Delta = \{k2^{-n} \mid n \in \mathbb{Z}_{\geqslant 1}, 0 < k < 2^n\}$ be the set of dyadic rational numbers in $(0,1)$. There is a family $\{U_r \mid r \in \Delta\}$ of open sets in X such that $A \subset U_r \subset B^c$ for all $r \in \Delta$ and $\overline{U}_r \subset U_s$ for $r < s$.

Proof. By normality, there exist disjoint open sets V, W such that $A \subset V$, $B \subset W$. Let $U_{1/2} = V$. Then since W^c is closed,

$$A \subset U_{1/2} \subset \overline{U}_{1/2} \subset W^c \subset B^c.$$

We now select U_r for $r = k2^{-n}$ by induction on n. Suppose that we have chosen U_r for $r = k2^{-n}$ when $0 < k < 2^n$ and $n \leqslant N-1$. To find U_r for $r = (2j+1)2^{-N}$ $(0 \leqslant j < 2^{N-1})$, observe that $\overline{U}_{j2^{1-N}}$ and $(U_{(j+1)2^{1-N}})^c$ are disjoint closed sets (where we set $\overline{U}_0 = A$ and $U_1^c = B$), so as above we can choose an open U_r with

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

These U_rs have the desired properties. □

Theorem 4.47: 4.15: Urysohn's Lemma.

Let X be a normal space. If A and B are disjoint closed sets in X, there exists $f \in C(X, [0,1])$ such that $f = 0$ on A and $f = 1$ on B.

Proof. Let U_r be as in Lemma 46 for $r \in \Delta$, and set $U_1 = X$. For $x \in X$, define $f(x) = \inf\{r \mid x \in U_r\}$. Since $A \subset U_r \subset B^c$ for $0 < r < 1$, we clearly have $f(x) = 0$ for $x \in A$ and $f(x) = 1$ for $x \in B$, and $0 \leqslant f(x) \leqslant 1$ for all $x \in X$. It remains to show that f is continuous. To this end, observe that $f(x) < \alpha$ if and only if $x \in U_r$ for some $r < \alpha$ if and only if $x \in \bigcup_{r<\alpha} U_r$, so $f^{-1}((-\infty, \alpha)) = \bigcup_{r<\alpha} U_r$ is open. Also $f(x) > \alpha$ if and only if $x \notin U_r$ for some $r > \alpha$ if and only if $x \notin \overline{U}_s$ for some $s > \alpha$ (since $\overline{U}_s \subset U_r$ for $s < r$) if and only if $x \in \bigcup_{s>\alpha} (\overline{U}_s)^c$, so $f^{-1}((a, \infty)) = \bigcup_{s>\alpha} (\overline{U}_s)^c$ is open. Since the open half-lines generate the topology on $\mathbb{R}$, f is continuous by Proposition 34. □

The proof of Urysohn's lemma may seem somewhat opaque at first, but there is a simple geometric intuition behind it. If one pictures X as the plane $\mathbb{R}^2$ and the sets U_r as regions bounded by curves, the curves ∂U_r form a "topographic map" of the function f:

Theorem 4.48: 4.16: The Tietze Extension Theorem.

Let X be a normal space. If A is a closed subset of X and $f \in C(A, [a,b])$, there exists $F \in C(X, [a,b])$ such that $F|_A = f$.

Proof. Replacing f by $(f-a)/(b-a)$, we may assume that $[a,b] = [0,1]$. We claim that there is a sequence $\{g_n\}$ of continuous functions on X such that $0 \leqslant g_n \leqslant 2^{n-1}/3^n$ on X and $0 \leqslant f - \sum_1^n g_j \leqslant (2/3)^n$ on A. To begin with, let $B = f^{-1}([0,1/3])$ and $C = f^{-1}([2/3,1])$. These are closed subsets of A, and since A itself is closed, they are closed in X. By Urysohn's Lemma there is a continuous $g_1 \colon X \to [0,1/3]$ with $g_1 = 0$ on B and $g_1 = 1/3$ on C; it follows that $0 \leqslant f - g_1 \leqslant 2/3$ on A. Having found $g_1, \ldots, g_{n-1}$, by the same reasoning we can find $g_n \colon X \to [0, 2^{n-1}/3^n]$ such that $g_n = 0$ on the set where $f - \sum_1^{n-1} g_j \leqslant 2^{n-1}/3^n$ and $g_n = 2^{n-1}/3^n$ on the set where $f - \sum_1^{n-1} g_j \geqslant (2/3)^n$. Let $F = \sum_1^\infty g_n$. Since $\|g_n\|_u \leqslant 2^{n-1}/3^n$, the partial sums of this series converge uniformly, so F is continuous by Proposition 45. Moreover, on A we have $0 \leqslant f - F \leqslant (2/3)^n$ for all n, whence $F = f$ on A. □

Corollary 4.49: 4.17.

If X is normal, $A \subset X$ is closed, and $f \in C(A)$, there exists $F \in C(X)$ such that $F|_A = f$.

Proof. By considering real and imaginary parts separately, it suffices to assume that f is real-valued. Let $g = f/(1+|f|)$. Then $g \in C(A, (-1,1))$, so there exists $G \in C(X, [-1,1])$

with $G|_A = g$. Let $B = G^{-1}(\{-1, 1\})$. By Urysohn's lemma there exists $h \in C(X, [0, 1])$ with $h = 1$ on $A, h = 0$ on B. Then $hG = G$ on A and $|hG| < 1$ everywhere, so $F = hG/(1 - |hG|)$ does the job. □

Definition 50. *A topological space X is called* ***completely regular*** *if X is T_1 and for each closed $A \subset X$ and each $x \notin A$ there exists $f \in C(X, [0, 1])$ such that $f(x) = 1$ and $f = 0$ on A. Completely regular spaces are also called* ***Tychonoff*** *or $T_{3^1\!/\!_2}$.*

The latter terminology is justified, for every completely regular space is T_3 (if A, x, f are as above, then $f^{-1}\big((\frac{1}{2}, \infty)\big)$ and $f^{-1}\big((-\infty, \frac{1}{2})\big)$ are disjoint neighborhoods of x and A), and Urysohn's lemma shows that every T_4 space is completely regular.

Exercise 4.51: Folland Exercise 4.14.

If X and Y are topological spaces, $f : X \to Y$ is continuous if and only if $f(A) \subset f(A)$ for all $A \subset X$ if and only if $\overline{f^{-1}(B)} \subset f^{-1}(\overline{B})$ for all $B \subset Y$.

Solution. Proof. Suppose f is continuous and consider any $A \subset X$. Then $f^{-1}(\overline{f(A)})$ is a closed set in X. Since $A \subset f^{-1}(f(A)) \subset f^{-1}(\overline{f(A)})$, one concludes that $\overline{A} \subset f^{-1}(\overline{f(A)})$ as the latter is a closed set and the former is the smallest closed set containing A. Finally, this allows one to see that for every $A \subset X$,

$$f(\overline{A}) \subset f\Big[f^{-1}(\overline{f(A)})\Big] \subset \overline{f(A)}.$$

(Note that the inclusions used regarding images and preimages are in general strict unless more assumptions are made on the map f.)

Next, suppose $f(\overline{A}) \subset \overline{f(A)}$ for all $A \subset X$. Given $B \subset Y, f^{-1}(B) \subset X$ and one can write

$$f\Big(\overline{f^{-1}(B)}\Big) \subset \overline{f(f^{-1}(B))} \subset \overline{B}.$$

Taking the inverse image on both sides and using the fact that $\overline{f^{-1}(B)} \subset f^{-1}\Big[f\Big(\overline{f^{-1}(B)}\Big)\Big]$ yields

$$\overline{f^{-1}(B)} \subset f^{-1}\Big[f\Big(\overline{f^{-1}(B)}\Big)\Big] \subset f^{-1}(\overline{B}).$$

For the final implication, suppose $\overline{f^{-1}(B)} \subset f^{-1}(\overline{B})$ for all $B \subset Y$. Given a closed set $D \subset Y$ one can then write

$$\overline{f^{-1}(D)} \subset f^{-1}(\overline{D}) = f^{-1}(D) \subset \overline{f^{-1}(D)},$$

so that $f^{-1}(D) = \overline{f^{-1}(D)}$, that is, $f^{-1}(D)$ is closed in X for every closed set $D \subset Y$. Therefore f is continuous. □

Exercise 4.52: Folland Exercise 4.15.

If X is a topological space, $A \subset X$ is closed, and $g \in C(A)$ satisfies $g = 0$ on ∂A, then the extension of g to X defined by $g(x) = 0$ for $x \in A^c$ is continuous.

Solution. Let $\widetilde{g} = 0$ on A^c and g on A. Let U be open in $\mathbb{R}$. Since $g \in C(A)$, we may assume $0 \nu U$ since otherwise $\widetilde{g}^{-1}(\{0\}) \subset A$, hence $\widetilde{g}^{-1}(\{0\}) = g^{-1}(\{0\})$, hence closed by continuity of g (and the fact $\{0\}$ is closed in $\mathbb{C}$ since $\mathbb{C}$ is T_1). So assume $0 \in U$. Then $\widetilde{g}^{-1}(\{0\}) = \underbrace{A^c \cup \partial A}_{=(A^c)^\circ=\text{closed}} \cup \underbrace{g^{-1}(\{0\})}_{\text{closed}}$, which is closed. Hence $\widetilde{g}$ is continuous. □

Exercise 4.53: Folland Exercise 4.16.

Let X be a topological space, Y a Hausdorff space, and f, g continuous maps from X to Y.

(a) $\{x \mid f(x) = g(x)\}$ is closed.
(b) If $f = g$ on a dense subset of X, then $f = g$ on all of X.

Exercise 4.54: Folland Exercise 4.17.

If X is a set, $\mathfrak{F}$ a collection of real-valued functions on X, and $\mathfrak{T}$ the weak topology generated by $\mathfrak{F}$, then $\mathfrak{T}$ is Hausdorff if and only if for every $x, y \in X$ with $x \neq y$ there exists $f \in \mathfrak{F}$ with $f(x) \neq f(y)$.

Exercise 4.55: Folland Exercise 4.18.

If X and Y are topological spaces and $y_0 \in Y$, then X is homeomorphic to $X \times \{y_0\}$ where the latter has the relative topology as a subset of $X \times Y$.

Exercise 4.56: Folland Exercise 4.19.

If $\{X_\alpha\}$ is a family of topological spaces, $X = \prod_\alpha X_\alpha$ (with the product topology) is uniquely determined up to homeomorphism by the following property: There exist continuous maps $\pi_\alpha : X \to X_\alpha$ such that if Y is any topological space and $f_\alpha \in C(Y, X_\alpha)$ for each α, there is a unique $F \in C(Y, X)$ such that $f_\alpha = \pi_\alpha \circ F$. (Thus X is the category-theoretic product of the X_αs in the category of topological spaces.)

Exercise 4.57: Folland Exercise 4.20.

If A is a countable set and X_α is a first (resp. second) countable space for each $\alpha \in A$, then $\prod_{\alpha \in A} X_\alpha$ is first (resp. second) countable.

Exercise 4.58: Folland Exercise 4.21.

If X is an infinite set with the cofinite topology, then every $f \in C(X)$ is constant.

Solution. Proof. First, we will show that if X is any infinite set equipped with the cofinite topology, then no two nonempty open sets are disjoint (this property is called ***hyperconnected***). To see this, consider two open sets U and U' which can be written as $U = F^c$ and $U' = (F')^c$ for some finite sets F, F' in this topology. As $F \cup F'$ is also a finite set, the set $(F \cup F')^c$ is an open set, and any point in this set is also in both U and U', that is, $U \cap U' \neq \varnothing$.

Now, suppose that f is continuous but not constant. Then it takes on at least two distinct values, say $p, q \in \mathbb{C}$. Hence we can find $\varepsilon > 0$ such that $B_\varepsilon(p) \cap B_\varepsilon(q) = \varnothing$. As these sets are disjoint nonempty open sets in $\mathbb{C}$, the preimage of the sets would be disjoint nonempty open sets in X which cannot happen by the previous paragraph. Thus f must be constant. □

Exercise 4.59: Folland Exercise 4.22.

Let X be a topological space, (Y, ρ) a complete metric space, and $\{f_n\}$ a sequence in Y^X such that $\sup_{x \in X} \rho(f_n(x), f_m(x)) \to 0$ as $m, n \to \infty$. There exists a unique $f \in Y^X$ such that $\sup_{x \in X} \rho(f_n(x), f(x)) \to 0$ as $n \to \infty$. If each f_n is continuous, so is f.

Exercise 4.60: Folland Exercise 4.23.

Give an elementary proof of the Tietze extension theorem for the case $X = \mathbb{R}$.

Exercise 4.61: Folland Exercise 4.24.

A Hausdorff space X is normal if and only if X satisfies the conclusion of Urysohn's lemma if and only if X satisfies the conclusion of the Tietze extension theorem.

Exercise 4.62: Folland Exercise 4.25.

If $(X, \mathfrak{T})$ is completely regular, then $\mathfrak{T}$ is the weak topology generated by $C(X)$.

Exercise 4.63: Folland Exercise 4.26.

Let X and Y be topological spaces.

(a) If X is connected (see Folland Exercise 4.10) and $f \in C(X, Y)$, then $f(X)$ is connected.

(b) X is called ***arcwise connected*** if for all $x_0, x_1 \in X$ there exists $f \in C([0,1], X)$ with $f(0) = x_0$ and $f(1) = x_1$. Every arcwise connected space is connected.

(c) Let $X = \{(0,0)\} \cup \{(s,t) \in \mathbb{R}^2 \mid t = \sin(1/s)\}$, with the relative topology induced from $\mathbb{R}^2$. Then X is connected but not arcwise connected.

Exercise 4.64: Folland Exercise 4.27.

If X_α is connected for each $\alpha \in A$ (see Folland Exercise 4.10), then $X = \prod_{\alpha \in A} X_\alpha$ is connected. (Fix $x \in X$ and let Y be the connected component of x in X. Show that Y includes $\{y \in X \mid \pi_\alpha(y) = \pi_\alpha(x)$ for all but finitely many $\alpha\}$ and that the latter set is dense in X. Use Folland Exercise 4.10,Folland Exercise 4.18.)

Exercise 4.65: Folland Exercise 4.28.

Let X be a topological space equipped with an equivalence relation, $\widetilde{X}$ the set of equivalence classes, $\pi \colon X \to \widetilde{X}$ the map taking each $x \in X$ to its equivalence class, and $\mathfrak{T} = \left\{ U \subset \widetilde{X} \,\middle|\, \pi^{-1}(U) \text{ is open in } X \right\}$.

(a) $\mathfrak{T}$ is a topology on $\widetilde{X}$. (It is called the quotient topology.)
(b) If Y is a topological space, $f \colon \widetilde{X} \to Y$ is continuous if and only if $f \circ \pi$ is continuous.
(c) $\widetilde{X}$ is T_1 if and only if every equivalence class is closed.

Exercise 4.66: Folland Exercise 4.29.

If X is a topological space and G is a group of homeomorphisms from X to itself, G induces an equivalence relation on X, namely, $x \sim y$ if and only if $y = g(x)$ for some $g \in G$. Let $X = \mathbb{R}^2$; describe the quotient space $\widetilde{X}$ and the quotient topology on it (as in Folland Exercise 4.28) for each of the following groups of invertible linear maps. In particular, show that in (a) the quotient space is homeomorphic to $[0,\infty)$; in (b) it is T_1 but not Hausdorff; in (c) it is T_0 but not T_1, and in (d) it is not T_0. (In fact, in (d) $\widetilde{X}$ is uncountable, but there are only six open sets and there are points $p \in \widetilde{X}$ such that $\overline{\{p\}} = \widetilde{X}$.)

(a) $\left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \,\middle|\, \theta \in \mathbb{R} \right\}$
(b) $\left\{ \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \,\middle|\, a \in \mathbb{R} \right\}$
(c) $\left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \,\middle|\, a > 0, b \in \mathbb{R} \right\}$
(d) $\left\{ \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \,\middle|\, a, b \in \mathbb{Q} \setminus \{0\} \right\}$.

4.3 Nets

As we have hinted above, sequential convergence does not play the same central role in general topological spaces as it does in metric spaces. The reasons for this may be illustrated by the following example. Consider the space $\mathbb{C}^{\mathbb{C}}$ of all complex-valued functions on $\mathbb{C}$, with the product topology (i.e., the topology of pointwise convergence), and its subspace $C(\mathbb{C})$. On the one hand, by Corollary 11, if $\{f_n\} \subset C(\mathbb{C})$ and $f_n \to f$ pointwise, then f is Borel measurable, so the set of limits of convergent sequences in $C(\mathbb{C})$ is a proper subset of $\mathbb{C}^{\mathbb{C}}$. Nonetheless, $C(\mathbb{C})$ is dense in $\mathbb{C}^{\mathbb{C}}$. Indeed, if $f \in \mathbb{C}^{\mathbb{C}}$, the sets

$$\{g \in \mathbb{C}^{\mathbb{C}} \mid |g(x_j) - f(x_j)| < \varepsilon \text{ for } j = 1, \ldots, n\} \text{ where } n \in \mathbb{Z}_{\geqslant 0}, x_1, \ldots, x_n \in \mathbb{C}, \varepsilon > 0,$$

form a neighborhood base at f, and each of these sets clearly contains continuous functions.

There is, however, a generalization of the notion of sequence that works well in arbitrary topological spaces; the key idea is to use index sets more general than $\mathbb{Z}_{\geqslant 0}$. The precise definitions are as follows.

Definition 67. *A* ***directed set*** *is a set A equipped with a binary relation $\lesssim$ such that*

- *$\alpha \lesssim \alpha$ for all $\alpha \in A$;*
- *if $\alpha \lesssim \beta$ and $\beta \lesssim \gamma$ then $\alpha \lesssim \gamma$;*
- *for any $\alpha, \beta \in A$ there exists $\gamma \in A$ such that $\alpha \lesssim \gamma$ and $\beta \lesssim \gamma$.*

If $\alpha \lesssim \beta$, we shall also write $\beta \gtrsim \alpha$.

Definition 68. *A* ***net*** *in a set X is a mapping $\alpha \mapsto x_\alpha$ from a directed set A into X. We shall usually denote such a mapping by $\langle x_\alpha \rangle_{\alpha \in A}$, or just by $\langle x_\alpha \rangle$ if A is understood, and we say that $\langle x_\alpha \rangle$ is* ***indexed by*** *A.*

Example 69. *Here are some examples of directed sets:*

(i) The set of positive integers $\mathbb{Z}_{\geqslant 0}$, with $j \lesssim k$ if and only if $j \leqslant k$.
(ii) The set $\mathbb{R} \setminus \{a\} (a \in \mathbb{R})$, with $x \lesssim y$ if and only if $|x - a| \geqslant |y - a|$.
(iii) The set of all partitions $\{x_j\}_0^n$ of the interval $[a, b]$ (i.e., $a = x_0 < \cdots < x_n = b$), with $\{x_j\}_0^n \lesssim \{y_k\}_0^m$ if and only if $\max(x_j - x_{j-1}) \geqslant \max(y_k - y_{k-1})$.
(iv) The set $\mathcal{N}$ of all neighborhoods of a point x in a topological space X, with $U \lesssim V$ if and only if $U \supset V$. (We say that $\mathcal{N}$ is ***directed by reverse inclusion****.)*
(v) The Cartesian product $A \times B$ of two directed sets, with $(\alpha, \beta) \lesssim (\alpha', \beta')$ if and only if $\alpha \lesssim \alpha'$ and $\beta \lesssim \beta'$. (This is always the way we make $A \times B$ into a directed set.)

Examples (i)-(iii) occur in elementary analysis: A net indexed by $\mathbb{Z}_{\geqslant 0}$ is just a sequence, and the nets indexed by the sets in (ii) and (iii) occur in defining limits of real variables and Riemann integrals. Example (iv) is of fundamental importance in topology, and we shall see several uses of the construction in (v).

Definition 70. *Let X be a topological space and E a subset of X.*

A net $\langle x_\alpha \rangle_{\alpha \in A}$ is ***eventually in*** *E if there exists $\alpha_0 \in A$ such that $x_\alpha \in E$ for all $\alpha \gtrsim \alpha_0$, and $\langle x_\alpha \rangle$ is* ***frequently in*** *E if for every $\alpha \in A$ there exists $\beta \gtrsim \alpha$ such that $x_\beta \in E$.*

A point $x \in X$ is a ***limit*** *of $\langle x_\alpha \rangle$ (or $\langle x_\alpha \rangle$* ***converges*** *to x, or $x_\alpha \to x$) if for every neighborhood U of $x, \langle x_\alpha \rangle$ is eventually in U, and x is a* ***cluster point*** *of $\langle x_\alpha \rangle$ if for every neighborhood U of x, $\langle x_\alpha \rangle$ is frequently in U.*

The next three propositions show that nets are a good substitute for sequences.

Proposition 4.71: 4.18.

If X is a topological space, $E \subset X$, and $x \in X$, then the following hold.

$$x \in \operatorname{Acc}(E) \iff \text{there exists a net in } E \setminus \{x\} \text{ that converges to } x.$$
$$x \in \overline{E} \iff \text{there exists a net in } E \text{ that converges to } x.$$

Proof. If x is an accumulation point of E, let $\mathcal{N}$ be the set of neighborhoods of x, directed by reverse inclusion. For each $U \in \mathcal{N}$, pick $x_U \in (U \setminus \{x\}) \cap E$. Then $x_U \to x$. Conversely, if $x_\alpha \in E \setminus \{x\}$ and $x_\alpha \to x$, then every punctured neighborhood of x contains some x_α, so x is an accumulation point of E. Likewise, if $x_\alpha \to x$ where $x_\alpha \in E$, then $x \in \overline{E}$, and the converse follows from Proposition 3. □

Proposition 4.72: 4.19.

If X and Y are topological spaces, $f\colon X \to Y$, and $x \in X$, then

$$f \text{ is continuous at } x \iff \text{for every net } \langle x_\alpha \rangle \text{ converging to } x, \langle f(x_\alpha) \rangle \text{ converges to } f(x).$$

Proof. If f is continuous at x and V is a neighborhood of $f(x)$, then $f^{-1}(V)$ is a neighborhood of x. Hence, if $x_\alpha \to x$ then $\langle x_\alpha \rangle$ is eventually in $f^{-1}(V)$, so $\langle f(x_\alpha) \rangle$ is eventually in V, and thus $f(x_\alpha) \to f(x)$. On the other hand, if f is not continuous at x, there is a neighborhood V of $f(x)$ such that $f^{-1}(V)$ is not a neighborhood of x, that is, $x \notin (f^{-1}(V))^\circ$, or equivalently, $x \in \overline{f^{-1}(V^c)}$. By Proposition 71, there is a net $\langle x_\alpha \rangle$ in $f^{-1}(V^c)$ that converges to x. But then $f(x_\alpha) \notin V$, so $f(x_\alpha) \not\to f(x)$. □

Definition 73. *A* ***subnet*** *of a net $\langle x_\alpha \rangle_{\alpha \in A}$ is a net $\langle y_\beta \rangle_{\beta \in B}$ together with a map $\beta \mapsto \alpha_\beta$ from B to A satisfying the following properties.*

- *For every $\alpha_0 \in A$ there exists $\beta_0 \in B$ such that $\alpha_\beta \gtrsim \alpha_0$ whenever $\beta \gtrsim \beta_0$.*
- *$y_\beta = x_{\alpha_\beta}$.*

Clearly if $\langle x_\alpha \rangle$ converges to a point x, then so does any subnet $\langle x_{\alpha_\beta} \rangle$.

Warning 4.74.

The name "subnet" is used because subnets perform much the same functions as subsequences, but it should not be taken too literally, as the mapping $\beta \mapsto \alpha_\beta$ need not be injective. In particular, the index set B may well have larger cardinality than the index set A, and *a subnet of a sequence need not be a subsequence.*

The following example demonstrates how different subnets can be from subsequences.

Example 75. *Consider* $\{x_n\}_{n\in\mathbb{Z}_{\geqslant 1}}, x_n = n$*. Then* $(1, 1, 2, 2, 3, 3, \dots)$ *is a subnet of* $\{x_n\}_{n\in\mathbb{Z}_{\geqslant 1}}$ *(but not a subsequence) defined by* $\{s_{\phi(n)}\}$ *where* $\phi(n) = \lfloor (n+1)/2 \rfloor$.

Proposition 4.76: 4.20.

If $\langle x_\alpha \rangle_{\alpha\in A}$ is a net in a topological space X and $x \in X$, then

$$x \text{ is a cluster point of } \langle x_\alpha \rangle \iff \langle x_\alpha \rangle \text{ has a subnet converging to } x.$$

Proof. If $\langle y_\beta \rangle = \langle x_{\alpha_\beta} \rangle$ is a subnet converging to x and U is a neighborhood of x, choose $\beta_1 \in B$ such that $y_\beta \in U$ for $\beta \gtrsim \beta_1$. Also, given $\alpha \in A$, choose $\beta_2 \in B$ such that $\alpha_\beta \gtrsim \alpha$ for $\beta \gtrsim \beta_2$. Then there exists $\beta \in B$ with $\beta \gtrsim \beta_1$ and $\beta \gtrsim \beta_2$, and we have $\alpha_\beta \gtrsim \alpha$ and $x_{\alpha_\beta} = y_\beta \in U$. Thus $\langle x_\alpha \rangle$ is frequently in U, so x is a cluster point of $\langle x_\alpha \rangle$. Conversely, if x is a cluster point of $\langle x_\alpha \rangle$, let $\mathcal{N}$ be the set of neighborhoods of x and make $\mathcal{N} \times A$ into a directed set by declaring that $(U, \alpha) \lesssim (U', \alpha')$ if and only if $U \supset U'$ and $\alpha \lesssim \alpha'$. For each $(U, \gamma) \in \mathcal{N} \times A$ we can choose $\alpha_{(U,\gamma)} \in A$ such that $\alpha_{(U,\gamma)} \gtrsim \gamma$ and $x_{\alpha_{(U,\gamma)}} \in U$. Then if $(U', \gamma') \gtrsim (U, \gamma)$ we have $\alpha_{(U',\gamma')} \gtrsim \gamma' \gtrsim \gamma$ and $x_{\alpha(U',\gamma')} \in U' \subset U$, whence it follows that $\langle x_{\alpha(U,\gamma)} \rangle$ is a subnet of $\langle x_\alpha \rangle$ that converges to x. □

Exercise 4.77: Folland Exercise 4.30.

If A is a directed set, a subset B of A is called ***cofinal*** in A if for each $\alpha \in A$ there exists $\beta \in B$ such that $\alpha \lesssim \beta$.

(a) If B is cofinal in A and $\langle x_\alpha \rangle_{\alpha\in A}$ is a net, the inclusion map $B \to A$ makes $\langle x_\beta \rangle_{\beta\in B}$ a subnet of $\langle x_\alpha \rangle_{\alpha\in A}$.

(b) If $\langle x_\alpha \rangle_{\alpha\in A}$ is a net in a topological space, then $\langle x_\alpha \rangle$ converges to x if and only if for every cofinal $B \subset A$ there is a cofinal $C \subset B$ such that $\langle x_\gamma \rangle_{\gamma\in C}$ converges to x.

Exercise 4.78: Folland Exercise 4.31.

Let $\langle x_n \rangle_{n\in\mathbb{Z}_{\geqslant 0}}$ be a sequence.

(a) If $k \to n_k$ is a map from $\mathbb{Z}_{\geqslant 0}$ to itself, then $\langle x_{n_k} \rangle_{k\in\mathbb{Z}_{\geqslant 0}}$ is a subnet of $\langle x_n \rangle$ if and only if $n_k \to \infty$ as $k \to \infty$, and it is a subsequence (as defined in Folland Section 0.1) if and only if n_k is strictly increasing in k.

(b) There is a natural one-to-one correspondence between the subsequences of $\langle x_n \rangle$ and the subnets of $\langle x_n \rangle$ defined by cofinal sets as in Folland Exercise 4.30.

Exercise 4.79: Folland Exercise 4.32.

A topological space X is Hausdorff if and only if every net in X converges to at most one point. (If X is not Hausdorff, let x and y be distinct points with no disjoint neighborhoods, and consider the directed set $\mathcal{N}_x \times \mathcal{N}_y$ where $\mathcal{N}_x, \mathcal{N}_y$ are the families of neighborhoods of x, y.)

Exercise 4.80: Folland Exercise 4.33.

Let $\langle x_\alpha \rangle_{\alpha \in A}$ be a net in a topological space, and for each $\alpha \in A$ let $E_\alpha = \{x_\beta \mid \beta \gtrsim \alpha\}$. Then x is a cluster point of $\langle x_\alpha \rangle$ if and only if $x \in \bigcap_{\alpha \in A} \overline{E}_\alpha$.

Exercise 4.81: Folland Exercise 4.34.

If X has the weak topology generated by a family $\mathfrak{F}$ of functions, then $\langle x_\alpha \rangle$ converges to $x \in X$ if and only if $\langle f(x_\alpha) \rangle$ converges to $f(x)$ for all $f \in \mathfrak{F}$. (In particular, if $X = \prod_{i \in I} X_i$, then $x_\alpha \to x$ if and only if $\pi_i(x_\alpha) \to \pi_i(x)$ for all $i \in I$.)

Exercise 4.82: Folland Exercise 4.35.

Let X be a set and $\mathcal{A}$ the collection of all finite subsets of X, directed by inclusion. Let $f \colon X \to \mathbb{R}$ be an arbitrary function, and for $A \in \mathcal{A}$, let $z_A = \sum_{x \in A} f(x)$. Then the net $\langle z_A \rangle$ converges in $\mathbb{R}$ if and only if $\{x \mid f(x) \neq 0\}$ is a countable set $\{x_n\}_{n \in \mathbb{Z}_{\geqslant 0}}$ and $\sum_1^\infty |f(x_n)| < \infty$, in which case $z_A \to \sum_1^\infty f(x_n)$. (Cf. Folland Proposition 21.)

Exercise 4.83: Folland Exercise 4.36.

Let X be the set of Lebesgue measurable complex-valued functions on $[0, 1]$. There is no topology $\mathcal{T}$ on X such that a sequence $\langle f_n \rangle$ converges to f with respect to $\mathcal{T}$ if and only if $f_n \to f$ a.e. (Use **??** and Folland Exercise 2.30,Folland Exercise 2.31(b).)

Remark 84. *The theory of nets is sometimes called the Moore-Smith theory of convergence, after its originators. Another general theory of convergence, invented by H. Cartan and publicized by Bourbaki, is based on the notion of filters. A filter in a set X is a family $\mathcal{F} \subset \mathcal{F}(X)$ with the following properties:*

If X is a topological space, a filter $\mathcal{F}$ in X converges to $x \in X$ if every neighborhood of x belongs to $\mathcal{F}$. Filters and nets are related as follows. If $\langle x_\alpha \rangle_{\alpha \in A}$ is a net in X, its derived filter is the collection of all $E \subset X$ such that $\langle x_\alpha \rangle$ is eventually in E. On the other hand, if $\mathcal{F}$ is a filter, then $\mathcal{F}$ is a directed set under reverse inclusion, and a net $\langle x_F \rangle_{F \in \mathcal{F}}$ indexed by $\mathcal{F}$ is said to be associated to $\mathcal{F}$ if $x_F \in F$ for all $F \in \mathcal{F}$. It is then easy to verify that a net $\langle x_\alpha \rangle$ converges to x if and only if its derived filter converges to x, and a filter $\mathcal{F}$ converges to x if and only if all of its associated nets converge to x.

4.4 Compact Spaces

In Folland Section 0.6 there are three equivalent characterizations of compactness for metric spaces: the Heine-Borel property, the Bolzano-Weierstrass property, and completeness plus total boundedness. Only the first two of these make sense for general topological spaces, and it is the first one that turns out to be the most useful.

Definition 85. *We call a topological space X* ***compact*** *if whenever $\{U_\alpha\}_{\alpha\in A}$ is an* ***open cover*** *of X—that is, a collection of open sets such that $X = \bigcup_{\alpha\in A} U_\alpha$—there exists a finite subset B of A such that $X = \bigcup_{\alpha\in B} U_\alpha$.*

A subset Y of a topological space X is called ***compact*** *if it is compact in the relative topology; thus $Y \subset X$ is compact if and only if whenever $\{U_\alpha\}_{\alpha\in A}$ is a collection of open subsets of X with $Y \subset \bigcup_{\alpha\in A} U_\alpha$, there is a finite $B \subset A$ with $Y \subset \bigcup_{\alpha\in B} U_\alpha$. Furthermore, Y is called* ***precompact*** *if its closure is compact.*

To be brief (although somewhat sylleptic, since the adjectives "open" and "finite" refer to different things), we say "X is compact if every open cover of X has a finite subcover."

DeMorgan's laws lead to the following characterization of compactness in terms of closed sets.

Definition 86. *A family $\{F_\alpha\}_{\alpha\in A}$ of subsets of X is said to have the* ***finite intersection property*** *if $\bigcap_{\alpha\in B} F_\alpha \neq \varnothing$ for all finite $B \subset A$.*

Proposition 4.87: 4.21.

A topological space X is compact if and only if for every family $\{F_\alpha\}_{\alpha\in A}$ of closed sets with the finite intersection property, $\bigcap_{\alpha\in A} F_\alpha \neq \varnothing$.

Proof. Let $U_\alpha = (F_\alpha)^c$. Then U_α is open, $\bigcap_{\alpha\in A} F_\alpha \neq \varnothing$ if and only if $\bigcup_{\alpha\in A} U_\alpha \neq X$, and $\{F_\alpha\}$ has the finite intersection property if and only if no finite subfamily of $\{U_\alpha\}$ covers X. The result follows. □

We now list several basic facts about compact spaces.

Proposition 4.88: 4.22.

A closed subset of a compact space is compact.

Proof. If X is compact, $F \subset X$ is closed, and $\{U_\alpha\}_{\alpha\in A}$ is a family of open sets in X with $F \subset \cup_{\alpha\in A} U_\alpha$, then $\{U_\alpha\}_{\alpha\in A} \cup \{F^c\}$ is an open cover of X. It has a finite subcover, so by discarding F^c from the latter if necessary, we obtain a finite subcollection of $\{U_\alpha\}_{\alpha\in A}$ that covers F. □

Proposition 4.89: 4.23.

If F is a compact subset of a Hausdorff space X and $x \notin F$, there are disjoint open sets U, V such that $x \in U$ and $F \subset V$.

Proof. For each $y \in F$, choose disjoint open U_y and V_y with $x \in U_y$ and $y \in V_y$. $\{V_y\}_{y \in F}$ is an open cover of F, so it has a finite subcover $\{V_{y_j}\}_1^n$. Then $U = \bigcap_1^n U_{y_j}$ and $V = \bigcup_1^n V_{y_j}$ have the desired properties. □

Proposition 4.90: 4.24.

Every compact subset of a Hausdorff space is closed.

Proof. According to Proposition 89, if F is compact then F^c is a neighborhood of each of its points,and hence is open. □

Warning 4.91.

In a non-Hausdorff space, compact sets need not be closed (for example, every subset of a space with the trivial topology is compact), and the intersection of compact sets need not be compact; see Folland Exercise 4.37.

Remark 92. *Despite Warning 91, in a Hausdorff space the intersection of any family of compact sets is compact by Propositions 88 and 90. Moreover, in an arbitrary topological space a finite union of compact sets is always compact. (If $K_1, \ldots K_n$ are compact and $\{U_\alpha\}$ is an open cover of $\bigcup_1^n K_j$, choose a finite subcover of each K_j and combine them.)*

Proposition 4.93: 4.25.

Every compact Hausdorff space is normal.

Proof. Suppose that X is compact Hausdorff and E, F are disjoint closed subsets of X. By Proposition 89, for each $x \in E$ there exist disjoint open sets U_x, V_x with $x \in U_x, F \subset V_x$. By Proposition 88, E is compact, and $\{U_x\}_{x \in E}$ is an open cover of E, so there is a finite subcover $\{U_{x_j}\}_1^n$. Let $U = \bigcup_1^n U_{x_j}$ and $V = \bigcap_1^n V_{x_j}$. Then U and V are disjoint open sets with $E \subset U$ and $F \subset V$. □

Proposition 4.94: 4.26.

If X is compact and $f\colon X \to Y$ is continuous, then $f(X)$ is compact.

Proof. Let $\{V_\alpha\}$ be an open cover of $f(X)$ in Y. Then $\{f^{-1}(V_\alpha)\}$ is an open cover of X, so it has a finite subcover $\{f^{-1}(V_{\alpha_j})\}$, and $\{V_{\alpha_j}\}$ is then a finite subcover of $f(X)$. □

Corollary 4.95: 4.27.

If X is compact, then $C(X) = BC(X)$.

Proposition 4.96: 4.28.

If X is compact and Y is Hausdorff, then any continuous bijection $f\colon X \to Y$ is a homeomorphism.

Proof. If $E \subset X$ is closed, then E is compact, hence $f(E)$ is compact, hence $f(E)$ is closed, by Propositions 88, 90 and 94. This means that f^{-1} is continuous, so f is a homeomorphism. □

We now show that a version of the Bolzano-Weierstrass property holds for compact topological spaces. As one might suspect, it is merely necessary to replace sequences by nets.

Theorem 4.97: 4.29.

If X is a topological space, the following are equivalent.

(a) X is compact.
(b) Every net in X has a cluster point.
(c) Every net in X has a convergent subnet.

Proof. The equivalence of (b) and (c) follows from Proposition 76. If X is compact and $\langle x_\alpha \rangle$ is a net in X, let $E_\alpha = \{x_\beta \mid \beta \gtrsim \alpha\}$. Since for any $\alpha, \beta \in A$ there exists $\gamma \in A$ with $\gamma \gtrsim \alpha$ and $\gamma \gtrsim \beta$, the family $\{E_\alpha\}_{\alpha \in A}$ has the finite intersection property, so by Proposition 87, $\bigcap_{\alpha \in A} \overline{E}_\alpha \neq \varnothing$. If $x \in \bigcap_{\alpha \in A} \overline{E}_\alpha$ and U is a neighborhood of x, then U intersects each E_α, which means that $\langle x_\alpha \rangle$ is frequently in U, so x is a cluster point of $\langle x_\alpha \rangle$. On the other hand, if X is not compact, let $\{U_\beta\}_{\beta \in B}$ be an open cover of X with no finite subcover. Let $\mathcal{A}$ be the collection of finite subsets of B, directed by inclusion, and for each $A \in \mathcal{A}$ let x_A be a point in $(\bigcup_{\beta \in A} U_\beta)^c$. Then $\langle x_A \rangle_{A \in \mathcal{A}}$ is a net with no cluster point. Indeed, if $x \in X$, choose $\beta \in B$ with $x \in U_\beta$. If $A \in \mathcal{A}$ and $A \gtrsim \{\beta\}$ then $x_A \notin U_\beta$, so x is not a cluster point of $\langle x_A \rangle$. □

We conclude by mentioning two other useful concepts related to compactness.

Definition 98. *Let X be a topological space.*

- *We call X **countably compact** if every countable open cover of X has a finite subcover.*
- *We call X **sequentially compact** if every sequence in X has a convergent subsequence.*

Of course, every compact space is countably compact, and for metric spaces compactness and sequential compactness are equivalent. However, in general there is no relation between compactness and sequential compactness. See Exercises Folland Exercise 4.39,Folland Exercise 4.40,Folland Exercise 4.41,Folland Exercise 4.42,Folland Exercise 4.43 for further results and examples.

Exercise 4.99: Folland Exercise 4.37.

Let $0'$ denote a point that is is not an element of $(-1,1)$, and let $X = (-1,1) \cup \{0'\}$. Let $\mathfrak{T}$ be the topology on X generated by the sets $(-1,a), (a,1), [(-1,b) \setminus \{0\}] \cup \{0'\}$, and $[(c,1) \setminus \{0\}] \cup \{0'\}$ where $-1 < a < 1, 0 < b < 1$, and $-1 < c < 0$. (One should picture X as $(-1,1)$ with the point 0 split in two.)

(a) Define $f, g\colon (-1,1) \to X$ by $f(x) = x$ for all x, $g(x) = x$ for $x \neq 0$, and $g(0) = 0'$. Then f and g are homeomorphisms onto their ranges.

(b) X is T_1 but not Hausdorff, although each point of X has a neighborhood that is homeomorphic to $(-1,1)$ (and hence is Hausdorff).

(c) The sets $\left[-\frac{1}{2}, \frac{1}{2}\right]$ and $\left(\left[-\frac{1}{2}, \frac{1}{2}\right] \setminus \{0\}\right) \cup \{0'\}$ are compact but not closed in X, and their intersection is not compact.

Exercise 4.100: Folland Exercise 4.38.

Suppose that $(X, \mathfrak{T})$ is a compact Hausdorff space and $\mathfrak{T}'$ is another topology on X. If $\mathfrak{T}'$ is strictly stronger than $\mathfrak{T}$, then $(X, \mathfrak{T}')$ is Hausdorff but not compact. If $\mathfrak{T}'$ is strictly weaker than $\mathfrak{T}$, then $(X, \mathfrak{T}')$ is compact but not Hausdorff.

Solution. If $\mathfrak{T}$ is strictly stronger than $\mathfrak{T}$ but is compact then $(X, \mathfrak{T}') \to (X, \mathfrak{T})$ via $x \mapsto x$ is a continuous bijection of a compact space onto a Hausdorff space, hence is a homeomorphism. But then f is a bijection of $\mathfrak{T}$ onto $\mathfrak{T}'$, a contradiction. But $\mathfrak{T}'$ is Hausdorff, since the same separating sets from $\mathfrak{T} \subset \mathfrak{T}'$ are available for use.

If $\mathfrak{T}'$ is strictly weaker then $\mathfrak{T}$, then $(X, \mathfrak{T}) \to (X, \mathfrak{T}')$ via $x \mapsto x$ is a continuous bijection of a compact space onto a Hausdorff space, hence is a homeomorphism, so we reach the same contradiction as above. But $(X, \mathfrak{T}')$ is compact, since any open cover of X in $\mathfrak{T}'$ is an open cover of X in $\mathfrak{T}$, which by compactness of $(X, \mathfrak{T})$ means there exists a finite subcover. □

Exercise 4.101: Folland Exercise 4.39.

Every sequentially compact space is countably compact.

Solution. Suppose for a contradiction X is sequentially compact but not countably compact, so there exists a countable open cover $\{U_j\}_{j=1}^{\infty}$ of X without a finite subcover.

For all j, pick

$$x_j \in X \setminus \Big(\bigcup_{r=1}^{j-1} U_r\Big).$$

Since X is sequentially compact, there exists a subsequence $\{x_{j_j}\}_{k=1}^{\infty}$ converging to some $x \in X$. But then x is not in any of the U_j, since if $x \in U_\ell$ for some ℓ then so are x_{j_j} for all $k \geqslant 0$ (by definition of convergence), contradicting $x_{j_{k+1}} \in X \setminus (\bigcup_{r=1}^{j_k-1} U_r) \subset U_k^c$ (since then $x_{j_k} \in U$ and $x_{j_k} \in U_k^c$). □

Exercise 4.102: Folland Exercise 4.40.

If X is countably compact, then every sequence in X has a cluster point. If X is also first countable, then X is sequentially compact.

Exercise 4.103: Folland Exercise 4.41.

A T_1 space X is countably compact if and only if every infinite subset of X has an accumulation point.

Exercise 4.104: Folland Exercise 4.42.

The set of countable ordinals (see Folland Section 0.4) with the order topology (Proposition 34) is sequentially compact and first countable but not compact. (To prove sequential compactness, use Folland Proposition 20.)

Exercise 4.105: Folland Exercise 4.43.

For $x \in [0, 1)$, let $\sum_1^\infty a_n(x)2^{-n}$, where $a_n(x) \in \{0, 1\}$ be the base-2 decimal expansion of x. (If x is a dyadic rational, choose the expansion such that $a_n(x) = 0$ for n large.) Then the sequence $\langle a_n \rangle$ in $\{0, 1\}^{[0,1)}$ has no pointwise convergent subsequence. Hence $\{0, 1\}^{[0,1)}$, with the product topology arising from the discrete topology on $\{0, 1\}$, is not sequentially compact. (It is, however, compact, as we shall show in next section.)

Exercise 4.106: Folland Exercise 4.44.

If X is countably compact and $f : X \to Y$ is continuous, then $f(X)$ is countably compact.

Solution. Note that the proof of this result is the same as the classical result for compact spaces, after replacing "open cover" with "countable open cover." But for the sake of completeness, we give the proof here. Consider any countable open cover, $\{U_n\}_{n \in \mathbb{Z}_{\geqslant 1}}$, of $f(X)$. Since f is continuous, $f^{-1}(U_n)$ is open for each $n \in \mathbb{Z}_{\geqslant 1}$ so that $\{f^{-1}(U_n)\}_{n \in \mathbb{Z}_{\geqslant 1}}$ is a countable open cover of X. But since X is countably compact, there exists a finite subcover $\{f^{-1}(U_{n_k})\}_{k=1}^m$ such that

$$X \subset \bigcup_{k=1}^m f^{-1}(U_{n_k}) = f^{-1}\Big(\bigcup_{k=1}^m U_{n_k}\Big).$$

This implies that

$$f(X) \subset f\Big(f^{-1}\Big(\bigcup_{k=1}^m U_{n_k}\Big)\Big) \subset \bigcup_{k=1}^m U_{n_k}$$

hence, the finite collection $\{U_{n_k}\}_{k=1}^m$ also covers $f(X)$, finishing the proof. ☐

Exercise 4.107: Folland Exercise 4.45.

If X is normal, then X is countably compact if and only if $C(X) = BC(X)$. (Use Folland Exercise 4.40 and Folland Exercise 4.44. If $\langle x_n \rangle$ is a sequence in X with no cluster point, then $\{x_n \mid n \in \mathbb{Z}_{\geqslant 0}\}$ is closed, and Corollary 49 applies.)

4.5 Locally Compact Hausdorff Spaces

Definition 108.

A topological space is called ***locally compact*** *if every point has a compact neighborhood.*

We shall be mainly concerned with locally compact Hausdorff spaces, which we call ***LCH*** spaces.

Proposition 4.109: 4.30.

If X is an LCH space, $U \subset X$ is open, and $x \in U$, there is a compact neighborhood N of x such that $N \subset U$.

Proof. We may assume $\overline{U}$ is compact; otherwise, replace U by $U \cap F^\circ$ where F is a compact neighborhood of x. By Proposition 89, there are disjoint relatively open sets V, W in $\overline{U}$ with $x \in V$ and $\partial U \subset W$. Then V is open in X since $V \subset U$, and $\overline{V}$ is a closed and hence compact subset of $U \setminus W$. Thus we may take $N = \overline{V}$. ☐

Proposition 4.110: 4.31.

If X is an LCH space and $K \subset U \subset X$ where K is compact and U is open, there exists a precompact open V such that $K \subset V \subset \overline{V} \subset U$.

Proof. By Proposition 109, for each $x \in K$ we can choose a compact neighborhood N_x of x with $N_x \subset U$. Then $\{N_x^\circ\}_{x \in K}$ is an open cover of K, so there is a finite subcover $\{N_{x_j}^\circ\}_1^n$. Let $V = \bigcup_1^n N_{x_j}^\circ$; then $K \subset V$ and $\overline{V} = \bigcup_1^n N_{x_j}$ is compact and contained in U. ☐

Lemma 4.111: 4.32: Urysohn's Lemma—Locally Compact Version.

If X is an LCH space and $K \subset U \subset X$ where K is compact and U is open, there exists $f \in C(X, [0,1])$ such that $f = 1$ on K and $f = 0$ outside a compact subset of U.

Proof. Let V be as in Proposition 110. Then $\overline{V}$ is normal by Proposition 93, so by Urysohn's lemma (Theorem 47) there exists $f \in C(\overline{V}, [0,1])$ such that $f = 1$ on K and $f = 0$ on ∂V. We extend f to X by setting $f = 0$ on $\overline{V}^c$. Suppose that $E \subset [0,1]$ is closed. If $0 \notin E$ we have $f^{-1}(E) = (f|_{\overline{V}})^{-1}(E)$, and if $0 \in E$ we have $f^{-1}(E) = (f|_{\overline{V}})^{-1}(E) \cup \overline{V}^c = (f|_{\overline{V}})^{-1}(E) \cup V^c$ since $(f|_{\overline{V}})^{-1}(E) \supset \partial V$. In either case, $f^{-1}(E)$ is closed, so f is continuous. □

Corollary 4.112: 4.33.

Every LCH space is completely regular.

Theorem 4.113: 4.34: Tietze Extension Theorem—Locally Compact Version.

Suppose that X is an LCH space and $K \subset X$ is compact. If $f \in C(K)$, there exists $F \in C(X)$ such that $F|_K = f$. Moreover, F may be taken to vanish outside a compact set.

The proof is similar to that of Lemma 111; details are left to the reader (Folland Exercise 4.46).

The preceding results show that LCH spaces have a rich supply of continuous functions that vanish outside compact sets. Let us introduce some terminology:

Definition 114. *If X is a topological space and $f \in C(X)$, the **support** of f, denoted by $\operatorname{supp}(f)$, is the smallest closed set outside of which f vanishes, that is, the closure of $\{x \in X \mid f(x) \neq 0\}$. If $\operatorname{supp}(f)$ is compact, we say that f is **compactly supported**, and we define*

$$C_c(X) = \{f \in C(X) \mid f \text{ is compactly supported}\}.$$

Definition 115. *If $f \in C(X)$, we say that f **vanishes at infinity** if for every $\varepsilon > 0$ the set $\{x \in X \mid |f(x)| \geqslant \varepsilon\}$ is compact, and we define*

$$C_0(X) = \{f \in C(X) \mid f \text{ vanishes at infinity}\}.$$

Clearly $C_c(X) \subset C_0(X)$. Moreover, $C_0(X) \subset BC(X)$, because for $f \in C_0(X)$ the image of the set $\{x \mid |f(x)| \geqslant \varepsilon\}$ is compact, and $|f| < \varepsilon$ on its complement.

Proposition 4.116: 4.35.

If X is an LCH space, $C_0(X)$ is the closure of $C_c(X)$ in the uniform metric.

Proof. If $\{f_n\}$ is a sequence in $C_c(X)$ that converges uniformly to $f \in C(X)$, for each $\varepsilon > 0$ there exists $n \in \mathbb{Z}_{\geqslant 0}$ such that $\|f_n - f\|_u < \varepsilon$. Then $|f(x)| < \varepsilon$ if $x \notin \operatorname{supp}(f_n)$, so $f \in C_0(X)$. Conversely, if $f \in C_0(X)$, for $n \in \mathbb{Z}_{\geqslant 0}$ let $K_n = \{x \mid |f(x)| \geqslant n^{-1}\}$. Then K_n is compact, so by Lemma 111 there exists $g_n \in C_c(X)$ with $0 \leqslant g_n \leqslant 1$ and $g_n = 1$ on K_n. Let $f_n = g_n f$. Then $f_n \in C_c(X)$ and $\|f_n - f\|_u \leqslant n^{-1}$, so $f_n \to f$ uniformly. □

If X is a noncompact LCH space, it is possible to make X into a compact space by adding a single point "at infinity" in such a way that the functions in $C_0(X)$ are precisely those continuous functions f such that $f(x) \to 0$ as x approaches the point at infinity. More precisely, let ∞ denote a point that is not an element of X, let $X^* = X \cup \{\infty\}$, and let $\mathfrak{T}$ be the collection of all subsets of X^* such that either (i) U is an open subset of X, or (ii) $\infty \in U$ and U^c is a compact subset of X.

Proposition 4.117: 4.36.

If X, X^*, and $\mathfrak{T}$ are as above, then $(X^*, \mathfrak{T})$ is a compact Hausdorff space, and the inclusion map $i\colon X \to X^*$ is an embedding. Moreover, if $f \in C(X)$, then f extends continuously to X^* if and only if $f = g + c$ where $g \in C_0(X)$ and c is a constant, in which case the continuous extension is given by $f(\infty) = c$.

The proof is straightforward and is left to the reader (Folland Exercise 4.47).

Definition 118. *The space X^* is called the* ***one-point compactification*** *or* ***Alexandroff compactification*** *of X.*

If X is a topological space, the space $\mathbb{C}^X$ of all complex-valued functions on X can be topologized in various ways. One way, of course, is the product topology, that is, the topology of pointwise convergence. Another is the topology of uniform convergence, which is generated by the sets

$$\{g \in \mathbb{C}^X \mid \sup_{x \in X} |g(x) - f(x)| < n^{-1}\} \quad (n \in \mathbb{Z}_{\geqslant 0}, f \in \mathbb{C}^X).$$

The proof of Proposition 45 shows that $C(X)$ is a closed subspace of $\mathbb{C}^X$ in the topology of uniform convergence. Intermediate between these two topologies is the following topology.

Definition 119. *If X is a topological space, then the* ***topology of uniform convergence on compact sets*** *on $\mathbb{C}^X$ is the topology generated by the sets*

$$\{g \in \mathbb{C}^X \mid \sup_{x \in K} |g(x) - f(x)| < 1/n, n \in \mathbb{Z}_{\geqslant 0}, f \in \mathbb{C}^X, K \subset X \text{ compact}\}$$

We shall now examine this topology in the case where X is an LCH space.

Lemma 4.120: 4.37.

If X is an LCH space and $E \subset X$, then E is closed if and only if $E \cap K$ is closed for every compact $K \subset X$.

Proof. If E is closed, then $E \cap K$ is closed by Propositions 88 and 90. If E is not closed, pick $x \in \overline{E} \setminus E$ and let K be a compact neighborhood of x. Then x is an accumulation point of $E \cap K$ but is not in $E \cap K$, so by Proposition 3 $E \cap K$ is not closed. □

Proposition 4.121: 4.38.

If X is an LCH space, $C(X)$ is a closed subspace of $\mathbb{C}^X$ in the topology of uniform convergence on compact sets.

Proof. If f is in the closure of $C(X)$, then f is a uniform limit of continuous functions on each compact $K \subset X$, so $f \mid K$ is continuous. If $E \subset \mathbb{C}$ is closed, $f^{-1}(E) \cap K = (f \mid K)^{-1}(E)$ is thus closed for each compact K, so by Lemma 120 $f^{-1}(E)$ is closed, whence f is continuous. □

A topological space X is called ***σ-compact*** if it is a countable union of compact sets. To appreciate the significance of the next two propositions, see Folland Exercise 4.54.

Proposition 4.122: 4.39.

If X is a σ-compact LCH space, there is a sequence $\{U_n\}$ of precompact open sets such that $\overline{U}_n \subset U_{n+1}$ for all n and $X = \bigcup_1^\infty U_n$.

Proof. Suppose $X = \bigcup_1^\infty K_n$ where each K_n is compact. Every compact subset of X has a precompact open neighborhood by Proposition 110. Thus we may take U_1 to be a precompact open neighborhood of K_1, and then, proceeding inductively, take U_n to be a precompact open neighborhood of $\overline{U}_{n-1} \cup K_n$. □

Proposition 4.123: 4.40.

If X is a σ-compact LCH space and $\{U_n\}$ is as in Proposition 122, then for each $f \in \mathbb{C}^X$ the sets

$$\{g \in \mathbb{C}^X \mid \sup |g(x) - f(x)| < 1/m\} \quad (m, n \in \mathbb{Z}_{\geqslant 0})$$

form a neighborhood base for f in the topology of uniform convergence on compact sets. Hence this topology is first countable, and $f_j \to f$ uniformly on compact sets if and only if $f_j \to f$ uniformly on each $\overline{U}_n$.

Proof. These assertions follow easily from the observation that if $K \subset X$ is compact, then $\{U_n\}_1^\infty$ is an open cover of K and hence $K \subset \overline{U}_n$ for some n. Details are left to the reader. (Folland Exercise 4.48). □

We close this section with a construction that is useful in a number of situations.

Definition 124. *If X is a topological space and $E \subset X$, a* **partition of unity** *on E is a collection $\{h_\alpha\}_{\alpha \in A}$ of functions in $C(X, [0,1])$ such that*

- *each $x \in X$ has a neighborhood on which only finitely many h_αs are nonzero, and*
- *$\sum_{\alpha \in A} h_\alpha(x) = 1$ for $x \in E$.*

A partition of unity $\{h_\alpha\}$ is ***subordinate to an open cover U*** *of E if for each α there exists $U \in \mathcal{U}$ with $\operatorname{supp}(h_\alpha) \subset U$.*

Proposition 4.125: 4.41.

Let X be an LCH space, K a compact subset of X, and $\{U_j\}_1^n$ an open cover of K. There is a partition of unity on K subordinate to $\{U_j\}_1^n$ consisting of compactly supported functions.

Proof. By Proposition 109, each $x \in K$ has a compact neighborhood N_x such that $N_x \subset U_j$ for some j. Since $\{N_x^\circ\}$ is an open cover of K, there exist $x_1, \ldots, x_m$ such that $K \subset \bigcup_1^m N_{x_k}$. Let F_j be the union of those N_{x_k}s that are subsets of U_j. Then F_j is a compact subset of U_j, so by Urysohn's lemma there exist $g_1, \ldots, g_n \in C_c(X, [0,1])$ with $g_j = 1$ on F_j and $\operatorname{supp}(g_j) \subset U_j$. Since the F_js cover K we have $\sum_1^n g_k \geqslant 1$ on K, so by Urysohn again there exists $f \in C_c(X, [0,1])$ with $f = 1$ on K and $\operatorname{supp}(f) \subset \{x \mid \sum_1^n g_k(x) > 0\}$. Let $g_{n+1} = 1 - f$, so that $\sum_1^{n+1} g_k > 0$ everywhere, and for $j = 1, \ldots, n$ let $h_j = g_j / \sum_1^{n+1} g_k$. Then $\operatorname{supp}(h_j) = \operatorname{supp}(g_j) \subset U_j$ and $\sum_1^n h_j = 1$ on K. □

A generalization of this result may be found in Folland Exercise 4.57.

Exercise 4.126: Folland Exercise 4.46.

Prove Theorem 113.

Exercise 4.127: Folland Exercise 4.47.

Prove Proposition 117. Also, show that if X is Hausdorff but not locally compact, Proposition 117 remains valid except that X^* is not Hausdorff.

Exercise 4.128: Folland Exercise 4.48.

Complete the proof of Proposition 123.

Exercise 4.129: Folland Exercise 4.49.

Let X be a compact Hausdorff space and $E \subset X$.

(a) If E is open, then E is locally compact in the relative topology.
(b) If E is dense in X and locally compact in the relative topology, then E is open.

(Use Folland Exercise 4.13.)
(c) E is locally compact in the relative topology if and only if E is relatively open in $\overline{E}$.

Solution. (a) Let $x \in X$. Since X is an LCH space and E is open in X, by Proposition 109 x has a compact neighborhood K in E. Since a compact subspace is compact in the whole space, K is compact in X, so E is locally compact.

(b) Let $x \in E$. Since E is locally compact in the subspace topology, there exists a compact neighborhood K of x in E. Then K is compact in X (again since a compact subspace is compact in the whole space). Note $x \in K^\circ$ and K° is open in the subspace topology E, so there exists an open set U in X such that $K^\circ = U \cap E$. K is a compact subset of a Hausdorff space X, hence is closed in X. Then by the density of E and

$$x \in K = \overline{K^\circ} = \overline{U \cap E} \overset{(29)}{=} \overline{U},$$

so $x \in U \subset E$, hence U is open.

(c) If E is a locally compact subspace then E is dense in $\overline{E}$ by definition, so by part (b) E is open in $\overline{E}$. Conversely, if E is open in $\overline{E}$, then since $\overline{E}$ is a CH space, by part (a) E is a locally compact subspace. □

Exercise 4.130: Folland Exercise 4.50.

Let U be an open subset of a compact Hausdorff space X and U^* its one-point compactification (see Folland Exercise 4.49(a)). If $\phi\colon X \to U^*$ is defined by $\phi(x) = x$ if $x \in U$ and $\phi(x) = \infty$ if $x \in U^c$, then ϕ is continuous.

Exercise 4.131: Folland Exercise 4.51.

If X and Y are topological spaces, $\phi \in C(X, Y)$ is called ***proper*** if $\phi^{-1}(K)$ is compact in X for every compact $K \subset Y$. Suppose that X and Y are LCH spaces and X^* and Y^* are their one-point compactifications. If $\phi \in C(X, Y)$, then ϕ is proper if and only if ϕ extends continuously to a map from X^* to Y^* by setting $\phi(\infty_X) = \infty_Y$.

Exercise 4.132: Folland Exercise 4.52.

The one-point compactification of $\mathbb{R}^n$ is homeomorphic to the n-sphere $\{x \in \mathbb{R}^{n+1} \mid |x| = 1\}$.

Exercise 4.133: Folland Exercise 4.53.

Lemma 120 remains true if the assumption that X is locally compact is replaced by the assumption that X is first countable.

Exercise 4.134: Folland Exercise 4.54.

Let $\mathbb{Q}$ have the relative topology induced from $\mathbb{Q}$.

(a) $\mathbb{Q}$ is not locally compact,

(b) $\mathbb{Q}$ is σ-compact (it is a countable union of singleton sets), but uniform convergence on singletons (i.e pointwise convergence) does not imply uniform convergence on compact subsets of $\mathbb{Q}$.

Exercise 4.135: Folland Exercise 4.55.

Every open set in a second countable LCH space is σ-compact.

Exercise 4.136: Folland Exercise 4.56.

Define $\Phi\colon [0,\infty] \to [0,1]$ by $\Phi(t) = t/(t+1)$ for $t \in [0,\infty)$ and $\Phi(\infty) = 1$.

(a) Φ is strictly increasing and $\Phi(t+s) \leqslant \Phi(t) + \Phi(s)$.

(b) If (Y,ρ) is a metric space, then $\Phi \circ \rho$ is a bounded metric on Y that defines the same topology as ρ.

(c) If X is a topological space, the function $\rho(f,g) = \Phi(\sup_{x\in X} |f(x) - g(x)|)$ is a metric on $\mathbb{C}^X$ whose associated topology is the topology of uniform convergence.

(d) If X is a σ-compact LCH space and $\{U_n\}_1^\infty$ is as in Proposition 122, the function

$$\rho(f,g) = \sum\nolimits_1^\infty 2^{-n}\Phi(\sup\nolimits_{x\in\overline{U}_n} |f(x) - g(x)|)$$

is a metric on $\mathbb{C}^X$ whose associated topology is the topology of uniform convergence on compact sets.

Exercise 4.137: Folland Exercise 4.57.

An open cover U of a topological space X is called locally finite if each $x \in X$ has a neighborhood that intersects only finitely many members of U. If U, v are open covers of X, V is a refinement of U if for each $V \in \mathcal{V}$ there exists $U \in U$ with $V \subset U$. A topological space X is called ***paracompact*** if every open cover of X has a locally finite refinement.

(a) If X is a σ-compact LCH space, then X is paracompact. In fact, every open cover U has locally finite refinements $\{V_\alpha\}, \{W_\alpha\}$ such that $\overline{V}_\alpha$ is compact and $\overline{W}_\alpha \subset V_\alpha$ for all α. (Let $\{U_n\}_1^\infty$ be as in Proposition 122. For each n, $\left\{E \cap \left(U_{n+2} \setminus \overline{U}_{n-1}\right) \,\middle|\, E \subset \mathfrak{U}\right\}$ is an open cover of $\overline{U}_{n+1} \setminus U_n$. Choose a finite subcover to obtain $\{V_\alpha\}$ and mimic the beginning of the proof of Proposition 71 to obtain $\{W_\alpha\}$.)

(b) If X is a σ-compact LCH space, for any open cover U of X there is a partition of unity on X subordinate to U and consisting of compactly supported functions.

4.6 Two Compactness Theorems

The geometric objects on which one does analysis (Euclidean spaces, manifolds, and so on) tend to be compact or locally compact. However, in infinite-dimensional spaces such as spaces of functions, compactness is a rather rare phenomenon and is to be greatly prized when it is available. Almost all compactness results in such situations are obtained via two basic theorems, Tychonoff's theorem and the Arzelà-Ascoli theorem, which we present in this section.

Tychonoff's theorem has to do with compactness of Cartesian products. To prepare for it, we introduce some notation. Recall that an element x of $X = \prod_{\alpha\in A} X_\alpha$ is, strictly speaking, a mapping from A into $\bigcup_{\alpha\in A} X_\alpha$; namely, $x(\alpha) \in X_\alpha$ is the αth coordinate of x, which we generally denote by $\pi_\alpha(x)$. If $B \subset A$, there is a natural map $\pi_B : X \to \prod_{\alpha\in B} X_\alpha$; namely, $\pi_B(x)$ is the restriction of the map x to B. (In particular, $\pi_{\{\alpha\}}$ is essentially identical to π_α, and we shall not distinguish between them.) If $p \in \prod_{\alpha\in B} X_\alpha$ and $q \in \prod_{\alpha\in C} X_\alpha$, we shall say that q is an extension of p if q extends p as a mapping, that is, if $B \subset C$ and $p(\alpha) = q(\alpha)$ for $\alpha \in B$.

Theorem 4.138: 4.42: Tychonoff's Theorem.

If $\{X_\alpha\}_{\alpha\in A}$ is any family of compact topological spaces, then $X = \prod_{\alpha\in A} X_\alpha$ (with the product topology) is compact.

Proof. By Theorem 97, it is enough to show that any net $\langle x_i\rangle_{i\in I}$ in X has a cluster point. We shall do this by examining cluster points of the nets $\langle \pi_B(x_i)\rangle$ in the subproducts of X. To wit, let

$$\mathcal{P} = \bigcup_{B\subset A}\left\{p \in \prod_{\alpha\in B} X_\alpha \;\middle|\; p \text{ is a cluster point of } \langle \pi_B(x_i)\rangle\right\}$$

$\mathcal{P}$ is nonempty, because each X_α is compact and so $\langle \pi_B(x_i)\rangle$ has cluster points when $B = \{\alpha\}$. Moreover, $\mathcal{P}$ is partially ordered by extension; that is, $p \leqslant q$ if q is an extension of p as defined above.

Suppose that $\{p_l \mid l \in L\}$ is a linearly ordered subset of $\mathcal{P}$, where $p_l \in \prod_{\alpha\in B_l} X_\alpha$. Let $B^* = \bigcup_{l\in L} B_l$, and let p^* be the unique element of $\prod_{\alpha\in B^*} X_\alpha$ that extends every p_l. We claim that $p^* \in \mathcal{P}$. Indeed, from the definition of the product topology, any neighborhood of p^* contains a set of the form $\prod_{\alpha\in B^*} U_\alpha$ where each U_α is open in X_α and $U_\alpha = X_\alpha$ for all but finitely many α, say $\alpha_1, \dots, \alpha_n$. Each of these α_js belongs to some B_l, so by linearity of the ordering they all belong to a single B_l. But then $\prod_{\alpha\in B_l} U_\alpha$ is a neighborhood of p_l, so $\langle \pi_{B_l}(x_i)\rangle$ is frequently in $\prod_{\alpha\in B_l} U_\alpha$; hence $\langle \pi_{B^*}(x_i)\rangle$ is frequently in $\prod_{\alpha\in B^*} U_\alpha$, so p^* is a cluster point of $\langle \pi_{B^*}(x_i)\rangle$. Therefore p^* is an upper bound for $\{p_l\}$ in $\mathcal{P}$.

By Zorn's lemma, then, $\mathcal{P}$ has a maximal element $\overline{p} \in \prod_{\alpha\in\overline{B}} X_\alpha$. We claim that $\overline{B} = A$. If not, pick $\gamma \in A \setminus \overline{B}$. By Proposition 76 there is a subnet $\langle \pi_{\overline{B}}(x_{i(j)})\rangle_{j\in J}$ of $\langle \pi_{\overline{B}}(x_i)\rangle$ that converges to $\overline{p}$, and since X_γ is compact, there is a subnet $\langle \pi_\gamma(x_{i(j(k))})\rangle_{k\in K}$ of $\langle \pi_\gamma(x_{i(j)})\rangle$ that converges to some $p_\gamma \in X_\gamma$. Let q be the unique element of $\prod_{\alpha\in\overline{B}\cup\{\gamma\}} X_\alpha$ that extends both $\overline{p}$ and p_γ; then the net $\langle \pi_{\overline{B}\cup\{\gamma\}}(x_{i(j(k))})\rangle_{k\in K}$ converges to q and hence q is a cluster point of $\langle \pi_{\overline{B}\cup\{\gamma\}}(x_i)\rangle$, contradicting the maximality of $\overline{p}$. Therefore $\overline{p}$ is a cluster point of $\langle x_i\rangle$, and we are done. □

We now turn to the Arzelà-Ascoli theorem, which has to do with compactness in spaces of continuous mappings. There are several variants of this result; the theorems below are two of the most useful ones. See also Folland Exercise 4.61.

Definition 139. *If X is a topological space, $x \in X$, and let $\mathfrak{F}$ be any subset of $C(X)$.*

- *$\mathfrak{F}$ is called **equicontinuous at x** if for every $\varepsilon > 0$ there is a neighborhood U of x such that $|f(y) - f(x)| < \varepsilon$ for all $y \in U$ and all $f \in \mathfrak{F}$, and $\mathfrak{F}$ is called **equicontinuous** if it is equicontinuous at each $x \in X$.*
- *$\mathfrak{F}$ is said to be **pointwise bounded** if $\{f(x) \mid f \in \mathfrak{F}\}$ is a bounded subset of $\mathbb{C}$ for each $x \in X$.*

Theorem 4.140: 4.43: Arzelà-Ascoli Theorem I.

Let X be a compact Hausdorff space. If $\mathfrak{F}$ is an equicontinuous, pointwise bounded subset of $C(X)$, then $\mathfrak{F}$ is totally bounded in the uniform metric, and the closure of $\mathfrak{F}$ in $C(X)$ (with respect to the uniform metric) is compact.

Proof. Suppose $\varepsilon > 0$. Since $\mathfrak{F}$ is equicontinuous, for each $x \in X$ there is an open neighborhood U_x of x such that $|f(y) - f(x)| < \varepsilon/4$ for all $y \in U_x$ and all $f \in \mathfrak{F}$. Since X is compact, we can choose $x_1, \dots, x_n \in X$ such that $\bigcup_1^n U_{x_j} = X$. Then by pointwise boundedness, $\{f(x_j) \mid f \in \mathfrak{F}, 1 \leqslant j \leqslant n\}$ is a bounded subset of $\mathbb{C}$, so there is a finite set $\{z_1, \dots, z_m\} \subset \mathbb{C}$ that is $\varepsilon/4$-dense in it—that is, each $f(x_j)$ is at a distance less than $\varepsilon/4$ from some z_k. Let $A = \{x_1, \dots, x_n\}$ and $B = \{z_1, \dots, z_m\}$; then the set B^A of functions from A to B is finite. For each $\phi \in B^A$, let

$$\mathfrak{F}_\phi = \{f \in \mathfrak{F} \mid |f(x_j) - \phi(x_j)| < \varepsilon/4 \text{ for } 1 \leqslant j \leqslant n\}.$$

Then clearly $\bigcup_{\phi\in B^A} \mathfrak{F}_\phi = \mathfrak{F}$, and we claim that each $\mathfrak{F}_\phi$ has diameter at most ε, so we obtain a finite ε-dense subset of $\mathfrak{F}$ by picking one f from each nonempty $\mathfrak{F}_\phi$. To prove the claim, suppose $f, g \in \mathfrak{F}_\phi$. Since $|f - \phi| < \varepsilon/4$ and $|g - \phi| < \varepsilon/4$ on A, we have $|f - g| < \varepsilon/2$ on A. If $x \in X$, we have $x \in U_{x_j}$ for some j, and then

$$|f(x) - g(x)| \leqslant |f(x) - f(x_j)| + |f(x_j) - g(x_j)| + |g(x_j) - g(x)| < \varepsilon.$$

This shows that $\mathfrak{F}$ is totally bounded. Since the closure of a totally bounded set is totally bounded and $C(X)$ is complete, the theorem is proved. □

Theorem 4.141: 4.44: Arzelà-Ascoli Theorem II.

Let X be a σ-compact LCH space. If $\{f_n\}$ is an equicontinuous, pointwise bounded sequence in $C(X)$, there exist $f \in C(X)$ and a subsequence of $\{f_n\}$ that converges to

f uniformly on compact sets.

Proof. By Proposition 122 there is a sequence $\{U_k\}$ of precompact open sets such that $\overline{U}_k \subset U_{k+1}$ and $X = \bigcup_1^\infty U_k$. By Theorem 140 there is a subsequence $\{f_{n_j}\}_{j=1}^\infty$ of $\{f_n\}$ that is uniformly Cauchy on $\overline{U}_1$; we denote it by $\{f_j^1\}_{j=1}^\infty$. Proceeding inductively, for $k \in \mathbb{Z}_{\geqslant 0}$ we obtain a subsequence $\{f_j^k\}_{j=1}^\infty$ of $\{f_j^{k-1}\}_{j=1}^\infty$ that is uniformly Cauchy on $\overline{U}_k$. Let $g_k = f_k^k$; then $\{g_k\}$ is a subsequence of $\{f_n\}$ which is (except for the first $k-1$ terms) a subsequence of $\{f_j^k\}$ and hence is uniformly Cauchy on each $\overline{U}_k$. Let $f = \lim g_k$. Then $f \in C(X)$ and $g_k \to f$ uniformly on compact sets by Propositions 121 and 123. □

Exercise 4.142: Folland Exercise 4.58.

If $\{X_\alpha\}_{\alpha \in A}$ is a family of topological spaces of which infinitely many are noncompact, then every closed compact subset of $\prod_{\alpha \in A} X_\alpha$ is nowhere dense.

Exercise 4.143: Folland Exercise 4.59.

The product of finitely many locally compact spaces is locally compact.

Exercise 4.144: Folland Exercise 4.60.

The product of countably many sequentially compact spaces is sequentially compact. (Use the "diagonal trick" as in the proof of Theorem 141.)

Exercise 4.145: Folland Exercise 4.61.

Theorem 140 remains valid for maps from a compact Hausdorff space X into a complete metric space Y provided the hypothesis of pointwise boundedness is replaced by pointwise total boundedness. (Make this statement precise and then prove it.)

Exercise 4.146: Folland Exercise 4.62.

Rephrase Theorem 141 in a form similar to Theorem 140 by using the metric in Folland Exercise 4.56(d).

Exercise 4.147: Folland Exercise 4.63.

Let $K \in C([0,1] \times [0,1])$. For $f \in C([0,1])$, let $Tf(x) = \int_0^1 K(x,y)f(y)dy$. Then $Tf \in C([0,1])$, and $\{Tf \mid \|f\|_u \leqslant 1\}$ is precompact in $C([0,1])$.

Solution. If $f = 0$ then $Tf = 0 \in C([0,1])$. For $f \neq 0$, let $\varepsilon > 0$ and choose $\delta > 0$ such that $|K(z_1) - K(z_2)| < \frac{\varepsilon}{\|f\|_u}$ for all $z_1, z_2 \in [0,1]^2$ with $|z_1 - z_2| < \delta$. Then, for $x_1, x_2 \in [0,1]$

satisfying $|x_1 - x_2| < \delta$,

$$|Tf(x_1) - Tf(x_2)| \leqslant \int_0^1 |K(x_1,y) - K(x_2,y)||f(y)|\,dy \leqslant \int_0^1 \frac{\varepsilon}{\|f\|_u}|f(y)|\,dy = \varepsilon,$$

showing $Tf \in C([0,1])$. For $\varepsilon > 0$, choose $\delta > 0$ such that $|K(z_1) - K(z_2)| < \varepsilon$ for $|z_1 - z_2| < \delta$. Then, for $f \in C([0,1])$ with $0 < \|f\|_u \leqslant 1$,

$$|Tf(x)| \leqslant \int_0^1 |K(x,y)||f(y)|\,dy \leqslant \int_0^1 |K(x,y)|\,dy,$$

implying $\{Tf \mid \|f\|_u \leqslant 1\}$ is pointwise bounded. Hence, by the Arzelà-Ascoli theorem, it is precompact. □

Exercise 4.148: Folland Exercise 4.64.

Let (X,ρ) be a metric space. A function $f \in C(X)$ is called Hölder continuous of exponent $\alpha(\alpha > 0)$ if the quantity

$$N_\alpha(f) = \sup_{x \neq y} \frac{|f(x) - f(y)|}{\rho(x,y)^\alpha}$$

is finite. If X is compact, $\{f \in C(X) \mid \|f\|_u \leqslant 1 \text{ and } N_\alpha(f) \leqslant 1\}$ is compact in $C(X)$.

Solution. We will use the Arzelà-Ascoli theorem. Let $\alpha > 0$ and $\mathfrak{F} = \{f \in C(X) \mid \|f\|_u \leqslant 1, N_\alpha(f) \leqslant 1\}$.

- $\mathfrak{F}$ is pointwise bounded: This is immediate because $\|f\|_u \leqslant 1$ for all $f \in \mathfrak{F}$.
- $\mathfrak{F}$ is equicontinuous: Let $\varepsilon > 0$. We want $\delta > 0$ such that for all $f \in \mathfrak{F}$, $|f(x) - f(y)| < \varepsilon$ whenever $\rho(x,x_0) < \delta$. For all $x \neq x_0$ and all $f \in \mathfrak{F}$, we have

$$\frac{|f(x) - f(x_0)|}{\rho(x,x_0)^\alpha} \leqslant N_\alpha(f) \leqslant 1 \implies |f(x) - f(x_0)| \leqslant \rho(x,x_0)^\alpha,$$

  so choosing $\delta = \varepsilon^{1/\alpha}$ works.

Hence by Arzelà-Ascoli $\mathfrak{F}$ is precompact in $C(X)$ with respect to the uniform norm, so it suffices (since $(C(X), \|-\|_u)$ is a metric space) to show any uniform limit of elements of $\mathfrak{F}$ is in $\mathfrak{F}$.

To that end suppose $\{f_n\}_{n=1}^\infty \subset \mathfrak{F}$ and $f_n \to f$ uniformly. Now $f \in C(X)$ as a uniform limit of continuous functions, so we need to show $\|f\|_u, N_\alpha(f) \leqslant 1$. Let $\varepsilon > 0$. For all $n \gg 0$ $\|f\|_u \leqslant \|f_n\|_u + \varepsilon = 1 + \varepsilon$ by the triangle inequality, hence $\|f\|_u \leqslant 1$.

It remains to show $N_\alpha(f) \leqslant 1$. For all sufficiently large n, by the triangle inequality we have for all $x \neq y$ that

$$\begin{aligned}|f(x) - f(y)| &\leqslant |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| \\ &< \varepsilon/2 + |f_n(x) - f_n(y)| + \varepsilon/2 = |f_n(x) - f_n(y)| + \varepsilon \leqslant \rho(x,x_0)^\alpha + \varepsilon,\end{aligned}$$

so $|f(x) - f(y)|/\rho(x,x_0)^\alpha \leqslant 1$. Taking the supremum over all $x \neq y$, we conclude $N_\alpha(f) \leqslant 1$, hence $f \in \mathfrak{F}$. $\mathfrak{F}$ is closed with respect to the uniform norm, so since $\mathfrak{F}$ is precompact we conclude $\mathfrak{F}$ is compact. □

Exercise 4.149: Folland Exercise 4.65.

Let U be an open subset of $\mathbb{C}$, and let $\{f_n\}$ be a sequence of holomorphic functions on U. If $\{f_n\}$ is uniformly bounded on compact subsets of U, there is a subsequence that converges uniformly to a holomorphic function on compact subsets of U. (Use the Cauchy integral formula to obtain equicontinuity.)

4.7 The Stone-Weierstrass Theorem

In this section we prove a far-reaching generalization of the well-known theorem of Weierstrass to the effect that any continuous function on a compact interval $[a, b]$ is the uniform limit of polynomials on $[a, b]$. Throughout this section, X will denote a compact Hausdorff space, and we equip the space $C(X)$ with the uniform metric.

Definition 150. *A subset $\mathcal{A}$ of $C(X, \mathbb{R})$ or $C(X)$ is said to* ***separate points*** *if for every $x, y \in X$ with $x \neq y$ there exists $f \in \mathcal{A}$ such that $f(x) \neq f(y)$.*

$\mathcal{A}$ is called an ***algebra*** *if it is a real (resp. complex) vector subspace of $C(X, \mathbb{R})$ (resp. $C(X)$) such that $fg \in \mathcal{A}$ whenever $f, g \in \mathcal{A}$.*

If $\mathcal{A} \subset C(X, \mathbb{R}), \mathcal{A}$ is called a ***lattice*** *if $\max(f, g)$ and $\min(f, g)$ are in $\mathcal{A}$ whenever $f, g \in \mathcal{A}$.*

Since the algebra and lattice operations are continuous, one easily sees that if $\mathcal{A}$ is an algebra or a lattice, so is its closure $\overline{\mathcal{A}}$ in the uniform metric.

Theorem 4.151: 4.45: The Stone-Weierstrass Theorem.

Let X be a compact Hausdorff space. If $\mathcal{A}$ is a closed subalgebra of $C(X, \mathbb{R})$ that separates points, then either $\mathcal{A} = C(X, \mathbb{R})$ or $\mathcal{A} = \{f \in C(X, \mathbb{R}) \mid f(x_0) = 0\}$ for some $x_0 \in X$. The first alternative holds if and only if $\mathcal{A}$ contains the constant functions.

Before proving Theorem 151, it will be helpful to demonstrate some of its applications.

Exercise 4.152.

If f is a continuous function on $[0, 1]$ such that

$$\int_0^1 x^n f(x)dx = 0, \quad n = 0, 1, \ldots,$$

then $f(x) = 0$ for all $x \in [0, 1]$.

Solution. Note first that for any polynomial, p_n, that converges uniformly to f (Stone-Weierstrass), we have that $\{p_n f\}$ converges to f^2 and by the DCT,

$$\int_0^1 f^2(x)dx = \lim_{n\to\infty} \int_0^1 p_n(x)f(x)dx = 0$$

by our assumption. Hence we can conclude that $f(x) = 0$ for all $x \in [0,1]$.

We can prove a similar result for bounded, continuous functions on $[1, \infty)$ with

$$\int_1^\infty x^{-n} f(x)dx = 0, \quad n = 2, 3, 4, \ldots$$

using the change of variables $x = u^{-1}$. □

Note that the form of Stone-Weierstrass presented as Corollary 161 below is useful for certain applications.

Exercise 4.153.

The algebra generated by $[1, x^2]$ is dense in $C([0,1])$.

Solution. This follows from Stone-Weierstrass since the algebra contains the constant functions and x^2 separates points in $[0,1]$. □

Remark 154. *This is not true if we replace* $[0,1]$ *with* $[-1,1]$ *as the algebra no longer separates points.*

Exercise 4.155.

Let X be a compact subset of $\mathbb{R}$. Show that $C(X)$ is a separable metric space.

Solution. The polynomials with rational coefficients form a countable set that is dense in the real coefficient polynomials, hence $C(X)$. □

You can generalize this problem as follows:

Exercise 4.156.

If (X, ρ) is a compact metric space, then $C(X)$ is a separable metric space.

Solution. Since (X, ρ) is a compact metric space, it is a separable metric space. Fix a countable dense subset $\{x_n\}_{n \in \mathbb{Z}_{\geqslant 1}}$ of X and for each $n \in \mathbb{Z}_{\geqslant 1}$, define $f_n(t) = \rho(t, x_n)$ for every $t \in X$. To complete the proof, consider the following:

(i) Show $f_n(t)$ separates points.
(ii) Then the algebra generated by $\{1, f_1(t), f_2(t), \ldots\}$ is dense by Stone-Weierstrass.
(iii) Now approximate using rational coefficients instead of real coefficients. □

The proof of Stone-Weierstrass (Theorem 151) will require several lemmas. The first one, in effect, proves the theorem when X consists of two points, and the second one is a special case of the classical Weierstrass theorem for $X = [-1,1]$. After these two we return to the general case.

Lemma 4.157: 4.46.

Consider $\mathbb{R}^2$ as an algebra under coordinate-wise addition and multiplication. Then the only subalgebras of $\mathbb{R}^2$ are $\mathbb{R}^2, \{(0,0)\}$, and the linear spans of $(1,0), (0,1)$, and $(1,1)$.

Proof. The subspaces of $\mathbb{R}^2$ listed above are evidently subalgebras. If $\mathcal{A} \subset \mathbb{R}^2$ is a nonzero algebra and $(0,0) \neq (a,b) \in \mathcal{A}$, then $(a^2, b^2) \in \mathcal{A}$. If $a \neq 0, b \neq 0$, and $a \neq b$, then (a,b) and (a^2, b^2) are linearly independent, so $\mathcal{A} = \mathbb{R}^2$. The other possibilities—$a \neq 0 = b, a = 0 \neq b$, and $a = b \neq 0$ for all nonzero $(a,b) \in \mathcal{A}$—give the other three subalgebras. □

Lemma 4.158: 4.47.

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 Taylor series for $(1-t)^{1/2}$ about the point $t = 0$:

$$(1-t)^{1/2} = 1 + \sum_1^\infty \left(-\frac{1}{2}\right)\left(\frac{1}{2}\right)\cdots\left(\frac{2n-3}{2}\right)\frac{t^n}{n!} = 1 - \sum_1^\infty c_n t^n,$$

where $c_n > 0$. By the ratio test, this series converges for $|t| < 1$; a proof that its sum is actually $(1-t)^{1/2}$ is outlined in Folland Exercise 4.66. Moreover, by the monotone convergence theorem (applied to counting measure on $\mathbb{Z}_{\geqslant 0}$),

$$\sum_1^\infty c_n = \lim_{t \nearrow 1} \sum_1^\infty c_n t^n = 1 - \lim_{t \nearrow 1}(1-t)^{1/2} = 1.$$

It follows from the finiteness of $\sum_1^\infty c_n$ that the series $1 - \sum_1^\infty c_n t^n$ converges absolutely and uniformly on $[-1,1]$, and its sum is $(1-t)^{1/2}$ there. Therefore, given $\varepsilon > 0$, by taking a suitable partial sum of this series we obtain a polynomial Q such that $|(1-t)^{1/2} - Q(t)| < \varepsilon/2$ for $t \in [-1,1]$. Setting $t = 1 - x^2$ and $R(x) = Q(1-x^2)$, we obtain a polynomial R such that $||x| - R(x)| < \varepsilon/2$ for all $x \in [-1,1]$. In particular, $|R(0)| < \varepsilon/2$, so 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 $x \in [-1,1]$. □

Lemma 4.159: 4.48.

If $\mathcal{A}$ is a closed subalgebra of $C(X, \mathbb{R})$, then $|f| \in \mathcal{A}$ whenever $f \in \mathcal{A}$, and $\mathcal{A}$ is a lattice.

Proof. If $f \in \mathcal{A}$ and $f \neq 0$, let $h = f/\|f\|_u$. Then h maps X into $[-1,1]$, so if $\varepsilon > 0$ and P is as in Lemma 158, we have $\||h| - P \circ h\|_u < \varepsilon$. Since $P(0) = 0$, P has no constant term, so $P \circ h \in \mathcal{A}$ since $\mathcal{A}$ is an algebra. Since $\mathcal{A}$ is closed and ε is arbitrary, we have $|h| \in \mathcal{A}$ and hence $|f| = \|f\|_u |h| \in \mathcal{A}$. This proves the first assertion, and the second one follows because

$$\max(f,g) = \frac{1}{2}(f + g + |f-g|), \quad \min(f,g) = \frac{1}{2}(f + g - |f-g|).$$ □

Lemma 4.160: 4.49.

Suppose $\mathcal{A}$ is a closed lattice in $C(X, \mathbb{R})$ and $f \in C(X, \mathbb{R})$. If for every $x, y \in X$ there exists $g_{xy} \in \mathcal{A}$ such that $g_{xy}(x) = f(x)$ and $g_{xy}(y) = f(y)$, then $f \in \mathcal{A}$.

Proof. Given $\varepsilon > 0$, for each $x, y \in X$ let $U_{xy} = \{z \in X \mid f(z) < g_{xy}(z) + \varepsilon\}$ and $V_{xy} = \{z \in X \mid f(z) > g_{xy}(z) - \varepsilon\}$. These sets are open and contain x and $V_{xy} = \{z \in X \mid f(z) > g_{xy}(z) - \varepsilon\}$. These sets are open and contain x and y. Fix y; then $\{U_{xy} \mid x \in X\}$ covers X, so there is a finite subcover $\{U_{x_j y}\}_1^n$. Let $g_y = \max(g_{x_1 y}, \ldots, g_{x_n y})$; then $f < g_y + \varepsilon$ on X and $f > g_y - \varepsilon$ on $V_y = \bigcap_1^n V_{x_j y}$, which is open and contains y. Thus $\{V_y\}_{y \in X}$ is another open cover of X, so there is a finite subcover $\{V_{y_j}\}_1^m$. Let $g = \min(g_{y_1}, \ldots, g_{y_m})$; then $\|f - g\|_u < \varepsilon$. Since $\mathcal{A}$ is a lattice, $g \in \mathcal{A}$, and since $\mathcal{A}$ is closed and ε is arbitrary, $f \in \mathcal{A}$. □

Proof of 151. Given $x \neq y \in X$, let $\mathcal{A}_{xy} = \{(f(x), f(y)) \mid f \in \mathcal{A}\}$. Then $\mathcal{A}_{xy}$ is a subalgebra of $\mathbb{R}^2$ as in Lemma 157 because $f \mapsto (f(x), f(y))$ is an algebra homomorphism. If $\mathcal{A}_{xy} = \mathbb{R}^2$ for all x, y, then Lemmas 159 and 160 imply that $\mathcal{A} = C(X, \mathbb{R})$. Otherwise, there exist x, y for which $\mathcal{A}_{xy}$ is a proper subalgebra of $\mathbb{R}^2$. It cannot be $\{(0, 0)\}$ or the linear span of $(1, 1)$ because $\mathcal{A}$ separates points, so by Lemma 157 $\mathcal{A}_{xy}$ is the linear span of $(1, 0)$ or $(0, 1)$. In either case there exists $x_0 \in X$ such that $f(x_0) = 0$ for all $f \in \mathcal{A}$. There is only one such x_0 since $\mathcal{A}$ separates points, so if neither x nor y is x_0, we have $\mathcal{A}_{xy} = \mathbb{R}^2$. Lemmas 159 and 160 now imply that $\mathcal{A} = \{f \in C(X, \mathbb{R}) \mid f(x_0) = 0\}$. Finally, if $\mathcal{A}$ contains constant functions, there is no x_0 such that $f(x_0) = 0$ for all $f \in \mathcal{A}$, so $\mathcal{A}$ must equal $C(X, \mathbb{R})$. □

We have stated the Stone-Weierstrass theorem in the form that is most natural for the proof. However, in applications one is typically dealing with a subalgebra $\mathcal{B}$ of $C(X, \mathbb{R})$ that is not closed, and one applies the theorem to $\mathcal{B} = \overline{\mathcal{B}}$. The resulting restatement of the theorem is as follows:

Corollary 4.161: 4.50.

Suppose $\mathcal{B}$ is a subalgebra of $C(X, \mathbb{R})$ that separates points. If there exists $x_0 \in X$ such that $f(x_0) = 0$ for all $f \in \mathcal{B}$, then $\mathcal{B}$ is dense in $\{f \in C(X, \mathbb{R}) \mid f(x_0) = 0\}$. Otherwise, $\mathcal{B}$ is dense in $C(X, \mathbb{R})$.

The classical Weierstrass approximation theorem is the special case of this corollary where X is a compact subset of $\mathbb{R}^n$ and $\mathcal{B}$ is the algebra of polynomials on $\mathbb{R}^n$ (restricted to X); here $\mathcal{B}$ contains the constant functions, so the conclusion is that it is dense in $C(X, \mathbb{R})$.

Warning 4.162.

The Stone-Weierstrass theorem, as it stands, is false for complex-valued functions, as the following example shows.

Example 163. *The algebra of polynomials in one complex variable is not dense in $C(K)$ for most compact subsets K of $\mathbb{C}$. (In particular, if $K^\circ \neq \varnothing$, any uniform limit of polynomials on K must be holomorphic on K°.) Here we shall give a simple proof that the function $f(z) = \overline{z}$ cannot be approximated uniformly by polynomials on the unit circle $\{e^{it} \mid t \in [0, 2\pi]\}$. If $P(z) = \sum_0^n a_j z^j$, then*

$$\int_0^{2\pi} \overline{f}(e^{it})P(e^{it})dt = \sum\nolimits_0^n a_j \int_0^{2\pi} e^{i(j+1)t}dt = 0.$$

Thus, abbreviating $f(e^{it})$ and $P(e^{it})$ by f and P, since $|f| = 1$ on the unit circle we have

$$\begin{aligned} 2\pi = \left|\int_0^{2\pi} f\overline{f}dt\right| &\leqslant \left|\int_0^{2\pi} (f-P)\overline{f}dt\right| + \left|\int_0^{2\pi} \overline{f}Pdt\right| \\ &= \left|\int_0^{2\pi} (f-P)\overline{f}dt\right| \leqslant \int_0^{2\pi} |f-P|dt \leqslant 2\pi\|f-P\|_u. \end{aligned}$$

Therefore, $\|f - P\|_u \geqslant 1$ for any polynomial P.

There is, however, a complex version of the Stone-Weierstrass theorem.

Theorem 4.164: 4.51: The Complex Stone-Weierstrass Theorem.

Let X be a compact Hausdorff space. If $\mathcal{A}$ is a closed subalgebra of $C(X)$ that separates points and is closed under complex conjugation, then either $\mathcal{A} = C(X)$ or $\mathcal{A} = \{f \in C(X) \mid f(x_0) = 0\}$ for some $x_0 \in X$.

Proof. Since $\operatorname{Re} f = (f + \overline{f})/2$ and $\operatorname{Im} f = (f - \overline{f})/2i$, the set $\mathcal{A}_{\mathbb{R}}$ of real and imaginary parts of functions in $\mathcal{A}$ is a subalgebra of $C(X, \mathbb{R})$ to which the StoneWeierstrass theorem applies. Since $\mathcal{A} = \{f + ig \mid f, g \in \mathcal{A}_{\mathbb{R}}\}$, the desired result follows. □

There is also a version of the Stone-Weierstrass theorem for noncompact LCH spaces. We state this result for real functions; the corresponding analogue of Theorem 164 for complex functions is an immediate consequence.

Theorem 4.165: 4.52.

Let X be a noncompact LCH space. If $\mathcal{A}$ is a closed subalgebra of $C_0(X, \mathbb{R}) (= C_0(X) \cap C(X, \mathbb{R}))$ that separates points, then either $\mathcal{A} = C_0(X, \mathbb{R})$ or $\mathcal{A} = \{f \in C_0(X, \mathbb{R}) \mid f(x_0) = 0\}$ for some $x_0 \in X$.

The proof is outlined in Folland Exercise 4.67.

Exercise 4.166.

Let $1 - \sum_1^\infty c_n t^n$ be the Maclaurin series for $(1-t)^{1/2}$.

(a) The series converges absolutely and uniformly on compact subsets of $(-1,1)$, as does the termwise differentiated series $-\sum_1^\infty nc_n t^{n-1}$. Thus, if $f(t) = 1 - \sum_1^\infty c_n t^n$, then $f'(t) = -\sum_1^\infty nc_n t^{n-1}$.

(b) By explicit calculation, $f(t) = -2(1-t)f'(t)$, from which it follows that $(1-t)^{-1/2}f(t)$ is constant. Since $f(0) = 1$, $f(t) = (1-t)^{1/2}$.

Exercise 4.167: Folland Exercise 4.67.

Prove Theorem 165. (If there exists $x_0 \in X$ such that $f(x_0) = 0$ for all $f \in \mathcal{A}$, let Y be the one-point compactification of $X \setminus \{x_0\}$; otherwise let Y be the one-point compactification of X. Apply Proposition 117 and the Stone-Weierstrass theorem on Y.)

Exercise 4.168: Folland Exercise 4.68.

Let X and Y be compact Hausdorff spaces. The algebra generated by functions of the form $f(x,y) = g(x)h(y)$, where $g \in C(X)$ and $h \in C(Y)$, is dense in $C(X \times Y)$.

Solution. Let X, Y be CH spaces and let $\mathcal{A}$ be the given algebra. We want to apply Stone-Weierstrass to $\overline{\mathcal{A}}$.

- $\mathcal{A}$ (hence $\overline{\mathcal{A}}$) contains constant functions, as we can set $g(x) \coloneqq z$, $h(y) \coloneqq 1$ for any $z \in \mathbb{C}$ to get $f(x,y) = z$.
- $\mathcal{A}$ (hence $\overline{\mathcal{A}}$) separates points: if $(x,y) \neq (x',y')$ then without loss of generality $x \neq x'$, so because CH spaces are normal we can apply Urysohn's lemma to get a continuous $g \in C(X)$ such that $g(x) \neq g(x')$. Then set $f(x,y) = g(x) \cdot 1$ (so $h(y)$ here is the constant function 1), in which case
$$f(x,y) = g(x) \neq g(x') = f(x',y'),$$
so $\mathcal{A}$ separates points.
- $\mathcal{A}$ is closed under complex conjugation: Because $(C(X \times Y), \|-\|_u)$ is a metric space, any $f \in \overline{\mathcal{A}}$ takes form $f = \lim_{n\to\infty} \sum_{\text{finite}} z_j g_{n,j} h_{n,j}$. Since complex conjugation is continuous,
$$\overline{f} = \lim_{n\to\infty} \overline{\sum\nolimits_{\text{finite}} z_j g_{n,j} h_{n,j}} = \lim_{n\to\infty} \sum\nolimits_{\text{finite}} \overline{z_j g_{n,j}}\overline{h_{n,j}} \in \overline{\mathcal{A}},$$
hence $\overline{\mathcal{A}}$ is closed under complex conjugation.

Therefore, by Stone-Weierstrass, $\overline{\mathcal{A}} = C(X \times Y)$, so $\mathcal{A}$ is dense in $C(X \times Y)$. □

Exercise 4.169: Folland Exercise 4.69.

Let A be a nonempty set, and let $X = [0,1]^A$. The algebra generated by the coordinate maps $\pi_\alpha \colon X \to [0,1]$ $(\alpha \in A)$ and the constant function 1 is dense in $C(X)$.

Solution. Let $\mathcal{A}$ be the algebra generated by the coordinate maps $\pi_\alpha \colon X \to [0,1]$ $(\alpha \in A)$ and the constant function 1. Note X is compact by Tychonoff (and Hausdorff as a product of Hausdorff spaces), so we aim to apply Stone-Weierstrass to $\overline{\mathcal{A}}$.

- $\overline{\mathcal{A}}$ contains constant functions because $z \cdot 1 \in \mathcal{A}$ for all $z \in \mathbb{C}$.
- $\overline{\mathcal{A}}$ separates points: If $x := \{x_\alpha\}_\alpha$ and $y = \{y_\alpha\}_\alpha$ are distinct, then $x_{\alpha_0} \neq y_{\alpha_0}$ for some α_0. Hence $\pi_{\alpha_0}(x) = x_{\alpha_0} = y_{\alpha_0} = \pi_{\alpha_0}(y)$, so since $\pi_{\alpha_0} \in \mathcal{A} \subset \overline{\mathcal{A}}$ we know $\overline{\mathcal{A}}$ separates points.
- $\overline{\mathcal{A}}$ is closed under complex conjugation: Because $(C(X), \|-\|_u)$ is a metric space, the general form of an element $f \in \mathcal{A}$ is a limit of some $\{f_n\}_{n=1}[^\infty] \subset \mathcal{A}$, and each f_n finite linear combination of the form $\sum_{j=1}^{k_n} z_{n,j} f_{n,j}$ where each $f_{n,j}$ is in the generating set (the map 1 together with the coordinate maps π_α), so $f = \lim_{n\to\infty} \sum_{j=1}^{k_n} z_{n,j} f_{n,j}$. Thus
$$\overline{f} = \overline{\lim_{n\to\infty} \sum_{j=1}^{k_n} z_{n,j} f_{n,j}} = \lim_{n\to\infty} \sum_{j=1}^{k_n} \overbrace{\underbrace{\overline{z_{n,j}}}_{\in\mathcal{A}}\,\underbrace{\overline{f_{n,j}}}_{\in\mathcal{A}}}^{\in\mathcal{A}},$$
so $\overline{f} \in \overline{\mathcal{A}}$.

Hence by Stone-Weierstrass, $\overline{\mathcal{A}} = C(X)$, hence $\mathcal{A}$ is dense in $C(X)$. □

Exercise 4.170: Folland Exercise 4.70.

Let X be a compact Hausdorff space. An ideal in $C(X,\mathbb{R})$ is a subalgebra $\mathcal{J}$ of $C(X,\mathbb{R})$ such that if $f \in \mathcal{J}$ and $g \in C(X,\mathbb{R})$ then $fg \in \mathcal{J}$.

(a) If $\mathcal{J}$ is an ideal in $C(X,\mathbb{R})$, let $h(\mathcal{J}) = \{x \in X \mid f(x) = 0 \text{ for all } f \in \mathcal{J}\}$. Then $h(\mathcal{J})$ is a closed subset of X, called the hull of $\mathcal{J}$.
(b) If $E \subset X$, let $k(E) = \{f \in C(X,\mathbb{R}) \mid f(x) = 0 \text{ for all } x \in E\}$. Then $k(E)$ is a closed ideal in $C(X,\mathbb{R})$, called the kernel of E.
(c) If $E \subset X$, then $h(k(E)) = \overline{E}$.
(d) If $\mathcal{J}$ is an ideal in $C(X,\mathbb{R})$, then $k(h(\mathcal{J})) = \overline{\mathcal{J}}$. [a]
(e) The closed subsets of X are in one-to-one correspondence with the closed ideals of $C(X,\mathbb{R})$.

[a] *Hint: $k(h(\mathcal{J}))$ may be identified with a subalgebra of $C_0(U,\mathbb{R})$ where $U = X \setminus h(\mathcal{J})$.*

Exercise 4.171: Folland Exercise 4.71.

(This is a variation on the theme of Folland Exercise 4.70; it does not use the Stone-Weierstrass theorem.) Let X be a compact Hausdorff space, and let M be the set of all nonzero algebra homomorphisms from $C(X, \mathbb{R})$ to $\mathbb{R}$. Each $x \in X$ defines an element $\widehat{x}$ of M by $\widehat{x}(f) = f(x)$.

(a) If $\phi \in M$, then $\{f \in C(X, \mathbb{R}) \mid \phi(f) = 0\}$ is a maximal proper ideal in $C(X, \mathbb{R})$.
(b) If $\mathfrak{J}$ is a proper ideal in $C(X, \mathbb{R})$, there exists $x_0 \in X$ such that $f(x_0) = 0$ for all $f \in \mathfrak{J}$. (Suppose not; construct an $f \in \mathfrak{J}$ with $f > 0$ everywhere and conclude that $1 \in \mathfrak{J}$. This requires no deep theorems.)
(c) The map $x \to \widehat{x}$ is a bijection from X to M.
(d) If M is equipped with the topology of pointwise convergence, then the map $x \to \widehat{x}$ is a homeomorphism from X to M. (Since M is defined purely algebraically, it follows that the topological structure of X is completely determined by the algebraic structure of $C(X, \mathbb{R})$.)

5 Elements of Functional Analysis

Functional analysis is the traditional name for the study of infinite-dimensional vector spaces over $\mathbb{R}$ or $\mathbb{R}$ and the linear maps between them. What distinguishes this from mere linear algebra is the importance of topological considerations. On finite-dimensional vector spaces there is only one reasonable topology, and linear maps are automatically continuous, but in infinite dimensions things are not so simple. (As we have already observed, if $\{f_n\}$ is a sequence of functions on $\mathbb{R}$, there are many things one can mean by the statement "$f_n \to f$.") As our aim in this chapter is only to give a brief introduction to the subject, we shall restrict attention—except in §5.4—to topologies defined by norms on vector spaces.

5.1 Normed Vector Spaces

Let K denote either $\mathbb{R}$ or $\mathbb{C}$, and let X be a vector space over K.

Notation 1. *We denote the zero element of X simply by 0, relying on context to distinguish it from the scalar $0 \in K$. In this section we will always write* **subspace** *to mean a vector subspace. If $x \in X$, we denote by Kx the one-dimensional subspace spanned by x. Also, if $\mathcal{M}$ and $\mathcal{N}$ are subspaces of x, $\mathcal{M} + \mathcal{N}$ denotes the subspace $\{x + y \mid x \in \mathcal{M}, y \in \mathcal{N}\}$ of X.*

Definition 2. *A* ***seminorm*** *on X is a function $x \mapsto \|x\|$ from X to $[0, \infty)$ such that*

(i) $\|x + y\| \leqslant \|x\| + \|y\|$ for all $x, y \in X$ (the triangle inequality),
(ii) $\|\lambda x\| = |\lambda| \|x\|$ for all $x \in X$ and $\lambda \in K$.

The second property implies that $\|0\| = 0$. A seminorm such that $\|x\| = 0$ only when $x = 0$ is called a ***norm****, and a vector space equipped with a norm is called a* ***normed vector space*** *(or* ***normed linear space****).*

Example 3. *If X is a normed vector space, the function $\rho(x, y) = \|x - y\|$ is a metric on X, since*

$$\|x - z\| \leqslant \|x - y\| + \|y - z\|, \quad \|x - y\| = \|(-1)(x - y)\| = \|y - x\|.$$

The topology it induces is called the ***norm topology*** *on X.*

Definition 4. *Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on X are called* ***equivalent*** *if there exist $C_1, C_2 > 0$ such that for all $x \in X$,*

$$C_1\|x\|_1 \leqslant \|x\|_2 \leqslant C_2\|x\|_1.$$

Equivalent norms define equivalent metrics and hence the same topology and the same Cauchy sequences.

Definition 5. *A normed vector space that is complete with respect to the norm metric is called a* ***Banach space****.*

Remark 6. *Every normed vector space can be embedded in a Banach space as a dense subspace. One way to do this is to mimic the construction of $\mathbb{R}$ from $\mathbb{R}$ via Cauchy sequences; a simpler way is presented in Folland Section 5.2.*

Definition 7. *If $\{x_n\}$ is a sequence in x, the series $\sum_1^\infty x_n$ is said to* ***converge*** *to x if $\sum_1^N x_n \to x$ as $N \to \infty$, and it is called* ***absolutely convergent*** *if $\sum_1^\infty \|x_n\| < \infty$.*

The following is a useful criterion for completeness of a normed vector space.

Theorem 5.8: 5.1.

A normed vector space X is complete if and only if every absolutely convergent series in X converges.

Proof. If X is complete and $\sum_1^\infty \|x_n\| < \infty$, let $S_N = \sum_1^N x_n$. Then for $N > M$ we have

$$\|S_N - S_M\| \leqslant \sum_{M+1}^N \|x_n\| \to 0 \text{ as } M, N \to \infty,$$

so the sequence $\{S_N\}$ is Cauchy and hence convergent. Conversely, suppose that every absolutely convergent series converges, and let $\{x_n\}$ be a Cauchy sequence. We can choose $n_1 < n_2 < \cdots$ such that $\|x_n - x_m\| < 2^{-j}$ for $m, n \geqslant n_j$. Let $y_1 = x_{n_1}$ and $y_j = x_{n_j} - x_{n_{j-1}}$ for $j > 1$. Then $\sum_1^k y_j = x_{n_k}$, and

$$\sum_1^\infty \|y_j\| \leqslant \|y_1\| + \sum_1^\infty 2^{-j} = \|y_1\| + 1 < \infty$$

so $\lim x_{n_k} = \sum_1^\infty y_j$ exists. But since $\{x_n\}$ is Cauchy, it is easily verified that $\{x_n\}$ converges to the same limit as $\{x_{n_k}\}$. □

Example 9. *We have already seen some examples of Banach spaces:*

(1) First, if X is a topological space, $B(X)$ and $BC(X)$ are Banach spaces with the uniform norm $\|f\|_u = \sup_{x \in X} |f(x)|$.

(2) Second, if $(X, \mathcal{M}, \mu)$ is a measure space, $L^1(\mu)$ is a Banach space with the L^1 norm $\|f\|_1 = \int |f| d\mu$. (Observe that $\|\cdot\|_1$ is only a seminorm if we think of $L^1(\mu)$ as consisting of individual functions, but it becomes a norm if we identify functions that are equal a.e.) That $L^1(\mu)$ is complete follows from the MCT for series (Theorem 48) and Theorem 8. Indeed, if $\sum_1^\infty \|f_n\|_1 < \infty$, MCT for series shows that $f = \sum_1^\infty f_n$ exists a.e., and

$$\int \left| f - \sum_1^N f_n \right| d\mu \leqslant \sum_{N+1}^\infty \int |f_n| d\mu \to 0 \text{ as } N \to \infty.$$

More examples will be found in Folland Exercise 5.8,Folland Exercise 5.9,Folland Exercise 5.10,Folland Exercise 5.11 and in subsequent sections.

Example 10. *If X and Y are normed vector spaces, $x \times y$ becomes a normed vector space when equipped with the product norm*

$$\|(x, y)\| = \max(\|x\|, \|y\|).$$

(Here, of course, $\|x\|$ refers to the norm on x while $\|y\|$ refers to the norm on y.) Sometimes other norms equivalent to this one, such as $\|(x, y)\| = \|x\| + \|y\|$ or $\|(x, y)\| = (\|x\|^2 + \|y\|^2)^{1/2}$, are used instead.

Definition 11. *If $\mathcal{M}$ is a vector subspace of the vector space X, it defines an equivalence relation on X as*

$$x \sim y \iff x - y \in \mathcal{M}.$$

*The equivalence class of $x \in X$ is denoted by $X + \mathcal{M}$, and the set of equivalence classes, or **quotient space**, is denoted by $X/\mathcal{M}$. $X/\mathcal{M}$ is a vector space with vector operations $(x + \mathcal{M}) + (y + \mathcal{M}) = (x + y) + \mathcal{M}$ and $\lambda(x + \mathcal{M}) = (\lambda x) + \mathcal{M}$. If X is a normed vector space and $\mathcal{M}$ is closed, $X/\mathcal{M}$ inherits a norm from X called the **quotient norm**, namely*

$$\|x + \mathcal{M}\| := \inf_{y \in \mathcal{M}} \|x + y\|$$

See Folland Exercise 5.12 for a more detailed discussion.

Definition 12. *A linear map $T \colon X \to Y$ between two normed vector spaces is called **bounded** if there exists $C \geqslant 0$ such that*

$$\|Tx\| \leqslant C\|x\| \text{ for all } x \in X.$$

Warning 5.13.

This is different from the notion of boundedness for functions on a set, according to which T would be bounded if $\|Tx\| \leqslant C$ for all x. Clearly no nonzero linear map can satisfy the latter condition, since $T(\lambda x) = \lambda Tx$ for all scalars λ. The present definition means that T is bounded on bounded subsets of X.

Proposition 5.14: 5.2.

If X and Y are normed vector spaces and $T\colon X \to Y$ is a linear map, the following are equivalent:

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

Proof. That (a) implies (b) is trivial. If T is continuous at $0 \in X$, there is a neighborhood U of 0 such that $T(U) \subset \{y \in Y \mid \|y\| \leqslant 1\}$, and U must contain a ball $B = \{x \in X \mid \|x\| \leqslant \delta\}$ about 0 ; thus $\|Tx\| \leqslant 1$ when $\|x\| \leqslant \delta$. Since T commutes with scalar multiplication, it follows that $\|Tx\| \leqslant a\delta^{-1}$ whenever $\|x\| \leqslant a$, that is, $\|Tx\| \leqslant \delta^{-1}\|x\|$. This shows that (b) implies (c). Finally, if $\|T\overline{x}\| \leqslant C\|x\|$ for all x, then $\|Tx_1 - Tx_2\| = \|T(x_1 - x_2)\| \leqslant \varepsilon$ whenever $\|x_1 - x_2\| \leqslant C^{-1}\varepsilon$, so that T is continuous. □

Notation 15. *If X and Y are normed vector spaces, we denote the space of all bounded linear maps from X to Y by $L(X, Y)$.*

It is easily verified that $L(X, Y)$ is a vector space and that the function $T \mapsto \|T\|$ defined by

$$\begin{aligned} \|T\| &= \sup\{\|Tx\| \mid \|x\| = 1\} \\ &= \sup\left\{ \frac{\|Tx\|}{\|x\|} \,\middle|\, x \neq 0 \right\} \\ &= \inf\{C \mid \|Tx\| \leqslant C\|x\| \text{ for all } x\} \end{aligned} \tag{5.15.1}$$

is a norm on $L(X, Y)$, called the operator norm Folland Exercise 5.2. We always assume $L(X, Y)$ to be equipped with this norm unless we specify otherwise.

Proposition 5.16: 5.4.

If Y is complete, so is $L(X, Y)$.

Proof. Let $\{T_n\}$ be a Cauchy sequence in $L(X, Y)$. If $x \in X$, then $\{T_n x\}$ is Cauchy in Y because $\|T_n x - T_m x\| \leqslant \|T_n - T_m\|\|x\|$. Define $T\colon X \to Y$ by $Tx = \lim T_n x$. We leave it to the reader (Folland Exercise 5.3) to verify that $T \in L(X, Y)$ (in fact, $\|T\| = \lim\|T_n\|$) and that $\|T_n - T\| \to 0$. □

Example 17. *Another useful property of the operator norm is the following. If $T \in L(X, T)$ and $S \in L(Y, Z)$, then for each $x \in X$*

$$\|STx\| \leqslant \|S\|\|Tx\| \leqslant \|S\|\|T\|\|x\|,$$

so that $ST \in L(X, Z)$ and $\|ST\| \leqslant \|S\|\|T\|$. In particular, $L(X, X)$ is an algebra. If X is complete, $L(X, X)$ is in fact a ***Banach algebra****, which is defined as a Banach space that* *is also an algebra, such that the norm of a product is at most the product of the norms. (Another example of a Banach algebra is $BC(X)$, where X is a topological space, with pointwise multiplication and the uniform norm.)*

Definition 18. *If $T \in L(X,Y)$, T is said to be* ***invertible****, or an* ***isomorphism****, if T is bijective and T^{-1} is bounded (in other words, $\|Tx\| \geqslant C\|x\|$ for some $C > 0$). T is called an* ***isometry*** *if $\|Tx\| = \|x\|$ for all $x \in X$.*

Warning 5.19.

An isometry is injective but not necessarily surjective; it is, however, an isomorphism onto its range.

Exercise 5.20: Folland Exercise 5.1.

If X is a normed vector space over $K (= \mathbb{R}$ or $\mathbb{R})$, then addition and scalar multiplication are continuous from $X \times X$ and $K \times X$ to X. Moreover, the norm is continuous from X to $[0, \infty)$; in fact, $|\|x\| - \|y\|| \leqslant \|x - y\|$ for each $x, y \in X$.

Exercise 5.21: Folland Exercise 5.2.

$L(X,Y)$ is a vector space and the function $\|\cdot\|$ defined by Equation (5.15.1) is a norm on it. In particular, the three expressions on the right of Equation (5.15.1) are always equal.

Exercise 5.22: Folland Exercise 5.3.

Complete the proof of Proposition 16.

Exercise 5.23: Folland Exercise 5.4.

If X, Y are normed vector spaces, the map $(T, x) \mapsto Tx$ is continuous from $L(X,Y) \times X$ to Y. (That is, if $T_n \to T$ and $x_n \to x$ then $T_n x_n \to Tx$.)

Exercise 5.24: Folland Exercise 5.5.

If X is a normed vector space, the closure of any subspace of X is a subspace.

Exercise 5.25: Folland Exercise 5.6.

Suppose that X is a finite-dimensional vector space. Let $e_1, \ldots, e_n$ be a basis for X, and define $\|\sum_1^n a_j e_j\|_1 = \sum_1^n |a_j|$.

(a) $\|\cdot\|_1$ is a norm on X.

(b) The map $(a_1, \dots, a_n) \mapsto \sum_1^n a_j e_j$ is continuous from K^n with the usual Euclidean topology to X with the topology defined by $\|\cdot\|_1$.
(c) $\{x \in X \mid \|x\|_1 = 1\}$ is compact in the topology defined by $\|\cdot\|_1$.
(d) All norms on X are equivalent. (Compare any norm to $\|\cdot\|_1$.)

Solution. Let $x, y \in X$.

(1) We have $\|x\|_1 = 0 \Leftrightarrow$ each $|x_i| = 0 \Leftrightarrow$ each $x_i = 0$. For $\lambda \in K$, we have that $\|\lambda x\|_1 = \sum |\lambda x_i| e_i = \sum |\lambda||x_i| e_i = |\lambda| \sum |x_i| e_i = |\lambda| \|x\|_1$; finally, $\|x + y\|_1 = \sum |x_i + y_i| e_i \leqslant \sum |x_i| e_i + \sum |y_i| e_i = \|x\|_1 + \|y\|_1$.

(2) Let $\delta = \varepsilon/n$. Then $\max_{1 \leqslant i \leqslant n} |a_j - b_j| \leqslant \|(a_1 - b_1, \dots, a_n - b_n)\| < \varepsilon/n$, so $\sum_1^n |a_j - b_j| < \sum_1^n \varepsilon/n = \varepsilon$.

(3) $(X, \|\cdot\|_1)$ is a normed space, so since we've shown for normed spaces that a closed and bounded subset is compact if and only if the space is finite-dimensional, it suffices to show that $U_b \coloneqq oloneq\{x \in X \mid \|x\|_1 = 1\}$ is closed and bounded in X; U_b is obviously bounded, so it suffices to show that it is closed. Let $y \in X$ be in U_b^c. Then we have that $\|y\|_1 \neq 1$, so there's some $\varepsilon > 0$ for which $\|y - y'\|_1 > \varepsilon$ for each $y' \in U_b$. Therefore, there's an open neighborhood of y also contained in U_b^c (since any $x \in X$ has $B_\varepsilon(y) \to \|y - x\|_1 < \varepsilon \to x \notin U_b \to x \in U_b^c$). It follows that U_b^c is open, so its complement U_b is closed in X, which gives the result per our initial remarks.

(4) Any constant makes the claimed inequality work for $x = 0$, so we will only work with nonzero $x \in X$ henceforth. Let $\|\cdot\|_2 \colon X \to [0, \infty)$ be an arbitrary norm on X. Notice that $\|x\|_2 = \|\sum_1^n x_j e_j\| \leqslant \sum_1^n \|x_j e_j\| \leqslant \sum_1^n |x_j| \|e_j\| \leqslant C_2 \|x\|_1$, where $C_2 \coloneqq \max_{1 \leqslant i \leqslant n}\{\|e_i\|\}$. We now need C_1 for which $C_1 \|x\|_1 \leqslant \|x\|_2$ for all $x \in X$. Any arbitrary norm $\|\cdot\|_2$ on X is continuous on X. Indeed, this is clear from settings $\delta = \varepsilon/C_2$ and following the logical progression in (b) to establish continuity of $\|\cdot\|_1$, which works because we have established that $\|\cdot\|_2$ since $\|\cdot\|_2 \leqslant C_2 \|\cdot\|_2$. We now define $F \colon X \to [0, \infty)$ by $F(x) = \|x\|_2$, which is continuous by the above argument. We then observe that $F|_{U_b}$ is a continuous function on a compact set, so since U_b is compact by part (c), we have that $F|_{U_b}$ has and achieves its extrema. It follows that there exists some point $q \in U_b$ for which $\|q\|_2 \leqslant \|u\|_2$ for any $u \in U_b$. Now, fix some arbitrary norm $\|\cdot\|_n$ and nonzero $x \in X$. Here we will argue that $\|(x/\|x\|_n)\|_n = 1$. This warrants justification: observe that $\|(x/\|x\|_n)\|_n = |(1/\|x\|_n)| \|x\|_n = (1/\|x\|_n) \|x\|_n = 1$, as $\|x\|_n$ is positive (since $x \neq 0 \Leftrightarrow \|x\|_n > 0$) and so its reciprocal is positive, warranting the penultimate equality here. We can therefore conclude that for any $x \in X$, $\|(x/\|x\|_1)\|_1 = 1$, and thus $x/\|x\|_1 \in U_b$. We now put everything together. For all $u \in U_b$ we now have two things: (i) $\|q\|_1 = \|u\|_1 = 1$, and (ii) $\|q\|_2 \leqslant \|u\|_2$. It follows from (ii) that $\|u\|_1 \|q\|_2 \leqslant (1)\|u\|_2 = \|u\|_2$, so since for each $x \in X$ we have by the above argument that $x/\|x\|_1 \in U_b$, we conclude that $\|(x/\|x\|_1)\|_1 \|q\|_2 \leqslant \|(x/\|x\|_1)\|_2$. But we can multiply both sides by $\|x\|_1 > 0$, giving the result $\|x\|_1 C_1 \leqslant \|x\|_2$, where $C_1 \coloneqq oloneq\|q\|_2$.

This completes the proof. ☐

Exercise 5.26: Folland Exercise 5.7.

Let X be a Banach space.

(a) If $T \in L(X,X)$ and $\|I - T\| < 1$ where I is the identity operator, then T is invertible; in fact, the series $\sum_0^\infty (I-T)^n$ converges in $L(X,X)$ to T^{-1}.

(b) If $T \in L(X,X)$ is invertible and $\|S - T\| < \|T^{-1}\|^{-1}$, then S is invertible. Thus the set of invertible operators is open in $L(X,X)$.

Solution.

(1) X is a Banach space, so $L(X,X)$ is also a Banach space. $\sum_{k=0}^\infty (I-T)^k$ is absolutely convergent because

$$\sum_{k=0}^{\infty} \|(I-T)^k\| \leqslant \sum_{k=0}^{\infty} \|I-T\|^k$$

is a geometric series with ratio $\|I-T\|_{\text{op}} < 1$ as given. Therefore $\sum_{k=0}^\infty (I-T)^k$ converges to some $S \in L(X,X)$. Fix $\varepsilon > 0$. Then for sufficiently large n we have $\|S - \sum_{k=0}^n (I-T)^k\| < \varepsilon$, and

$$\begin{aligned}
\|S - I - S(I-T)\| &= \|S - \sum_{k=0}^{N+1}(I-T)^k + \sum_{k=1}^{N+1}(I-T)^k - S(I-T)\| \\
&\leqslant \|S - \sum_{k=0}^{N+1}(I-T)^k\| + \|\sum_{k=1}^{N+1}(I-T)^{k-1}(I-T) - S(I-T)\| \\
&< \varepsilon + \|(\sum_{k=0}^{N}(I-T)^k - S)(I-T)\|_{\text{op}} \\
&\leqslant \varepsilon + \varepsilon\|I-T\|_{\text{op}} < 2\varepsilon
\end{aligned}$$

It follows that $\|S - I - S(I-T)\| = 0$, so $S - I = S(I-T) = S - ST$ and hence $ST = I$. Similarly, $\|S - I - (I-T)S\| = 0$, so $TS = I$, giving $T^{-1} = S$, S as above, as claimed.

(2) We have

$$\begin{aligned}
\|I - T^{-1}S\| = \|T^{-1}S - I\| &= \|T^{-1}S - T^{-1}T\| \\
&\leqslant \|T^{-1}\|\|S-T\| < \|T^{-1}\|\|T^{-1}\|^{-1} = 1,
\end{aligned}$$

so it follows from part (a) that $T^{-1}S$ is invertible, and in particular that S is invertible, and tracing back we find where W is the inverse of $T^{-1}S$ that $S^{-1} = WT^{-1}$. Hence operators in $B_{1/\|T^{-1}\|}(T)$ are also invertible for all $T \in L(X,X)$. Thus the set of invertible $T \in L(X,X)$ is open.

☐

Exercise 5.27: Folland Exercise 5.8.

Let $(X, \mathcal{M})$ be a measurable space, and let $M(X)$ be the space of complex measures on $(X, \mathcal{M})$. Then $\|\mu\| = |\mu|(X)$ is a norm on $M(X)$ that makes $M(X)$ into a Banach space. (Use Theorem 8.)

Exercise 5.28: Folland Exercise 5.9.

Let $C^k([0,1])$ be the space of functions on $[0,1]$ possessing continuous derivatives up to order k on $[0,1]$, including one-sided derivatives at the endpoints.

(a) If $f \in C([0,1])$, then $f \in C^k([0,1])$ if and only if f is k times continuously differentiable on $(0,1)$ and $\lim_{x \searrow 0} f^{(j)}(x)$ and $\lim_{x \nearrow 1} f^{(j)}(x)$ exist for $j \leqslant k$. (The mean value theorem is useful.)

(b) $\|f\| = \sum_0^k \|f^{(j)}\|_u$ is a norm on $C^k([0,1])$ that makes $C^k([0,1])$ into a Banach space. (Use induction on k. The essential point is that if $\{f_n\} \subset C^1([0,1]), f_n \to f$ uniformly, and $f_n' \to g$ uniformly, then $f \in C^1([0,1])$ and $f' = g$. The easy way to prove this is to show that $f(x) - f(0) = \int_0^x g(t)dt$.)

Exercise 5.29: Folland Exercise 5.10.

Let $L_k^1([0,1])$ be the space of all $f \in C^{k-1}([0,1])$ such that $f^{(k-1)}$ is absolutely continuous on $[0,1]$ (and hence $f^{(k)}$ exists a.e. and is in $L^1([0,1])$). Then $\|f\| = \sum_0^k \int_0^1 |f^{(j)}(x)|dx$ is a norm on $L_k^1([0,1])$ that makes $L_k^1([0,1])$ into a Banach space. (See Folland Exercise 5.9 and its hint.)

Exercise 5.30: Folland Exercise 5.11.

If $0 < \alpha \leqslant 1$, let $\Lambda_\alpha([0,1])$ be the space of Hölder continuous functions of exponent α on $[0,1]$. That is, $f \in \Lambda_\alpha([0,1])$ if and only if $\|f\|_{\Lambda_\alpha} < \infty$, where

$$\|f\|_{\Lambda_\alpha} = |f(0)| + \sup_{x,y \in [0,1], x \neq y} \frac{|f(x) - f(y)|}{|x-y|^\alpha}.$$

(a) $\|\cdot\|_{\Lambda_\alpha}$ is a norm that makes $\Lambda_\alpha([0,1])$ into a Banach space.

(b) Let $\lambda_\alpha([0,1])$ be the set of all $f \in \Lambda_\alpha([0,1])$ such that

$$\frac{|f(x) - f(y)|}{|x-y|^\alpha} \to 0 \text{ as } x \to y, \text{ for all } y \in [0,1].$$

If $\alpha < 1, \lambda_\alpha([0,1])$ is an infinite-dimensional closed subspace of $\Lambda_\alpha([0,1])$. If $\alpha = 1, \lambda_\alpha([0,1])$ contains only constant functions.

Exercise 5.31: Folland Exercise 5.12.

Let X be a normed vector space and $\mathcal{M}$ a proper closed subspace of X.

(a) $\|x + \mathcal{M}\| = \inf\{\|x + y\| \mid y \in \mathcal{M}\}$ is a norm on $X/\mathcal{M}$.

(b) For any $\varepsilon > 0$ there exists $x \in X$ such that $\|x\| = 1$ and $\|x + \mathcal{M}\| \geqslant 1 - \varepsilon$.

(c) The projection map $\pi(x) = x + \mathcal{M}$ from X to $X/\mathcal{M}$ has norm 1.
(d) If x is complete, so is $x/\mathcal{M}$. (Use Theorem 8.)
(e) The topology defined by the quotient norm is the quotient topology as defined in Folland Exercise 4.28.

Exercise 5.32: Folland Exercise 5.13.

If $\|\cdot\|$ is a seminorm on the vector space X, let $\mathcal{M} = \{x \in X \mid \|x\| = 0\}$. Then $\mathcal{M}$ is a subspace, and the map $x + \mathcal{M} \mapsto \|x\|$ is a norm on $X/\mathcal{M}$.

Exercise 5.33: Folland Exercise 5.14.

If X is a normed vector space and $\mathcal{M}$ is a non-closed subspace, then $\|x + \mathcal{M}\|$, as defined in Folland Exercise 5.12, is a seminorm on $X/\mathcal{M}$. If one divides by its nullspace as in Folland Exercise 5.13, the resulting quotient space is isometrically isomorphic to $x/\overline{\mathcal{M}}$. (See Folland Exercise 5.5.)

Exercise 5.34: Folland Exercise 5.15.

Suppose that X and Y are normed vector spaces and $T \in L(X, Y)$. Let $\mathcal{N}(T) = \{x \in X \mid Tx = 0\}$.
(a) $\mathcal{N}(T)$ is a closed subspace of X.
(b) There is a unique $S \in L(X/\mathcal{N}(T), y)$ such that $T = S \circ \pi$ where $\pi \colon X \to X/\mathcal{M}$ is the projection (see Folland Exercise 5.12). Moreover, $\|S\| = \|T\|$.

Exercise 5.35: Folland Exercise 5.16.

The purpose of this exercise is to develop a theory of integration for functions with values in a separable Banach space. The integral we will develop is called the ***Bochner integral***. Let $(X, \mathcal{M}, \mu)$ be a measure space, Y a separable Banach space, and L_Y the space of all $(\mathcal{M}, B_Y)$-measurable maps from X to Y and F_Y the set of maps $f \colon X \to Y$ of the form $f(x) = \sum_1^n \chi_{E_j}(x) y_j$ where $n \in \mathbb{Z}_{\geqslant 1}$, $y_j \in Y$, $E_j \in \mathcal{M}$, and $\mu(E_j) < \infty$. If $f \in L_Y$, since $y \mapsto \|y\|$ is continuous Folland Exercise 5.1, $x \mapsto \|f(x)\|$ is $(\mathcal{M}, \mathcal{B}_{\mathbb{Z}_{\geqslant 1}})$-measurable, and we define $\|f\|_1 = \int \|f(x)\| d\mu(x)$. Finally, let $L_Y^1 = \{f \in L_y \mid \|f\|_1 < \infty\}$.
(a) L_Y is a vector space, F_Y and L_Y^1 are subspaces of it, $F_Y \subset L_Y^1$, and $\|\cdot\|_1$ is a seminorm on L_Y^1 that becomes a norm if we identify two functions that are equal a.e.
(b) Let $\{y_n\}_1^\infty$ be a countable dense set in Y. Given $\varepsilon > 0$, let $B_n^\varepsilon = \{y \in Y \mid \|y - y_n\| < \varepsilon \|y_n\|\}$. Then $\bigcup_1^\infty B_n^\varepsilon \supset Y \setminus \{0\}$.

(c) If $f \in L^1_Y$, there is a sequence $\{h_n\} \subset Fy$ with $h_n \to f$ a.e. and $\|h_n - f\|_1 \to 0$. (With notation as in (b), let $A_{nj} = B_n^{1/j} \setminus \bigcup_{m=1}^{n-1} B_m^{1/j}$ and $E_{nj} = f^{-1}(A_{nj})$, and consider $g_j = \sum_{n=1}^{\infty} y_n \chi_{E_{nj}}$.)
(d) There is a unique linear map $\int : L^1_Y \to Y$ such that $\int y\chi_E = \mu(E)y$ for $y \in Y$ and $E \in \mathcal{M}(\mu(E) < \infty)$, and $\|\int f\| \leqslant \|f\|_1$.
(e) The dominated convergence theorem: If $\{f_n\}$ is a sequence in L^1_Y such that $f_n \to f$ a.e., and there exists $g \in L^1$ such that $\|f_n(x)\| \leqslant g(x)$ for all n and a.e. x, then $\int f_n \to \int f$.
(f) If Z is a separable Banach space, $T \in L(y, Z)$, and $f \in L^1_Y$, then $T \circ f \in L^1_z$ and $\int T \circ f = T(\int f)$.

5.2 Linear Functionals

Let X be a vector space over K, where $K = \mathbb{R}$ or $\mathbb{R}$. A linear map from X to K is called a linear functional on X. If X is a normed vector space, the space $L(X, K)$ of bounded linear functionals on X is called the dual space of X and is denoted by X^*. According to Proposition 16,, X^* is a Banach space with the operator norm.

If X is a vector space over $\mathbb{C}$, it is also a vector space over $\mathbb{C}$, and we can consider both real and complex linear functionals on X, that is, maps $f : X \to \mathbb{C}$ that are linear over $\mathbb{C}$ and maps $f : X \to \mathbb{C}$ that are linear over $\mathbb{C}$. The relationship between the two is as follows:

Proposition 5.36: 5.5.

Let X be a vector space over $\mathbb{C}$. If f is a complex linear functional on X and $u = \operatorname{Re} f$, then u is a real linear functional, and $f(x) = u(x) - iu(ix)$ for all $x \in X$. Conversely, if u is a real linear functional on X and $f : X \to \mathbb{C}$ is defined by $f(x) = u(x) - iu(ix)$, then f is complex linear. In this case, if X is normed, we have $\|u\| = \|f\|$.

Proof. If f is complex linear and $u = \operatorname{Re} f, u$ is clearly real linear and $\operatorname{Im} f(x) = -\operatorname{Re}[if(x)] = -u(ix)$, so $f(x) = u(x) - iu(ix)$. On the other hand, if u is real linear and $f(x) = u(x) - iu(ix)$, then f is clearly linear over $\mathbb{R}$, and $f(ix) = u(ix) - iu(-x) = u(ix) + iu(x) = if(x)$, so f is also linear over $\mathbb{R}$. Finally, if x is normed, since $|u(x)| = |\operatorname{Re} f(x)| \leqslant |f(x)|$ we have $\|u\| \leqslant \|f\|$. On the other hand, if $f(x) \neq 0$, let $\alpha = \operatorname{sgn} f(x)$. Then $|f(x)| = \alpha f(x) = f(\alpha x) = u(\alpha x)$ (since $f(\alpha x)$ is real), so $|f(x)| \leqslant \|u\|\|\alpha x\| = \|u\|\|x\|$, whence $\|f\| \leqslant \|u\|$. □

It is not obvious that there are any nonzero bounded linear functionals on an arbitrary normed vector space. The fact that such functionals exist in great abundance is one of the fundamental theorems of functional analysis. We shall now present this result in a more general form that has other important applications.

Definition 37. *If X is a real vector space, a **sublinear functional** on X is a map $p\colon X \to \mathbb{R}$ such that for all $x, y \in X$ and all $\lambda \geqslant 0$,*

- *$p(x+y) \leqslant p(x) + p(y)$ and*
- *$p(\lambda x) = \lambda p(x)$.*

For example, every seminorm is a sublinear functional.

Theorem 5.38: 5.6: The Hahn-Banach Theorem.

Let X be a real vector space, p a sublinear functional on X, $\mathcal{M}$ a subspace of X, and f a linear functional on $\mathcal{M}$ such that $f(x) \leqslant p(x)$ for all $x \in \mathcal{M}$. Then there exists a linear functional F on X such that $F(x) \leqslant p(x)$ for all $x \in X$ and $F|_{\mathcal{M}} = f$.

Proof. We prove the theorem by induction. Pick $x \in X \setminus M$.

- *Step 1: Extend f from M to a linear functional on $g\colon M \oplus \mathbb{R}x \to \mathbb{R}$*: We have for all $y_1, y_2 \in M$ that
$$f(y_1) + f(y_2) = f(y_1 + y_2) \leqslant p(y_1 + y_2) \leqslant p(y_1 - x) + p(x + y_2).$$
Note that there's some α such that
$$\sup_{y \in M}\{f(y) - p(y-x)\} \leqslant \alpha \leqslant \inf_{y \in M}\{p(x+y) - f(y)\}.$$
Then let $g\colon M \oplus \mathbb{R}x \to \mathbb{R}$ by $g(y + \lambda x) \coloneqq oloneq f(y) + \lambda\alpha$. Clearly g is linear and extends f.
- *Step 2: Show that g preserves the bound*: For any $\lambda > 0$ and $y \in M$ we have $g(y + \lambda x) = \lambda f(y/\lambda) + \lambda(p(x + y/\lambda) - f(y/\lambda))$ (since $\lambda \neq 0$), and then multiply through, cancel, and use positive homogeneity to get that this is $= p(y + \lambda x)$, and hence $p(y + \lambda x) \leqslant g(y + \lambda x)$ for all positive λ.

  Similarly, for each $\lambda < 0$ and $y \in M$ we have $g(y + \lambda x) = |\lambda| f(y/|\lambda|) - |\lambda|(f(y/|\lambda|) - p(x + y/|\lambda|))$, and as we multiply through and cancel to get this is $= p(y + \lambda x)$, so $g(y + \lambda x) \leqslant p(y + \lambda x)$ for negative λ as well. Therefore, for all $y \in M \oplus \mathbb{R}x$, we have $g(y) \leqslant p(y)$.
- *Step 3: Invoke Zorn's lemma to get $F \in X^*$ preserving the bound on all of X*: Let $\mathfrak{F}$ be the family of all linear extensions F of f such that $F \leqslant p$ on the domain of f. Then equip $\mathfrak{F}$ with the partial ordering $\prec$ such that $F_1 \prec F_2$ if and only if f_2 extends F_1. Observe that every linearly ordered subset $\mathfrak{F}_0 \subset \mathfrak{F}$ is bounded above by just taking F^* with domain $\bigcup_{F \in \mathfrak{F}_0}(\text{domain of } F)$, and $F^*(x) \coloneqq oloneq F(x)$ for all x in the domain of F (where $F \in \mathfrak{F}_0$). We are then done by Kuratowski-Zorn. □

If p is a seminorm and $f\colon X \to \mathbb{R}$ is linear, the inequality $f \leqslant p$ is equivalent to the inequality $|f| \leqslant p$, because $|f(x)| = \pm f(x) = f(\pm x)$ and $p(-x) = p(x)$. In this situation the Hahn-Banach theorem also applies to complex linear functionals:

Theorem 5.39: 5.7: The Complex Hahn-Banach Theorem.

Let x be a complex vector space, p a seminorm on $X, \mathcal{M}$ a subspace of X, and f a complex linear functional on $\mathcal{M}$ such that $|f(x)| \leqslant p(x)$ for $x \in \mathcal{M}$. Then there exists a complex linear functional F on x such that $|F(x)| \leqslant p(x)$ for all $x \in X$ and $F|_{\mathcal{M}} = f$.

Proof. Let $u = \operatorname{Re} f$. By **??** there is a real linear extension U of u to X such that $|U(x)| \leqslant p(x)$ for all $x \in X$. Let $F(x) = U(x) - iU(ix)$ as in Proposition 36. Then F is a complex linear extension of f, and as in the proof of Proposition 36, if $\alpha = \overline{\operatorname{sgn} F(x)}$, we have $|F(x)| = \alpha F(x) = F(\alpha x) = U(\alpha x) \leqslant p(\alpha x) = p(x)$. □

Warning 5.40.

From now on until Folland Section 5.5, all of our results apply equally to real or complex vector spaces, but for the sake of definiteness we shall assume that the scalar field is $\mathbb{C}$.

The principal applications of the Hahn-Banach theorem to normed vector spaces are summarized in the following theorem.

Theorem 5.41: 5.8.

Let X be a normed vector space.

(a) If $\mathcal{M}$ is a closed subspace of X and $x \in X \setminus \mathcal{M}$, there exists $f \in X^*$ such that $f(x) \neq 0$ and $f_{\mathcal{M}} = 0$. In fact, if $\delta = \inf_{y \in \mathcal{M}} \|x - y\|, f$ can be taken to satisfy $\|f\| = 1$ and $f(x) = \delta$.

(b) If $x \neq 0 \in X$, there exists $f \in X^*$ such that $\|f\| = 1$ and $f(x) = \|x\|$.

(c) The bounded linear functionals on X separate points.

(d) If $x \in X$, define $\widehat{x} \colon X^* \to \mathbb{C}$ by $\widehat{x}(f) = f(x)$. Then the map $x \mapsto \widehat{x}$ is a linear isometry from X into X^{**} (the dual of X^*).

Proof. To prove (a), define f on $\mathcal{M} + \mathbb{C}x$ by $f(y + \lambda x) = \lambda\delta(y \in \mathcal{M}, \lambda \in \mathbb{C})$. Then $f(x) = \delta, f_{\mathcal{M}} = 0$, and for $\lambda \neq 0, |f(y + \lambda x)| = |\lambda|\delta \leqslant |\lambda|\|\lambda^{-1}y + x\| = \|y + \lambda x\|$. Thus the Hahn-Banach theorem can be applied, with $p(x) = \|x\|$ and $\mathcal{M}$ replaced by $\mathcal{M} + \mathbb{C}x$. (b) is the special case of (a) with $\mathcal{M} = \{0\}$, and (c) follows immediately: if $x \neq y$, there exists $f \in X^*$ with $f(x - y) \neq 0$, i.e., $f(x) \neq f(y)$. As for (d), obviously $\widehat{x}$ is a linear functional on X^* and the map $x \mapsto \widehat{x}$ is linear. Moreover, $|\widehat{x}(f)| = |f(x)| \leqslant \|f\|\|x\|$, so $\|\widehat{x}\| \leqslant \|x\|$. On the other hand, (b) implies that $\|\widehat{x}\| \geqslant \|x\|$. □

With notation as in Theorem 41(d), let $\widehat{X} = \{\widehat{x} \mid x \in X\}$. Since X^{**} is always complete, the closure $\overline{\widehat{x}}$ of $\widehat{x}$ in X^{**} is a Banach space, and the map $x \mapsto \widehat{x}$ embeds x into $\overline{\widehat{X}}$ as a dense subspace. $\overline{\widehat{x}}$ is called the completion of X. In particular, if X is itself a Banach space then $\overline{\widehat{X}} = \widehat{X}$.

If X is finite-dimensional, then of course $\widehat{X} = X^{**}$, since these spaces have the same dimension. For infinite-dimensional Banach spaces it may or may not happen that $\widehat{X} = X^{**}$; if it does, x is called reflexive. The examples of Banach spaces we have examined so far are not reflexive except in trivial cases where they turn out to be finite-dimensional. We shall prove some cases of this assertion and present examples of reflexive Banach spaces in later sections.

Notation 42. *Usually we shall identify $\widehat{X}$ with X and thus regard X^{**} as a superspace of X; reflexivity then means that $X^{**} = X$.*

Exercise 5.43: Folland Exercise 5.17.

A linear functional f on a normed vector space X is bounded if and only if $f^{-1}(\{0\})$ is closed. (Use Folland Exercise 5.12(b).)

Solution.

- $\Rightarrow$ A linear function f on a normed $\mathbb{F}$-vector space is bounded if and only if it is continuous, so f is continuous. Hence, $f^{-1}(\{0\})$ is closed since $\{0\}$ is closed (since the topological space $\mathbb{F}$ is T_1).
- $\Leftarrow$ Conversely, let $f^{-1}(\{0\})$ be closed. $M = \{0\}$ is a closed subspace of X, so by Folland's Theorem 41, we have that for any $x \notin M$ (i.e., $x \neq 0$), there's an $f_x \in X^*$ such that $f(x) \neq 0$ and $f|_M = 0$ (i.e., $f(0) = 0$). In fact, if $\delta = \inf_{y \in M} \|x - y\|$ (i.e., $\delta = \|x\|_X$), then f_x can be taken to satisfy $\|f_x\| = 1$ and $f_x(x) = \delta$. By 12(b), there's an $x \in X$ with unit norm and $\|x + f^{-1}(\{0\})\| \geqslant 1 - \frac{1}{2} = \frac{1}{2}$. Then for any $x \in X/f^{-1}(\{0\})$ and any $y \in X/(f^{-1}(\{0\}) \oplus \mathbb{C}x)$, we have
$$y = \frac{f(y)}{f(x)}x + \left(y - \frac{f(y)}{f(x)}x\right) \in \mathbb{C}x + f^{-1}(\{0\}) = f^{-1}(\{0\}) \oplus \mathbb{C}x,$$
so $X = f^{-1}(\{0\}) \oplus \mathbb{C}x$. But for any $x \in X$, we already know there's $y \in f^{-1}(\{0\})$ with
$$|f(\lambda x+y)| = |\lambda||f(x)| \leqslant 2|\lambda|\|x+f^{-1}(\{0\})\||f(x)| \leqslant 2|\lambda|\|x+y/\lambda\||f(x)| = 2|f(x)|\|\lambda x+y\|,$$
forcing the boundedness of f as desired.

□

Exercise 5.44: Folland Exercise 5.18.

Let X be a normed vector space.

(a) If $\mathcal{M}$ is a closed subspace and $x \in X \setminus \mathcal{M}$ then $\mathcal{M} + \mathbb{C}x$ is closed. (Use Folland Exercise 5.8(a).)

(b) Every finite-dimensional subspace of X is closed.

Exercise 5.45: Folland Exercise 5.19.

Let X be an infinite-dimensional normed vector space.

(a) There is a sequence $\{x_j\}$ in X such that $\|x_j\| = 1$ for all j and $\|x_j - x_k\| \geqslant \frac{1}{2}$ for $j \neq k$. (Construct x_j inductively, using Folland Exercise 5.12(b) and Folland Exercise 5.18.)

(b) X is not locally compact.

Exercise 5.46: Folland Exercise 5.20.

If $\mathcal{M}$ is a finite-dimensional subspace of a normed vector space X, there is a closed subspace $\mathcal{N}$ such that $\mathcal{M} \cap \mathcal{N} = \{0\}$ and $\mathcal{M} + \mathcal{N} = X$.

Exercise 5.47: Folland Exercise 5.21.

If X and Y are normed vector spaces, define $\alpha\colon X^* \times Y^* \to (x \times y)^*$ by $\alpha(f, g)(x, y) = f(x) + g(y)$. Then α is an isomorphism which is isometric if we use the norm $\|(x, y)\| = \max(\|x\|, \|y\|)$ on $x \times y$, the corresponding operator norm on $(x \times y)^*$, and the norm $\|(f, g)\| = \|f\| + \|g\|$ on $X^* \times Y^*$.

Exercise 5.48: Folland Exercise 5.22.

Suppose that X and Y are normed vector spaces and $T \in L(X, Y)$.

(a) Define $T^\dagger\colon Y^* \to X^*$ by $T^\dagger f = f \circ T$. Then $T^\dagger \in L(Y^*, X^*)$ and $\|T^\dagger\| = \|T\|.T^\dagger$ is called the adjoint or transpose of T.

(b) Applying the construction in (a) twice, one obtains $T^{\dagger\dagger} \in L(X^{**}, y^{**})$. If x and y are identified with their natural images $\widehat{x}$ and $\widehat{y}$ in X^{**} and y^{**}, then $T^{\dagger\dagger} \mid x = T$.

(c) $T^\dagger$ is injective if and only if the range of T is dense in y.

(d) If the range of $T^\dagger$ is dense in X^*, then T is injective; the converse is true if X is reflexive.

Exercise 5.49: Folland Exercise 5.23.

Suppose that X is a Banach space. If $\mathcal{M}$ is a closed subspace of X and $\mathcal{N}$ is a closed subspace of X^*, let $\mathcal{M}^0 = \{f \in X^* \mid f_{\mathcal{M}} = 0\}$ and $\mathcal{N}^\perp = \{x \in X \mid f(x) = 0 \text{ for all } f \in \mathcal{N}\}$. (Thus, if we identify x with its image in $X^{**}, \mathcal{N}^\perp = \mathcal{N}^0 \cap X$.)

(a) $\mathcal{M}^0$ and $\mathcal{N}^\perp$ are closed subspaces of X^* and X, respectively.

(b) $(\mathcal{M}^0)^\perp = \mathcal{M}$ and $(\mathcal{N}^\perp)^0 \supset \mathcal{N}$. If X is reflexive, $(\mathcal{N}^\perp)^0 = \mathcal{N}$.

(c) Let $\pi\colon X \to X/\mathcal{M}$ be the natural projection, and define $\alpha\colon (X/\mathcal{M})^* \to X^*$ by $\alpha(f) = f \circ \pi$. Then α is an isometric isomorphism from $(X/\mathcal{M})^*$ onto $\mathcal{M}^0$, where $X/\mathcal{M}$ has the quotient norm.
(d) Define $\beta\colon X^* \to \mathcal{M}^*$ by $\beta(f) = f_{\mathcal{M}}$; then β induces a map $\overline{\beta}\colon X^*/\mathcal{M}^0 \to \mathcal{M}^*$ as in Folland Exercise 5.15, and $\overline{\beta}$ is an isometric isomorphism.

Solution. (a) M^0 and $N^\perp$ are closed subspaces of X^* and X, respectively. M^0 is a subspace since $0|_M = 0$ and if $f|_M, g|_M = 0$ then for any $\alpha \in \mathbb{C}$ $(\alpha f + g)(M) = \alpha f(M) + g(M) = 0$. Now take a Cauchy sequence $\{f_n\} \subset M^0$. Since $\mathbb{C}$ is Banach, so is X^* by Proposition 16. Then $\{f_n\} \to f \in X^*$. Then $f|_M = \lim_{n\to\infty} f_n|_M \equiv 0$ so $f \in M^0$.

$N^\perp$ is a subspace since $0 \in \ker f$ for any $f \in X^*$ and if $x, y \in N^\perp$ then for any $\alpha \in \mathbb{C}, f \in N$, $f(\alpha x + y) = \alpha f(x) + f(y) = 0$ by linearity of f and definition of $N^\perp$. Now take Cauchy $\{x_n\} \subset N^\perp$, $\{x_n\} \to x \in X$ since X is Banach. Now since any $f \in N \subset X^*$ is linear $f(\lim x_n) = \lim f(x_n) \equiv 0$ so $x \in N^\perp$.

(b) First we show $(M^0)^\perp \subseteq M$. Suppose $x \in M^{0\perp}$. Then $x \in \cap_{f\in M^0} \ker(f)$. If $y \in X \setminus M$ then there is some $g \in X^*$ s.t. $g|_M = 0$ and $f(y) = 1$ since M is closed using Theorem 41(a). But then $y \notin \cap_{f\in M^0} \ker(f)$, so $x \in M$. Now we show $M \subseteq (M^0)^\perp$. Suppose $x \in M$. Then $x \in \cap_{f\in M^0} \ker(f) \equiv M^{0\perp}$.

We can see by expanding the definition that clearly $N \subset N^{\perp 0} = \{x \in X | x \in \cap_{f\in N} \ker(f)\}^0 = \{g \in X^* \mid g(x) = 0\ \forall\ x \text{s.t.} x \in \cap_{f\in N} \ker(f)\}$ since for any $f \in N$, $f(x) = 0$ for every x in its kernel. Now suppose X is reflexive and we work to show $N^{\perp 0} \subset N$. Fix $g \in N^\perp$. For the sake of contradiction suppose $g \in X^* \setminus N$. Then since N is closed, by Theorem 41(a) there is some $\hat{x} \in (X^*)^* = X$ (by reflexivity) such that $\hat{x}|_N = 0$ but $\hat{x}(g) \neq 0$. Then by the natural isomorphism between X, X^{**}, $f(x) = 0$ for every $f \in N$ but $g(x) \neq 0$, therefore $g \notin N^{\perp 0}$.

(c) We first check that this defines an isomorphism. Given $g \in \operatorname{im} f$ there is some $f \in (X/M)^*$ s.t. $g = f \circ \pi$ and so $g|_M = 0$ since $\pi|_M = 0$ and so $g \in M^0$. Given $f \in M^0$, $f(x + M) = f(x) + f(M) = f(x)$ by linearity and definition of M^0 so $f \in \operatorname{im} \alpha$ by taking the the map in $(X/M)^*$ agreeing with f on x.

To show isometry, we show that $||\alpha(f)|| = ||f||$ for every $f \in (X/M)^*$. First note that $||\pi|| = 1$:

$$||\pi|| \leqslant \frac{||\pi(x)||}{||x||} = \inf_{y\in M} \frac{||x+y||}{||x||} \leqslant \frac{||x||}{||x||} = 1$$

Where we used the definition of the quotient norm and the triangle inequality. By definition of quotient norm if $z \notin M$ then $||\pi(z)|| = ||z||$ since z is linearly independent from M, thus $||\pi|| = 1$ since 1 is attained in the supremum of the operator norm.

Now $||\alpha(f)|| = ||f \circ \pi|| \leqslant ||\pi||||f|| = ||f||$ by sublinearity of the operator norm. If $f \in (X/M)^*$ then for any $\tilde{x} = x + M \in X$, $x \in X/M$, we have

$$||f(x)|| = ||f(\pi(\tilde{x}))|| \leqslant ||f \circ \pi||\ ||\tilde{x}||$$

by the sublinearity of operator norm. This inequality holds taking the infimum over M in $\tilde{x}$ and so by the definition of the quotient norm and α, $||f(x)|| \leqslant ||\alpha(f)||\,||x||$. Rearranging we have $||f|| = \frac{||f(x)||}{||x||} \leqslant ||\alpha(f)||$ and thus equality. □

Exercise 5.50: Folland Exercise 5.24.

Suppose that X is a Banach space.

(a) Let $\widehat{x}, (X^*)$ be the natural images of x, X^* in X^{**}, x^{***}, and let $\widehat{x}^0 = \{F \in X^{***} \mid F|_{\hat{X}} = 0\}$. Then $(X^*) \cap \widehat{x}^0 = \{0\}$ and $(X^*)^2 + \widehat{x}^0 = x^{***}$.

(b) x is reflexive if and only if X^* is reflexive.

Exercise 5.51: Folland Exercise 5.25.

If X is a Banach space and X^* is separable, then X is separable. (Let $\{f_n\}_1^\infty$ be a countable dense subset of X^*. For each n choose $x_n \in X$ with $\|x_n\| = 1$ and $|f_n(x_n)| \geqslant \frac{1}{2}\|f_n\|$. Then the linear combinations of $\{x_n\}_1^\infty$ are dense in X.) Note: Separability of X does not imply separability of X^*.

Exercise 5.52: Folland Exercise 5.26.

Let X be a real vector space and let P be a subset of x such that (i) if $x, y \in P$, then $x + y \in P$, (ii) if $x \in P$ and $\lambda \geqslant 0$, then $\lambda x \in P$, (iii) if $x \in P$ and $-x \in P$, then $x = 0$. (Example: If X is a space of real-valued functions, P can be the set of nonnegative functions in x.)

(a) The relation $\leqslant$ defined by $x \leqslant y$ if and only if $y - x \in P$ is a partial ordering on X.

(b) (The Klein Extension Theorem) Suppose that $\mathcal{M}$ is a subspace of X such that for each $x \in X$ there exists $y \in \mathcal{M}$ with $x \leqslant y$. If f is a linear functional on $\mathcal{M}$ such that $f(x) \geqslant 0$ for $x \in \mathcal{M} \cap P$, there is a linear functional F on x such that $F(x) \geqslant 0$ for $x \in P$ and $F_{\mathcal{M}} = f$. (Consider $p(x) = \inf\{f(y) \mid y \in \mathcal{M} \text{ and } x \leqslant y\}$.)

5.3 The Baire Category Theorem and its Consequences

In this section we present an important theorem about complete metric spaces and use it to obtain some fundamental results concerning linear maps between Banach spaces.

If X is a topological space, a set $E \subset X$ is of the ***first category***, or ***meager***, if E is a countable union of nowhere dense sets (equivalently, a countable intersection of open dense sets); otherwise E is of the ***second category***. The complement of a meager set is called ***generic*** or ***residual***. It is useful to think of generic sets as corresponding to the situation of a typical set, and to think of meager sets as the exceptional situation.

Note 53.

(1) E is closed and nowhere dense if and only if E^c is open and dense.
(2) The countable union of meager sets is a meager set, and the countable intersection of generic sets is generic.
(3) Any dense open set is generic (for example, interiors) (by (1)).
(4) Thus, the notions of "big" and "small" captured by the Lebesgue measure does not transfer to meager or generic sets: there are meager (exceptional) subsets of $[0,1]$ with Lebesgue measure 1, and in particular uncountable subsets of $[0,1]$, and there exist generic subsets of $[0,1]$ of Lebesgue measure 0, as the following example shows.

Let $\{x_k\}_k$ be an enumeration of the rationals in $[0,1]$, and consider

$$E := \bigcap_{k=1}^{\infty} \underbrace{\bigcup_{k=1}^{\infty} \left(r_k - \frac{1}{2^k n}, r_k + \frac{1}{2^k n} \right)}_{\text{open and dense, hence generic}}.$$

Then E is a countable intersection of generic sense, and hence is generic. In addition, E has Lebesgue measure 0.

The Baire category theorem is often used to prove existence results: One shows that objects having a certain property exist by showing that the set of objects (within a suitable complete metric space) is generic. For example, one can prove the existence of nowhere differentiable continuous functions in this way; see Folland Exercise 5.42.

Theorem 5.54: 5.9: The Baire Category Theorem (BCT).

Let X be a complete metric space.
(a) If $\{U_n\}_1^\infty$ is a sequence of open dense subsets of X, then $\bigcap_1^\infty U_n$ is dense in X.
(b) X is not a countable union of nowhere dense sets, that is, X is of the second category in itself.

Proof. For part (a), we must show that if W is a nonempty open set in X, then W intersects $\bigcap_1^\infty U_n$. Since $U_1 \cap W$ is open and nonempty, it contains a ball $B(r_0, x_0)$, and we can assume that $0 < r_0 < 1$. For $n > 0$, we choose $x_n \in X$ and $r_n \in (0, \infty)$ inductively as follows: Having chosen x_j and r_j for $j < n$, we observe that $U_n \cap B(r_{n-1}, x_{n-1})$ is open and nonempty, so we can choose x_n, r_n so that $0 < r_n < 2^{-n}$ and $B(r_n, x_n) \subset U_n \cap B(r_{n-1}, x_{n-1})$. Then if $n, m \geqslant N$, we see that $x_n, x_m \in B(r_N, x_N)$, and since $r_n \to 0$, the sequence $\{x_n\}$ is Cauchy. As X is complete, $x = \lim x_n$ exists. Since $x_n \in B(r_N, x_N)$ for $n \geqslant N$ we have

$$x \in \overline{B(r_N, x_N)} \subset U_N \cap B(r_1, x_1) \subset U_N \cap W$$

for all N, and the proof is complete. As for (b), if $\{E_n\}$ is a sequence of nowhere dense sets in X, then $\left\{ \left(\overline{E}_n\right)^c \right\}$ is a sequence of open dense sets. Since $\cap \left(\overline{E}_n\right)^c \neq \varnothing$, we have $\bigcup E_n \subset \bigcup \overline{E}_n \neq X$. □

Note 55 (Strengthenings of Baire Category Theorem).

- *We remark that since the conclusions of the Baire category theorem are purely topological, it suffices for X to be homeomorphic to a complete metric space. For example, the theorem applies to $X = (0,1)$, which is not complete with the usual metric but is homeomorphic to $\mathbb{R}$.*
- *The Baire category theorem is also true for LCH spaces. (This is Folland Exercise 5.28.) Note that there exist complete metric spaces that are not LCH spaces and vice versa, so one is not a special case of the other. The proof is almost the same but part (a) has a slight modification as follows: Let B_0 be a nonempty open set in X, and choose nonempty open sets B_n inductively so that $\overline{B_n} \subset U_n \cap B_{n-1}$. If X is LCH, then we can take $\overline{B_n}$ to be compact, and by compactness we have $K = \bigcap \overline{B_n}$ is nonempty. Since $K \subset U_n \cap B_n$ for all n, $(B_0 \cap \bigcap U_n) \neq \varnothing$.*

5.3.1 First Applications of the Baire Category Theorem

Corollary 5.56.

In a complete metric space, generic sets are dense.

Proof. Suppose E is a generic subset that is not dense. Then there exists a closed ball $\overline{B} \subset E^c = \bigcup_{n=1}^\infty F_n$, where each F_n is nowhere dense. But then $\overline{B} = \bigcup_{n=1}^\infty (F_n \cap \overline{B})$ is a countable union of nowhere dense sets, contradicting the Baire category theorem (applied to $\overline{B}$). □

Theorem 5.57.

If X is a complete metric space and $\{f_n \colon X \to \mathbb{C}\}$ a sequence of continuous functions such that $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in X$, then the set E of points in X where f is continuous is a generic set.

Proof. First, we observe that E is a G_δ-set (a countable intersection of open sets). Fix $x \in X$ and define the ***oscillation*** of f at x as

$$\operatorname{osc}_f(x) := \lim_{r \searrow 0} \sup_{y,z\in B_r(x)} |f(y) - f(z)|.$$

This is well-defined, as the limit exists—indeed, $\sup_{y,z\in B_r(x)}|f(y) - f(z)|$ is nonnegative (hence bounded below by 0) and decreasing as r decreases.

Note that $\operatorname{osc}_f(x) = 0$ if and only if $x \in E$. Moreover, for every $\varepsilon > 0$, $\{x \in X \mid \operatorname{osc}_f(x) < \varepsilon\}$ is open. Indeed, if $\operatorname{osc}_f(x) < \varepsilon$, then there exists $r > 0$ such that $|f(y) - f(z)| < \varepsilon$ for all $y, z \in B_r(x)$, so by the triangle inequality $B(x, r/2) \subset \{x \in X \mid \operatorname{osc}_f(x) < \varepsilon\}$. Then $E = \bigcap_{n=1}^\infty \{x \in X \mid \operatorname{osc}_f(x) < 1/n\}$ is a G_δ-set, as claimed.

Now $E^c = \bigcup_{n=1}^{\infty} \underbrace{\{x \in X \mid \operatorname{osc}_f(x) \geqslant 1/n\}}_{=:F_n} = \bigcup_{n=1}^{\infty} F_n$. Note that each F_n is closed (their complements are open). We now show that each F_n has empty interior, so that E^c is meager.

Lemma 5.58.

For every open ball $B \subset X$ and $\varepsilon > 0$, there exists an open ball $B_0 \subset B$ and some $m \in \mathbb{Z}_{\geqslant 1}$ such that $|f(x) - f_m(x)| < \varepsilon$ for all $x \in B$.

Proof. Take a closed ball $Y \subset B$ and let $E_\ell = \{x \in Y \mid \sup_{j,k \geqslant \ell} |f_j(x) - f_k(x)| \leqslant \varepsilon\}$. Then $Y = \bigcup_{\ell=1}^{\infty} E_\ell$, since $\{f_k(x)\}$ converges for every x. Since Y is closed it is a complete metric space, so by the Baire category theorem E_m is *not* nowhere dense. Thus there exists an open ball $B_0 \subset \overline{E_m} = E_m$, where closure is by continuity of the f_ks. Thus $|f_j(x) - f_k(x)| \leqslant \varepsilon$ for all $x \in B$ whenever $j, k \geqslant m$. Letting $k \to \infty$ yields $|f_j(x) - f(x)| \leqslant \varepsilon$ for all $x \in B$, $j \geqslant m$. This proves the lemma. □

Finally, we show each F_n above has empty interior. Suppose that some F_n does *not* have empty interior, and take an open ball $B \subset F_n$. Apply the lemma with $\varepsilon = 1/4n$ to obtain an oen ball $B_0 \subset B$ and an integer $m \geqslant 1$ such that $|f(x) - f_n(x)| < 1/4n$ for all $x \in B_0$. By continuity, there exists a ball $B_0' \subset B_0$ such that $|f_n(y) - f_m(z)| < 1/4n$ for all $y, z \in B_0'$ (since f_m is continuous). Therefore, if $y, z \in B_0'$, then

$$\begin{aligned} |f(y) - f(z)| &\leqslant |f(y) - f_m(y)| + |f_m(y) - f_m(z)| + |f_m(z) - f(z)| \\ &\leqslant \frac{1}{4n} + \frac{1}{4n} + \frac{1}{4n} = \frac{3}{4n} < \frac{1}{n}. \end{aligned}$$

Thus $\operatorname{osc}_f(x') < 1/n$, where x' is the center of B_0'. This means $x' \notin F_n$, a contradiction since $x' \in B_0' \subset B_0 \subset B \subset F_n$ □

Example 59. *Does there exist a function on $\mathbb{R}$ that is*

(a) continuous precisely at the irrationals?

(b) continuous precisely at the rationals?

For (a) the answer is yes, and and an example of such a function is the ***stars over babylon function****, (also called the* ***Thomae function****, or the* ***popcorn function****), which is given by*

$$f(x) = \begin{cases} \frac{1}{q} & \text{if } x = p/q \text{ is rational in lowest terms (with } q > 0) , \\ 0 & \text{otherwise.} \end{cases}$$

For (b) the answer is no: There does not exist a function on $\mathbb{R}$ that is continuous precisely at the rationals, since $\mathbb{Q}$ is not a G_δ-set. Indeed, suppose to the contrary that $\mathbb{Q} = \bigcap_{n=1}^{\infty} U_n$ for open sets U_n. Since each U_n contains $\mathbb{Q}$ (which is dense), each U_n is dense; thus by assumption $\mathbb{Q} = \bigcap_{n=1}^{\infty}$ is an intersection of dense open sets U_n. Then U_n^c is closed and *does not contain any rationals, so U_n^c is nowhere dense. Let $\{x_n\}$ be an enumeration of the rationals. Then $\mathbb{R} = \bigcup_{n=1}^{\infty}(U_n^c \cup \{x_n\})$ is a countable union of nowhere dense sets.*

The Baire category theorem can also be used to show that the typical element of $C([0,1])$ is nowhere differentiable, in the sense that the set of all such functions is generic:

Theorem 5.60: Banach.

The set of nowhere differentiable functions in $C([0,1])$ is generic.

Proof. It suffices to show the set

$$D := \{f \in C([0,1]) \mid f' \text{ exists somewhere}\}$$

is meager. Define

$$C_n := \left\{ f \in C([0,1]) \;\middle|\; \begin{smallmatrix} \text{there exists } x \in [0,1-1/n] \text{ such that} \\ \left|\frac{f(x+h)-f(x)}{h}\right| \leqslant n \text{ for all } h \in (0,1/n) \end{smallmatrix} \right\}.$$

First we show $D \subset \bigcup_2^{\infty} C_n$: Indeed, if this weren't the case, then there exists some $f \in \bigcap_2^{\infty} C_n^c$ such that f' exists somewhere. But then for each $n \geqslant 2$ and all $x \in [0, 1-1/n]$, there is some $h \in (0, 1/n)$ for which the difference quotient is larger than n, meaning no real number could be the limit of the difference quotient as $h \to 0$, i.e. f' doesn't exist. This contradicts the definition of D, so D must be contained in the union $\bigcup_2^{\infty} C_n$.

We now show that C_n is closed for each n. Indeed, fix $n \in \mathbb{N}$ and pick some $\{f_k\}_{k=1}^{\infty} \subset C_n$ and some $f \in C_n$ such that $\lim_{k\to\infty} f_k = f$. with respect to the uniform norm $\|\cdot\|_{\infty}$. Now, because $f_k \subset C_n$ for each k, there's some subsequence of points $\{x_k\}_{k=1}^{\infty} \subset [0,1]$ such that

$$\left|\frac{f_k(x_k+h) - f_k(x_k)}{h}\right| \leqslant n$$

whenever $h \in (0, 1/n)$.

Now, $\{x_k\}_{k=1}^{\infty}$ is a sequence of real numbers that are bounded (they're in the interval $[0,1]$), so there is some convergent subsequence, say $\{x_{k_m}\}_{m=1}^{\infty}$, that converges to, say $x_0 \in [0,1]$. Moreover, note that because$f_k \to f$ in the uniform norm, we have that $f_{k_m} \to f$ in the uniform norm as well.

We now claim that $|f(x_0+h) - f(x_0)| \leqslant hn$ for any $h \in (0,1/n)$: Fix $\varepsilon > 0$ and $h \in (0, 1/n)$. Choose m sufficiently large such that

$$\begin{aligned} &\|f_{k_m} - f\|_{\infty} < \varepsilon h/4, \\ &|f(x_{k_m}) - f(x_0)| < \varepsilon h/4, \\ &|f(x_{k_m}+h) - f(x_0+h)| < \varepsilon h/4. \end{aligned}$$

Then we have that

$$\begin{aligned} |f(x_0+h) - f(x_0)| \leqslant\; &|f(x_0+h) - f(x_{k_m}+h)| + |f(x_{k_m}+h) - f_{k_m}(x_{k_m}+h)| \\ &+ |f_{k_m}(x_{k_m}+h) - f_{k_m}(x_{k_m})| + |f_{k_m}(x_{k_m}) - f(x_{k_m})| \\ &+ |f(x_{k_m}) - f(x_0)| \end{aligned}$$

$$< \varepsilon h/4 + \varepsilon h/4 + \varepsilon h/4 + \varepsilon h/4 + nh$$
$$= h(n+\varepsilon),$$

which goes to nh as $\varepsilon \searrow 0$.

Finally, we claim that C_n has empty interior for each n. Fix $n \in \mathbb{N}$. Suppose for a contradiction it wasn't empty—then there's some $f \in C_n$ and an $\varepsilon > 0$ such that

$$B_\varepsilon^{\|\cdot\|_\infty}(f) \subset C_n.$$

By Stone-Weierstrass, there' a polynomial p with $\|f-p\|_\infty > 0$ such that $B_\delta(p) \subset B)_\varepsilon(f) \subset C_n$ for some $\delta > 0$. Now we construct a continuous function φ so that $\|\varphi\|_\infty < \delta$ and for which $\varphi'_+(x)$, the right-hand derivative of φ at x, exists for each $x \in [0,1)$ and is such that $|\varphi'_+(x)| > n + \|f'\|_\infty$. Then $\varphi + p \in B_\delta(p)$, and for all $x \in [0,1)$ we have $|(\varphi+p)'_+(x)| = |\varphi'_+(x) + p'_+(x)| \geqslant |\varphi'_+(x)| - \|p\|_\infty > n$, which implies $\varphi + p \notin C_n$, contradicting that $\varphi + p \in B_\delta(p) (\subset C_n)$. Hence the interior or C_n must be empty for each n.

From this it follows that $D \subset \bigcup_2^\infty C_n$ is a countable union of nowhere dense sets, so it is nowhere dense by the Baire category theorem, meaning its complement is generic. □

5.3.2 Applications of Baire Category Theorem to Linear Maps

We turn to the applications of the Baire category theorem in the theory of linear maps. Some terminology:

If X and Y are topological spaces, a map $f\colon X \to Y$ is called ***open*** if $f(U)$ is open in Y whenever U is open in X.

If X and Y are metric spaces, amounts to requiring that if B is a ball centered at $x \in X$, then $f(B)$ contains a ball centered at $f(x)$.

If X and Y are in particular normed vector spaces and f is linear, then f commutes with translations and dilations; it follows that f is open if and only if $f(B)$ contains a ball centered at 0 in Y when B is the ball of radius 1 about 0 in X.

Theorem 5.61: 5.10: The Open Mapping Theorem.

If X and Y are Banach spaces, then surjective bounded maps $T \in L(X,Y)$ are open.

Proof. Let B_r denote the (open) ball of radius r about 0 in X. By the preceding remarks, it will suffice to show that $T(B_1)$ contains a ball about 0 in Y. Since $X = \bigcup_1^\infty B_n$ and T is surjective, we have $Y = \bigcup_1^\infty T(B_n)$. But Y is complete and the map $y \mapsto ny$ is a homeomorphism of Y that maps $T(B_1)$ to $T(B_n)$, so by Baire's category theorem $T(B_1)$ cannot be nowhere dense (since complete metric spaces are non-meager—that is, not of the first category—in themselves). That is, there exist $y_0 \in y$ and $r > 0$ such that the ball $B(4r, y_0)$ is contained in $\overline{T(B_1)}$. Pick $y_1 = Tx_1 \in T(B_1)$

$$y = Tx_1 + (y - y_1) \in \overline{T(x_1 + B_1)} \subset \overline{T(B_2)}.$$

Dividing both sides by 2, we conclude that there exists $r > 0$ such that if $\|y\| < r$ then $y \in \overline{T(B_1)}$. If we could replace $\overline{T(B_1)}$ by $T(B_1)$, perhaps shrinking r at the same time, the proof would be complete; we now proceed to accomplish this.

Since T commutes with dilations, it follows that if $\|y\| < r2^{-n}$, then $y \in$ and proceeding inductively, we can find $x_n \in B_{2^{-n}}$ such that $\|y - \sum_1^n Tx_j\| < r2^{-n-1}$. Since X is complete, by Theorem 8 the series $\sum_1^\infty x_n$ converges, say to x. But then $\|x\| < \sum_1^\infty 2^{-n} = 1$ and $y = Tx$. In other words, $T(B_1)$ contains all y with $\|y\| < r/2$, so we are done. □

Corollary 5.62: 5.11: The (Bounded) Inverse Mapping Theorem.

If X and Y are Banach spaces and $T \in L(X, Y)$ is bijective, then T is an isomorphism; that is, $T^{-1} \in L(Y, X)$.

Proof. If T is bijective, continuity of T^{-1} is equivalent to the openness of T. □

For the next results we need some more terminology. If X and Y are normed vector spaces and T is a linear map from X to Y, we define the ***graph*** of T to be

$$\Gamma(T) = \{(x, y) \in X \times Y \mid y = Tx\}$$

which is a subspace of $X \times Y$. (From a strict set-theoretic point of view, of course, T and $\Gamma(T)$ are identical; the distinction is a psychological one.) We say that T is ***closed*** if $\Gamma(T)$ is a closed subspace of $X \times Y$.

Clearly, if T is continuous, then T is closed, and if X and Y are complete the converse is also true:

Theorem 5.63: 5.12: The Closed Graph Theorem.

If X and Y are Banach spaces and $T: X \to Y$ is a closed linear map, then T is bounded.

Note 64. *Energy is not bounded, but wants to be symmetric, hence not everywhere defined by uncertainty principle. Unbdd symmetric operators cannot be everywhere defined. Thus you need unbounded operators such that the domain is not the enetire space but a dense subspace. And we are kind of forced into that by this theorem.*

here. The proof is to write T as a composition of two bounded operators. Let π_1 and π_2 be the projections of $\Gamma(T)$ onto X and Y, that is, $\pi_1(x, Tx) = x$ and $\pi_2(x, Tx) = Tx$. Obviously $\pi_1 \in L(\Gamma(T), x)$ and $\pi_2 \in L(\Gamma(T), y)$. Since X and Y are complete, so is $X \times Y$, and hence so is $\Gamma(T)$ since T is closed. The map π_1 is a bijection from $\Gamma(T)$ to X, so by Corollary 62, π_1^{-1} is bounded. But then $T = \pi_2 \circ \pi_1^{-1}$ is bounded. □

Remark 65. *Continuity of a linear map $T: X \to Y$ means that if $x_n \to x$ then $Tx_n \to Tx$, whereas closedness means that if $x_n \to x$ and $Tx_n \to Y$ then $y = Tx$. Thus the significance* *of the closed graph theorem is that in verifying that $Tx_n \to Tx$ when $x_n \to x$, we may assume that Tx_n converges to something, and we need only to show that the limit is the right thing. This frequently saves a lot of trouble.*

The completeness of x and y was used in a crucial way in proving the open mapping theorem and hence also in proving the closed graph theorem. In fact, the conclusions of both of these theorems may fail if either x or y is incomplete; see Folland Exercise 5.29, Folland Exercise 5.30, and Folland Exercise 5.31.

Our final result in this section is a theorem of almost magical power that allows one to deduce uniform estimates from pointwise estimates in certain situations.

Theorem 5.66: 5.13: The Uniform Boundedness Principle.

Suppose that X and Y are normed vector spaces and $\mathcal{A}$ is a subset of $L(X, Y)$.

(a) If $\sup_{T\in\mathcal{A}} \|Tx\| < \infty$ for all x in some nonmeager subset of X, then $\sup_{T\in\mathcal{A}} \|T\| = \infty$. ∞.

(b) If X is a Banach space and $\sup_{T\in\mathcal{A}} \|Tx\| < \infty$ for all $x \in X$, then $\sup_{T\in\mathcal{A}} \|T\| = \infty$.

Proof. Let

$$E_n = \{x \in X \mid \sup_{T\in\mathcal{A}} \|Tx\| \leqslant n\} = \bigcap_{T\in\mathcal{A}} \{x \in X \mid \|Tx\| \leqslant n\}.$$

Then the E_ns are closed, so under the hypothesis of (a) some E_n must contain a nontrivial closed ball $\overline{B(r, x_0)}$. But then $E_{2n} \supset \overline{B(r, 0)}$, for if $\|x\| \leqslant r$, then $x - x_0 \in E_n$ and hence

$$\|Tx\| \leqslant \|T(x - x_0)\| + \|Tx_0\| \leqslant 2n.$$

In other words, $\|Tx\| \leqslant 2n$ whenever $T \in \mathcal{A}$ and $\|x\| \leqslant r$, so $\sup_{T\in\mathcal{A}} \|T\| \leqslant 2n/r$. This proves (a), and (b) follows by the Baire category theorem. ☐

Corollary 5.67.

Let X and Y be Banach spaces and $\{T_n\}_{n=1}^{\infty} \subset L(X, Y)$ such that $\lim_{n\to\infty} T_n x$ exists for all $x \in X$. Then define $T\colon X \to Y$ by

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

Then $T \in L(X, Y)$.

Proof. The hypothesis of the UBP are satisfied, so there exist $M > 0$ (independent of n such that $\|T_n x\| \leqslant M\|X\|$) for all $x \in X$. Then $\|Tx\| \leqslant \|Tx - T_n x\| + \|T_n x\| \leqslant \|Tx - T_n x\| + M\|x\|$. Letting $n \to \infty$, we obtain $\|Tx\| \leqslant M\|x\|$ for all $x \in X$. ☐

Exercise 5.68: Folland Exercise 5.27.

There exist meager subsets of $\mathbb{R}$ whose complements have Lebesgue measure zero.

Exercise 5.69: Folland Exercise 5.28.

The Baire category theorem remains true if X is assumed to be an LCH space rather than a complete metric space. (The proof is similar; the substitute for completeness is Proposition 87.)

Solution. [here] See Note 55. □

Exercise 5.70: Folland Exercise 5.29.

Let $Y = L^1(\mu)$ where μ is counting measure on $\mathbb{Z}_{\geqslant 1}$, and let $x = \{f \in Y \mid \sum_1^\infty n|f(n)| < \infty\}$, equipped with the L^1 norm.

(a) X is a proper dense subspace of Y; hence X is not complete.
(b) Define $T\colon X \to Y$ by $Tf(n) = nf(n)$. Then T is closed but not bounded.
(c) Let $S = T^{-1}$. Then $S\colon Y \to X$ is bounded and surjective but not open.

Solution. Let $\mathbb{K}$ denote $\mathbb{R}$ or $\mathbb{C}$. As μ is the counting measure on $\mathbb{Z}_{\geqslant 1}$, we can make the identifications

$$Y = \left\{\{a_n\} \,\middle|\, a_n \in \mathbb{K} \text{ and } \sum\nolimits_1^\infty |a_n| < \infty\right\}$$

and

$$X = \left\{\{a_n\} \,\middle|\, a_n \in \mathbb{K} \text{ and } \sum\nolimits_1^\infty n|a_n| < \infty\right\}.$$

(a)
- X is properly contained in Y: First note X is contained in Y, since if $\sum_1^\infty n|a_n| < \infty$ then $\sum_1^\infty n|a_n| < \infty$. The containment is proper, since the sequence $a_n = 1/n^2$ has $\{a_n\}_{n=1}^\infty \in Y \smallsetminus X$. Hence $X \subsetneq Y$.
- X is a linear subspace of Y: Let $\{a_n\}, \{b_n\} \in X$ and $\lambda \in \mathbb{K}$. Then for any $N \in \mathbb{Z}_{\geqslant 1}$,
$$\sum\nolimits_1^N n|a_n + \lambda b_n| \leqslant \sum\nolimits_1^N (n|a_n| + n|b_n|) + \sum\nolimits_1^N n|a_n| + \sum\nolimits_1^N n|b_n|.$$
Sending $n \to \infty$, we obtain
$$\sum\nolimits_1^\infty n|a_n + \lambda b_n| \leqslant \sum\nolimits_1^\infty n|a_n| + \sum\nolimits_1^\infty n|b_n| < \infty,$$
where the last inequality is because $\{a_n\}, \{b_n\} \in X$. Hence $\{\lambda a_n + b_n\} \in X$, so X is a linear subspace.
- X is dense in Y: Since simple functions are dense in $Y = L^1(\mu)$, it suffices to show X contains all simple functions in $L^1(\mu)$. So let $g = \{b_n\} \in L^1(\mu)$ be a simple function, that is, $g = \sum_1^N z_j \chi_{E_j}$ for finitely many $E_j \in \mathcal{P}(\mathbb{Z}_{\geqslant 1})$. Note that there exist at most finitely many $n \in \mathbb{Z}_{\geqslant 1}$ such that $b_n \neq 0$: indeed, if there exists $k \in \{1, \ldots, N\}$ such that both $z_j \neq 0$ and E_k is an infinite set, then
$$\infty = \sum\nolimits_{\ell=1}^\infty c_k \mu(E_k) \leqslant \sum\nolimits_{\ell=1}^\infty c_\ell(E_\ell) = \int g \, \mathrm{d}\mu,$$

contradicting $g \in L^1(\mu)$. Thus $\int g = \sum_{n=1}^{\infty} n|b_n|$ is a finite sum, and hence is finite. It follows that $g \in Y$, so Y is dense in X.

(b) – T is not bounded: Fix an arbitrary $m \in \mathbb{Z}_{\geqslant 1}$ and define $f_m(n) = 1$ if $m = n$ and $f_m(n) = 0$ otherwise. Then $\sum_n n|f_m(n)| = n < \infty$, so $f_m \in X$. But $\|Tf_m\| = \sum_n n|Tf_m(n)| = \sum_n n^2|f_m(m)| = m^2 = m\|f_n\|$, so $\|T\|_{\text{op}} \leqslant m$. But m was an arbitrary nonnegative integer, so $\|T\|_{\text{op}} = \infty$. Hence T is not bounded.

– T is closed: Suppose $f(n) \to f$ in X and $Tf(n) \to g$ in Y. We claim $Tf = g$. First fix $\varepsilon > 0$. By our assumption, for all sufficiently large N we have $\sum_{n=N}^{\infty} n|f(n)| < \varepsilon/4$, $\sum_{n=N}^{\infty} |g(n)| < \varepsilon/4$, $\|g - Tf_n\| < \varepsilon/4$, and $\|f - f_n\| < \frac{\varepsilon}{4N}$. Then for all sufficiently large m and N, we have

$$\sum_{n=1}^{\infty} |Tf(n) - Tf_m(n)| = \sum_{n=1}^{N-1} |nf(n) - nf_m(n)| + \sum_{n=N}^{\infty} |nf(n) - Tf_m(n)|$$
$$< \sum_{n=1}^{N-1} |f(n) - f_m(n)| + \varepsilon/4 + \sum_{n=N}^{\infty} |Tf_m(n) - g(n)| + \sum_{n=N}^{\infty} |g(n)| < \varepsilon,$$

so $Tf_n \to Tf$ in L^1. Since $Tf(n) \to g$ by assumption, we conclude by uniqueness of limits in a normed (hence Hausdorff) vector space (namely, $L^1(\mu)$) that $Tf = g$.

(c) Fix $f \in Y$. Then $Sf(n) = n^{-1}f(n)$ for any $n \in \mathbb{Z}_{\geqslant 1}$, so

$$\|Sf\| = \sum_{n=1}^{\infty} n^{-1}|f(n)| \leqslant \sum_{n=1}^{\infty} |f(n)| = \|f\|.$$

Thus $\|S\|_{\text{op}} \leqslant 1$, so S is bounded. And S is surjective, since any $\{a_n\} \in X$ is the image under S of the sequence $\{\frac{a_n}{n}\}$ (since if $\sum n|a_n| < \infty$ then in particular $\sum \frac{1}{n}|a_n| < \infty$, meaning $\{\frac{a_n}{n}\} \in Y$). Lastly, if S were open, then $T = S^{-1}$ is continuous, which contradicts part (b). Thus S is not an open map, as claimed. □

Exercise 5.71: Folland Exercise 5.30.

Let $Y = C([0,1])$ and $X = C^1([0,1])$, both equipped with the uniform norm.

(a) x is not complete.

(b) The map $(d/dx)\colon X \to Y$ is closed (see Folland Exercise 5.9) but not bounded.

Exercise 5.72: Folland Exercise 5.31.

Let X, y be Banach spaces and let $S\colon X \to Y$ be an unbounded linear map (for the existence of which, see Folland Section 5.6). Let $\Gamma(S)$ be the graph of S, a subspace of $x \times y$.

(a) $\Gamma(S)$ is not complete.

(b) Define $T\colon X \to \Gamma(S)$ by $Tx = (x, Sx)$. Then T is closed but not bounded.

(c) $T^{-1}\colon \Gamma(S) \to X$ is bounded and surjective but not open.

Exercise 5.73: Folland Exercise 5.32.

Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be norms on the vector space X such that $\|\cdot\|_1 \leqslant \|\cdot\|_2$. If X is complete with respect to both norms, then the norms are equivalent.

Exercise 5.74: Folland Exercise 5.33.

There is no slowest rate of decay of the terms of an absolutely convergent series; that is, there is no sequence $\{a_n\}$ of positive numbers such that $\sum a_n|c_n| < \infty$ if and only if $\{c_n\}$ is bounded. (The set of bounded sequences is the space $B(\mathbb{Z}_{\geqslant 1})$ of bounded functions on $\mathbb{Z}_{\geqslant 1}$, and the set of absolutely summable sequences is $L^1(\mu)$ where μ is counting measure on $\mathbb{Z}_{\geqslant 1}$. If such an $\{a_n\}$ exists, consider $T\colon B(\mathbb{Z}_{\geqslant 1}) \to L^1(\mu)$ defined by $Tf(n) = a_n f(n)$. The set of f such that $f(n) = 0$ for all but finitely many n is dense in $L^1(\mu)$ but not in $B(\mathbb{Z}_{\geqslant 1})$.)

Solution. The linear operator $T\colon (X, \|-\|_2) \to (X, \|-\|_1)$ defined by $Tx \coloneqq x$ is bounded, since by hypothesis $\|Tx\|_1 = \|x\|_1 \leqslant \|x\|_2$ for all $x \in X$. Since T is a bijection of sets, $T^{-1} \in L((X, \|-\|_1), (X, \|-\|_2))$ by the bounded inverse mapping theorem. Hence there exists $C_2 > 0$ such that $\|x\|_2 = \|T^{-1}x\|_2 \leqslant C\|x\|_1$. Thus

$$\|x\|_1 \leqslant \|x\|_2 \leqslant C\|x\|_1$$

for all $x \in X$, so $\|-\|_1$ and $\|-\|_2$ are equivalent. □

Exercise 5.75: Folland Exercise 5.34.

With reference to Folland Exercise 5.9 and Folland Exercise 5.10, show that the inclusion map of $L^1_k([0,1])$ into $C^{k-1}([0,1])$ is continuous (a) by using the closed graph theorem, and (b) by direct calculation. (This is to illustrate the use of the closed graph theorem as a labor-saving device.)

Exercise 5.76: Folland Exercise 5.35.

Let X and Y be Banach spaces, $T \in L(X,Y), \mathcal{N}(T) = \{x \mid Tx = 0\}$, and $\mathcal{M} =$ range (T). Then $X/\mathcal{N}(T)$ is isomorphic to $\mathcal{M}$ if and only if $\mathcal{M}$ is closed. (See Folland Exercise 5.15.)

Exercise 5.77: Folland Exercise 5.36.

Let X be a separable Banach space and let μ be counting measure on $\mathbb{Z}_{\geqslant 1}$. Suppose that $\{x_n\}_1^\infty$ is a countable dense subset of the unit ball of X, and define $T\colon L^1(\mu) \to X$ by $Tf = \sum_1^\infty f(n)x_n$.

(a) T is bounded.

(b) T is surjective.
(c) X is isomorphic to a quotient space of $L^1(\mu)$. (Use Folland Exercise 5.35.)

Exercise 5.78: Folland Exercise 5.37.

Let X and Y be Banach spaces. If $T\colon X \to Y$ is a linear map such that $f \circ T \in X^*$ for every $f \in Y^*$, then T is bounded.

Solution. Suppose $x_n \to x$ and $Tx_n \to y$. We claim $y = Tx$. On one hand, by continuity of f we have

$$\lim_{n\to\infty} f \circ T(x_n) = f\Big(\lim_{n\to\infty} Tx_n\Big) = f(y).$$

On the other hand, $f \circ T \in X^*$ by hypothesis, so in particular $f \circ T$ is continuous; hence

$$\lim_{n\to\infty} f \circ T(x_n) = f \circ T\Big(\lim_{n\to\infty} x_n\Big) = f \circ T(x).$$

Thus

$$f(y) = f \circ T(x) \text{ for all } y \in Y^*. \tag{5.78.1}$$

It follows that $y = Tx$, since otherwise there exists $f \in Y^*$ such that $f(y) \neq f(Tx)$ (since by a corollary to the Hahn-Banach theorem X^* separates points), contradicting Equation (5.78.1). It then follows that the graph of T is closed, so by the closed graph theorem T is bounded. ☐

Exercise 5.79: Folland Exercise 5.38.

Let X and Y be Banach spaces, and let $\{T_n\}$ be a sequence in $L(X, Y)$ such that $\lim T_n x$ exists for every $x \in X$. Let $Tx = \lim T_n x$; then $T \in L(X, Y)$.

Exercise 5.80: Folland Exercise 5.39.

Let x, y, z be Banach spaces and let $B\colon x \times y \to z$ be a separately continuous bilinear map; that is, $B(x, \cdot) \in L(y, z)$ for each $x \in X$ and $B(\cdot, y) \in L(X, Z)$ for each $y \in Y$. Then B is jointly continuous, that is, continuous from $x \times y$ to z. (Reduce the problem to proving that $\|B(x, y)\| \leqslant C\|x\|\|y\|$ for some $C > 0$.)

Exercise 5.81: Folland Exercise 5.40: The Principle of Condensation of Singularities.

Let X and Y be Banach spaces and $\{T_{jk} \mid j, k \in \mathbb{Z}_{\geqslant 1}\} \subset L(X, Y)$. Suppose that for each k there exists $x \in X$ such that $\sup\{\|T_{jk}x\| \mid j \in \mathbb{N}\} = \infty$. Then there is an x (indeed, a residual set of xs) such that $\sup\{\|T_{jk}x\| \mid j \in \mathbb{Z}_{\geqslant 1}\} = \infty$ for all k.

Exercise 5.82: Folland Exercise 5.41.

Let X be a vector space of countably infinite dimension (that is, every element is a finite linear combination of members of a countably infinite linearly independent set). There is no norm on X with respect to which X is complete. (Given a norm on X, apply Folland Exercise 5.18(b) and the Baire category theorem.)

Exercise 5.83: Folland Exercise 5.42.

Let E_n be the set of all $f \in C([0,1])$ for which there exists $x_0 \in [0,1]$ (depending on f) such that $|f(x) - f(x_0)| \leqslant n|x - x_0|$ for all $x \in [0,1]$.

(a) E_n is nowhere dense in $C([0,1])$. (Any real $f \in C([0,1])$ can be uniformly approximated by a piecewise linear function g whose linear pieces, finite in number, have slope $\pm 2n$. If $\|h - g\|_u$ is sufficiently small, then $h \notin E_n$.)

(b) The set of nowhere differentiable functions is residual in $C([0,1])$.

Exercise 5.84.

Assume that T is a bounded linear map on $L^2([0,1])$ with the property that Tf is continuous on $[0,1]$ whenever f is continuous on $[0,1]$. Prove that the restriction of T to $C([0,1])$ is a bounded operator on $C([0,1])$, where as usual $C([0,1])$ is equipped with the uniform norm.

Solution. We will use the closed graph theorem. Suppose both $f_n \to f$ and $Tf_n \to g$ uniformly. We claim $Tf = g$. We first state and prove a useful lemma:

Lemma 5.85.

For all $f \in C([0,1])$ and all real numbers $p \in [1, \infty)$, $\|f\|_{L^p} \leqslant \|f\|_u$, where $\|-\|_u$ is the sup-norm.

Proof. Since $f \in C([0,1])$, $\|f\|_u$ is finite. Thus

$$\|f\|_{L^p}^p = \int_0^1 |f|^p \, \mathrm{d}y \leqslant \int_0^1 \|f\|_u^p \, \mathrm{d}y = \|f\|_u^p.$$

Taking the pth root of both sides, we obtain the desired inequality $\|f\|_{L^p} \leqslant \|f\|_u$. □

Since $T \in L(L^2([0,1]), L^2([0,1]))$, there exists $C > 0$ such that

$$\|Tf_n - Tf\|_{L^2} \leqslant C\|f_n - f\|_{L^2} \leqslant C\|f_n - f\|_u,$$

where the final inequality is by Lemma 85. Since $f_n \to f$ uniformly, it follows that $Tf_n \to Tf$ in $L^2([0,1])$. But also $Tf_n \to g$ uniformly by assumption, so in particular $Tf_n \to g$ in $L^2([0,1])$. And $L^2([0,1])$ is Hausdorff as a normed vector space, so by uniqueness of limits $Tf = g$. Thus, by the closed graph theorem, we conclude $T \in L(C([0,1]), C([0,1]))$. □

5.4 Topological Vector Spaces

It is frequently useful to consider topologies on vector spaces other than those defined by norms, the only crucial requirement being that the topology should be well behaved with respect to the vector operations. Precisely, a topological vector space is a vector space X over the field K ($= \mathbb{R}$ or $\mathbb{R}$) which is endowed with a topology such that the maps $(x, y) \to x + y$ and $(\lambda, x) \to \lambda x$ are continuous from $x \times x$ and $K \times x$ to x. A topological vector space is called locally convex if there is a base for the topology consisting of convex sets (that is, sets A such that if $x, y \in A$ then $tx + (1 - t)y \in A$ for $0 < t < 1$). Most topological vector spaces that arise in practice are locally convex and Hausdorff.

The most common way of defining locally convex topologies on vector spaces is in terms of seminorms. Namely, if we are given a family of seminorms on X, the "balls" that they define can be used to generate a topology in the same way that the balls defined by a norm generate the topology on a normed vector space. The precise result is as follows:

Theorem 5.86: 5.14.

Let $\{p_\alpha\}_{\alpha\in A}$ be a family of seminorms on the vector space X. If $x \in X, \alpha \in A$, and $\varepsilon > 0$, let

$$U_{x\alpha\varepsilon} = \{y \in X \mid p_\alpha(y - x) < \varepsilon\},$$

and let $\mathcal{T}$ be the topology generated by the sets $U_{x\alpha\varepsilon}$.

(a) For each $x \in X$, the finite intersections of the sets $U_{x\alpha\varepsilon}(\alpha \in A, \varepsilon > 0)$ form a neighborhood base at x.

(b) If $\langle x_i\rangle_{i\in I}$ is a net in X, then $x_i \to x$ if and only if $p_\alpha(x_i - x) \to 0$ for all $\alpha \in A$.

(c) $(X, \mathcal{T})$ is a locally convex topological vector space.

Proof. (a) If $x \in \bigcap_1^k U_{x_j\alpha_j\varepsilon_j}$, let $\delta_j = \varepsilon_j - p_\alpha(x - x_j)$. By the triangle inequality, we have $x \in \bigcap_1^k U_{x\alpha_j\delta_j} \subset \bigcap_1^k U_{x_j\alpha_j\varepsilon_j}$. Thus the assertion follows from Proposition 7.

(b) In view of (a), it suffices to observe that $p_\alpha(x_i - x) \to 0$ if and only if $\langle x_i\rangle$ is eventually in $U_{x\alpha\varepsilon}$ for every $\varepsilon > 0$.

(c) The continuity of the vector operations follows easily from Proposition 72 and part (b). Indeed, if $x_i \to x$ and $y_i \to Y$, then

$$p_\alpha((x_i + y_i) - (x + y)) \leqslant p_\alpha(x_i - x) + p_\alpha(y_i - y) \to 0,$$

so $x_i + y_i \to x + y$. If also $\lambda_i \to \lambda$, then eventually $|\lambda_i| \leqslant C = |\lambda| + 1$, so

$$p_\alpha(\lambda_i x_i - \lambda x) \leqslant p_\alpha(\lambda_i(x_i - x)) + p_\alpha((\lambda_i - \lambda)x) \leqslant Cp_\alpha(x_i - x) + |\lambda_i - \lambda|p_\alpha(x),$$

and it follows that $\lambda_i x_i \to \lambda x$. Moreover, the sets $U_{x\alpha\varepsilon}$ are convex, for if $y, z \in U_{x\alpha\varepsilon}$, then

$$p_\alpha(x - [ty + (1 - t)z]) \leqslant p_\alpha(tx - ty) + p_\alpha((1 - t)x + (1 - t)z) < t\varepsilon + (1 - t)\varepsilon = \varepsilon. \square$$

The local convexity of the topology therefore follows from (a).

In this context there is an analogue of Proposition 14:

Proposition 5.87: 5.15.

Suppose X and y are vector spaces with topologies defined, respectively, by the families $\{p_\alpha\}_{\alpha\in A}$ and $\{q_\beta\}_{\beta\in B}$ of seminorms, and $T\colon X \to Y$ is a linear map. Then T is continuous if and only if for each $\beta \in B$ there exist $\alpha_1, \ldots, \alpha_k \in A$ and $C > 0$ such that $q_\beta(Tx) \leqslant C\sum_1^k p_{\alpha_j}(x)$.

Proof. If the latter condition holds and $\langle x_i\rangle$ is a net converging to $x \in X$, by Theorem 86(b) we have $p_\alpha(x_i - x) \to 0$ for all α, hence $q_\beta(Tx_i - Tx) \to 0$ for all β, hence $Tx_i \to Tx$. By Proposition 72, T is continuous. Conversely, if T is continuous, for every $\beta \in B$ there is a neighborhood U of 0 in X such that $q_\beta(Tx) < 1$ for $x \in U$. By Theorem 86(a) we may assume that $U = \bigcap_1^k U_{x\alpha_j\varepsilon_j}$. Let $\varepsilon = \min(\varepsilon_1, \ldots, \varepsilon_k)$; then $q_\beta(Tx) < 1$ whenever $p_{\alpha_j}(x) < \varepsilon$ for all j. Now, given $x \in X$, there are two possibilities. If $p_{\alpha_j}(x) > 0$ for some j, let $y = \varepsilon x / \sum_1^k p_{\alpha_j}(x)$. Then $p_{\alpha_j}(y) < \varepsilon$ for all j, so

$$q_\beta(Tx) = \sum_1^k \varepsilon^{-1} p_{\alpha_j}(x) q_\beta(Ty) \leqslant \varepsilon^{-1}\sum_1^k p_{\alpha_j}(x)$$

On the other hand, if $p_{\alpha_j}(x) = 0$ for all j, then $p_{\alpha_j}(rx) = 0$ for all j and all $r > 0$, hence $rq_\beta(Tx) = q_\beta(T(rx)) < 1$ for all $r > 0$, hence $q_\beta(Tx) = 0$. Thus $q_\beta(Tx) \leqslant \varepsilon^{-1}\sum_1^k p_{\alpha_j}(x)$ in this case too, and we are done. □

The proof of the following proposition is left to the reader (Folland Exercise 5.43).

Proposition 5.88: 5.16.

Let x be a vector space equipped with the topology defined by a family $\{p_\alpha\}_{\alpha\in A}$ of seminorms.

(a) X is Hausdorff if and only if for each $x \neq 0$ there exists $\alpha \in A$ such that $p_\alpha(x) \neq 0$.

(b) If X is Hausdorff and A is countable, then X is metrizable with a translation-invariant metric (i.e., $\rho(x, y) = \rho(x + z, y + z)$ for all $x, y, z \in X$).

If X has the topology defined by the seminorms $\{p_\alpha\}_{\alpha\in A}$, by Proposition 87 a linear functional f on X is continuous if and only if $|f(x)| \leqslant C\sum_1^k p_{\alpha_j}(x)$ for some $C > 0$ and $\alpha_1, \ldots, \alpha_k \in A$. Since a finite sum of seminorms is again a seminorm, the HahnBanach theorem guarantees the existence of lots of continuous linear functionals on x—enough to separate points, if X is Hausdorff. The set of all such functionals is denoted, as before, by X^*. There are various ways of making X^* into a topological vector space, but we shall not consider this question systematically. The simplest way is to impose the weakest topology that makes all the evaluation maps $f \mapsto f(x)$ $(x \in X)$ continuous, an idea that we shall discuss further below.

In a topological vector space X the notion of Cauchy sequence or Cauchy net makes sense. Namely, a net $\langle x_i\rangle_{i\in I}$ in X is called Cauchy if the net $\langle x_i - x_j\rangle_{(i,j)\in I\times I}$ converges to zero. (Here $I \times I$ is directed in the usual way: $(i, j) \precsim (i', j')$ if and only if $i \precsim i'$ and

$j \precsim j'$.) Naturally, X is called complete if every Cauchy net converges. Completeness is of most interest when X is first countable, in which case it is equivalent to the condition that every Cauchy sequence converges (Folland Exercise 5.44). More particularly, if X is Hausdorff and its topology is defined by a countable family of seminorms, then this topology is first countable by Theorem 86(a); indeed, it is given by a translation-invariant metric ρ by Proposition 88(b), and a sequence is Cauchy according to the definition just given if and only if it is Cauchy with respect to ρ. A complete Hausdorff topological vector space whose topology is defined by a countable family of seminorms is called a Fréchet space.

Let us now consider some interesting examples of topological vector spaces whose topologies are defined by families of seminorms rather than by single norms. We have already seen some:

- Let X be an LCH space. On $\mathbb{C}^X$, the topology of uniform convergence on compact sets is defined by the seminorms $p_K(f) = \sup_{x\in K} |f(x)|$ as K ranges over compact subsets of X. If X is σ-compact and $\{U_n\}$ are as in Propositions 122 and 123, this topology is defined by the seminorms $p_n(f) = \sup_{x\in\overline{U}_n} |f(x)|$. In this case, $\mathbb{C}^X$ is easily seen to be complete, so it is a Fréchet space; by Proposition 121, so is $C(X)$.
- The space $L^1_{\text{loc}}(\mathbb{R}^n)$, defined in Folland Section 3.4, is a Fréchet space with the topology defined by the seminorms $p_k(f) = \int_{|x|\leqslant k} |f(x)|dx$. (Completeness follows easily from the completeness of L^1.) An obvious generalization of this construction yields a locally convex topological vector space $L^1_{\text{loc}}(X, \mu)$ where X is any LCH space and μ is a Borel measure on X that is finite on compact sets.

Another class of topological vector spaces arises naturally in connection with the theory of differential equations. One often wishes to study the operator d/dx, or more complicated operators constructed from it, acting on various spaces of functions. Unfortunately, it is virtually impossible to define norms on most infinite-dimensional functions spaces so that d/dx becomes a bounded operator. Here is one precise result along these lines: There is no norm on the space $C^\infty([0, 1])$ of infinitely differentiable functions on $[0, 1]$ with respect to which d/dx is bounded. Indeed, if $f_\lambda(x) = e^{\lambda x}$, then $(d/dx)f_\lambda = \lambda f_\lambda$, so $\|d/dx\| \geqslant |\lambda|$ for all λ no matter what norm is used on $C^\infty([0, 1])$.

In view of this difficulty, three courses of action are available. First, one can consider differentiation as an unbounded operator from x to y where y is a suitable Banach space and x is a dense subspace of y, as in Folland Exercise 5.30. Second, one can consider differentiation as a bounded linear map from one Banach space X to a different one y, such as $x = C^k([0, 1])$ and $y = C^{k-1}([0, 1])$ in Folland Exercise 5.9. Finally, one can consider differentiation as a continuous operator on a locally convex space X whose topology is not given by a norm. All of these points of view have their uses, but it is the last one that concerns us here. It is easy to construct families of seminorms on spaces of smooth functions such that differentiation becomes continuous almost by definition. For example, the seminorms $p_k(f) = \sup_{0<x<1} |f^{(k)}(x)| (k = 0, 1, 2, \ldots)$ make $C^\infty([0, 1])$ into a Fréchet space (the completeness is proved as in Folland Exercise 5.9), and d/dx is continuous on this space by Proposition 87 since $p_k(f') = p_{k+1}(f)$. Other examples are considered in Folland Folland Exercise 5.45 and in Folland Chapter 9.

One of the most useful procedures for constructing topologies on vector spaces is by requiring the continuity of certain linear maps. Namely, suppose that X is a vector space, y is a normed linear space, and $\{T_\alpha\}_{\alpha \in A}$ is a collection of linear maps from x to y. Then the weak topology $\mathcal{T}$ generated by $\{T_\alpha\}$ makes X into a locally convex topological vector space. Indeed, $\mathcal{T}$ is just the topology $\mathcal{T}'$ defined by the seminorms $p_\alpha(x) = \|T_\alpha x\|$ according to Theorem 86. ($\mathcal{T}$ is generated by sets of the form $\{x \mid \|T_\alpha x - y_0\| < \varepsilon\}$ with $y_0 \in Y$, whereas $\mathcal{T}'$ is generated by sets of the form $\{x \mid \|T_\alpha x - T_\alpha x_0\| < \varepsilon\}$ with $x_0 \in X$. If the T_α's are surjective, these are obviously the same; the general case is left as Folland Exercise 5.46.) The topology on $C^\infty([0,1])$ in the preceding paragraph is an example of this construction, with $y = C([0,1])$ and $T_k f = f^{(k)}$. We now present some more.

First, let X be a normed vector space. The weak topology generated by X^* is known simply as the weak topology on X, and convergence with respect to this topology is known as weak convergence. Thus, if $\langle x_\alpha \rangle$ is a net in $X, x_\alpha \to x$ weakly if and only if $f(x_\alpha) \to f(x)$ for all $f \in X^*$. When X is infinite-dimensional, the weak topology is always weaker than the norm topology; see Folland Exercise 5.49

Next, let X be a normed vector space, $X*$ its dual space. The weak topology on X^* as defined above is the topology generated by X^{**}; of more interest is the topology generated by X (considered as a subspace of X^{**}), which is called the ***weak* topology*** (read "weak star topology") on X^*. X^* is a space of functions on X, and the weak* topology is simply the topology of pointwise convergence: $f_\alpha \to f$ if and only if $f_\alpha(x) \to f(x)$ for all $x \in X$. The weak* topology is even weaker than the weak topology on X^*; the two coincide precisely when X is reflexive.

Finally, Let X and Y be Banach spaces. The topology on $L(X,Y)$ generated by the evaluation maps $T \mapsto Tx(x \in X)$ is called the strong operator topology on $L(X,Y)$, and the topology generated by the linear functionals $T \mapsto f(Tx)(x \in X,\ f \in Y^*)$ is called the weak operator topology on $L(X,Y)$. Again, these topologies are best understood in terms of convergence: $T_\alpha \to T$ strongly if and only if $T_\alpha x \to Tx$ in the norm topology of y for each $x \in X$, whereas $T_\alpha \to T$ weakly if and only if $T_\alpha x \to Tx$ in the weak topology of y for each $x \in X$. Thus the strong operator topology is stronger than the weak operator topology but weaker than the norm topology on $L(X,Y)$.

The following result concerning strong convergence is almost trivial but extremely useful:

Proposition 5.89: 5.17.

Suppose $\{T_n\}_1^\infty \subset L(X,Y)$, $\sup_n \|T_n\| < \infty$, and $T \in L(X,Y)$. If $\|T_n x - Tx\| \to 0$ for all x in a dense subset D of X, then $T_n \to T$ strongly.

Proof. Let $C = \sup\{\|T\|, \|T_1\ |\ , \|T_2\|, \ldots\}$. Given $x \in X$ and $\varepsilon > 0$, choose $x' \in D$ such that $\|x - x'\| < \varepsilon/3C$. If n is large enough so that $\|T_n x' - Tx'\| < \varepsilon/3$, we have

$$\begin{aligned}\|T_n x - Tx\| &\leqslant \|T_n x - T_n x'\| + \|T_n x' - Tx'\| + \|Tx' - Tx\| \\ &\leqslant 2C\|x - x'\| + \frac{1}{3}\varepsilon < \varepsilon,\end{aligned}$$

so that $T_n x \to Tx$. □

Our final result in this section is a compactness theorem that is one of the main reasons for the usefulness of the weak* topology on a dual space.

Theorem 5.90: 5.18: Alaoglu's Theorem.

If X is a normed vector space, the closed unit ball $B^* = \{f \in X^* \mid \|f\| \leqslant 1\}$ in X^* is compact in the weak k^* topology.

Proof. For each $x \in X$ let $D_x = \{z \in \mathbb{C} \mid |z| \leqslant \|x\|\}$, and let $D = \prod_{x \in X} D_x$. Then D is compact by Tychonoff's theorem. The elements of D are precisely those complex-valued functions ϕ on X such that $|\phi(x)| \leqslant \|x\|$ for all $x \in X$, and B^* consists of those elements of D that are linear. Moreover, the relative topologies that B^* inherits from the product topology on D and the weak* topology on X^* both coincide with the topology of pointwise convergence, so it suffices to see that B^* is closed in D. But this is easy: If $\langle f_\alpha \rangle$ is a net in B^* that converges to $f \in D$, for any $x, y \in X$ and $a, b \in \mathbb{C}$ we have

$$f(ax + by) = \lim f_\alpha(ax + by) = \lim[af_\alpha(x) + bf_\alpha(y)] = af(x) + bf(y),$$

so that $f \in B^*$. □

Warning 5.91.

Alaoglu's theorem does not imply that $X*$ is locally compact in the weak* topology; see Folland Exercise 5.49(b).

Exercise 5.92: Folland Exercise 5.43.

Prove Proposition 88. (For part (b), proceed as in Folland Exercise 4.56(d).)

Exercise 5.93: Folland Exercise 5.44.

If X is a first countable topological vector space and every Cauchy sequence in X converges, then every Cauchy net in X converges.

Exercise 5.94: Folland Exercise 5.45.

The space $C^\infty(\mathbb{R})$ of all infinitely differentiable functions on $\mathbb{R}$ has a Fréchet space topology with respect to which $f_n \to f$ if and only if $f_n^{(k)} \to f^{(k)}$ uniformly on compact sets for all $k \geqslant 0$.

Exercise 5.95: Folland Exercise 5.46.

If X is a vector space, y a normed linear space, $\mathcal{T}$ the weak topology on X generated by a family of linear maps $\{T_\alpha \mid X \to Y\}$, and $\mathcal{T}'$ the topology defined by the seminorms $\{x \mapsto \|T_\alpha x\|\}$, then $\mathcal{T} = \mathcal{T}'$.

Exercise 5.96: Folland Exercise 5.47.

Suppose that X and y are Banach spaces.

(a) If $\{T_n\}_1^\infty \subset L(X, Y)$ and $T_n \to T$ weakly (or strongly), then $\sup_n \|T_n\| < \infty$.

(b) Every weakly convergent sequence in x, and every weak*-convergent sequence in X^*, is bounded (with respect to the norm).

Exercise 5.97: Folland Exercise 5.48.

Suppose that X is a Banach space.

(a) The norm-closed unit ball $B = \{x \in X \mid \|x\| \leqslant 1\}$ is also weakly closed. (Use Theorem 41(d).)

(b) If $E \subset X$ is bounded (with respect to the norm), so is its weak closure.

(c) If $F \subset X^*$ is bounded (with respect to the norm), so is its weak* closure.

(d) Every weak*-Cauchy sequence in X^* converges. (Use Folland Exercise 5.38.)

Exercise 5.98: Folland Exercise 5.49.

Suppose that X is an infinite-dimensional Banach space.

(a) Every nonempty weakly open set in X, and every nonempty weak*-open set in X^*, is unbounded (with respect to the norm).

(b) Every bounded subset of x is nowhere dense in the weak topology, and every bounded subset of $x*$ is nowhere dense in the weak* topology. (Use Folland Exercise 4.48(b,c).)

(c) X is meager in itself with respect to the weak topology, and $X*$ is meager in itself with respect to the weak* topology.

(d) The weak* topology on $x*$ is not defined by any translation-invariant metric. (Use Folland Exercise 5.48(d).)

Exercise 5.99: Folland Exercise 5.50.

If x is a separable normed linear space, the weak* topology on the closed unit ball in X^* is second countable and hence metrizable. (But see Folland Exercise 5.49(d).)

Exercise 5.100: Folland Exercise 5.51.

A vector subspace of a normed vector space X is norm-closed if and only if it is weakly closed. (However, a norm-closed subspace of X^* need not be weak*-closed unless x is reflexive; see Folland Exercise 5.52(d).)

Exercise 5.101: Folland Exercise 5.52.

Let X be a Banach space and let $f_1, \ldots, f_n$ be linearly independent elements of X^*.

(a) Define $T\colon X \to \mathbb{C}^n$ by $Tx = (f_1(x), \ldots, f_n(x))$. If $\mathcal{N} = \{x \mid Tx = 0\}$ and $\mathcal{M}$ is the linear span of $f_1, \ldots, f_n$, then $\mathcal{M} = \mathcal{N}^0$ in the notation of Folland Exercise 5.23 and hence $\mathcal{M}^*$ is isomorphic to $(X/\mathcal{N})^*$.

(b) If $F \in X^{**}$, for any $\varepsilon > 0$ there exists $x \in X$ such that $F(f_j) = f_j(x)$ for $j = 1, \ldots, n$ and $\|x\| \leqslant (1+\varepsilon)\|F\|$. ($F_{\mathcal{M}}$ can be identified with an element of $(X/\mathcal{N})^{**}$ and hence with an element of $X/\mathcal{N}$ since the latter is finite-dimensional.)

(c) If X is considered as a subspace of $X**$, the relative topology on X induced by the weak* topology on X^{**} is the weak topology on x.

(d) In the weak* topology on X^{**}, X is dense in X^{**} and the closed unit ball in x is dense in the closed unit ball in X^{**}.

(e) x is reflexive if and only if its closed unit ball is weakly compact.

Exercise 5.102: Folland Exercise 5.53.

Suppose that X is a Banach space and $\{T_n\}, \{S_n\}$ are sequences in $L(X, X)$ such that $T_n \to T$ strongly and $S_n \to S$ strongly.

(a) If $\{x_n\} \subset X$ and $\|x_n - x\| \to 0$, then $\|T_n x_n - Tx\| \to 0$. (Use Folland Exercise 5.47(a).)

(b) $T_n S_n \to TS$ strongly.

5.5 Hilbert Spaces

The most important Banach spaces, and the ones on which the most refined analysis can be done, are the Hilbert spaces, which are a direct generalization of finite-dimensional Euclidean spaces. Before defining them, we need to introduce a few concepts.

Definition 103. *Let $\mathcal{H}$ be a complex vector space. An* ***inner product*** *(or* ***scalar product****) on $\mathcal{H}$ is a map $(x, y) \mapsto \langle x|y\rangle$ from $X \times X \to \mathbb{C}$ such that:*

(i) $\langle ax + by|z\rangle = a\langle x|z\rangle + b\langle y|z\rangle$ *for all* $x, y, z \in \mathcal{H}$ *and* $a, b \in \mathbb{C}$.
(ii) $\langle y|x\rangle = \overline{\langle x|y\rangle}$ *for all* $x, y \in \mathcal{H}$.
(iii) $\langle x|x\rangle \in (0, \infty)$ *for all nonzero* $x \in X$.

We observe that (i) and (ii) imply that

$$\langle x|ay + bz\rangle = \bar{a}\langle x|y\rangle + \bar{b}\langle x|z\rangle \text{ for all } x, y, z \in \mathcal{H} \text{ and } a, b \in \mathbb{C}.$$

(One can also define inner products on real vector spaces: $\langle x|y\rangle$ is then real, a and b are assumed real in (i), and (ii) becomes $\langle y|x\rangle = \langle x|y\rangle$.)

Definition 104. *A complex vector space equipped with an inner product is called a* ***pre-Hilbert space****. If* $\mathcal{H}$ *is a pre-Hilbert space, for* $x \in \mathcal{H}$ *we define*

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

Theorem 5.105: 5.19: The Schwarz Inequality.

$|\langle x|y\rangle| \leqslant \|x\|\|y\|$ for all $x, y \in \mathcal{H}$, with equality if and only if x and y are linearly dependent.

Proof. If $\langle x|y\rangle = 0$, the result is obvious. If $\langle x|y\rangle \neq 0$ (and in particular $y \neq 0$), let $\alpha = \operatorname{sgn}\langle x|y\rangle$ and $z = \alpha y$, so that $\langle x|z\rangle = \langle z|x\rangle = |\langle x|y\rangle|$ and $\|z\| = \|y\|$. Then for $t \in \mathbb{R}$ we have

$$0 \leqslant \langle x - tz|x - tz\rangle = \|x\|^2 - 2t|\langle x|y\rangle| + t^2\|y\|^2.$$

The expression on the right is a quadratic function of t whose absolute minimum occurs at $t = \|y\|^{-2}|\langle x|y\rangle|$. Setting t equal to this value, we obtain

$$0 \leqslant \|x - tz\|^2 = \|x\|^2 - \|y\|^{-2}|\langle x|y\rangle|^2$$

with equality if and only if $x - tz = x - \alpha ty = 0$, from which the desired result is immediate. □

Proposition 5.106: 5.20.

The function $x \mapsto \|x\|$ is a norm on $\mathcal{H}$.

Proof. That $\|x\| = 0$ if and only if $x = 0$ and that $\|\lambda x\| = |\lambda|\|x\|$ are obvious from the definition. As for the triangle inequality, we have

$$\|x + y\|^2 = \langle x + y|x + y\rangle = \|x\|^2 + 2\operatorname{Re}\langle x|y\rangle + \|y\|^2,$$

so by the Schwarz inequality,

$$\|x + y\|^2 \leqslant \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2,$$

as desired. A pre-Hilbert space that is complete with respect to the norm $\|x\| = \sqrt{\langle x|x\rangle}$ is called a Hilbert space. (One can also consider real Hilbert spaces with real inner products. However, Hilbert spaces are usually assumed to be complex unless otherwise specified.)

Example: Let $(X, \mathcal{M}, \mu)$ be a measure space, and let $L^2(\mu)$ be the set of all measurable functions $f\colon X \to \mathbb{C}$ such that $\int |f|^2 d\mu < \infty$ (where, as usual, we identify two functions that are equal a.e.). From the inequality $ab \leqslant \frac{1}{2}(a^2 + b^2)$, valid for all $a, b \geqslant 0$, we see that if $f, g \in L^2(\mu)$ then $|f\overline{g}| \leqslant \frac{1}{2}(|f|^2 + |g|^2)$, so that $f\overline{g} \in L^1(\mu)$. It follows easily that the formula

$$\langle f|g\rangle = \int f\overline{g} d\mu$$

defines an inner product on $L^2(\mu)$. In fact, $L^2(\mu)$ is a Hilbert space for any measure μ. (For a proof of completeness, see Folland Theorem 8; for the present we shall take this result for granted.)

An important special case of this construction is obtained by taking μ to be counting measure on $(A, \mathcal{P}(A))$, where A is any nonempty set; in this situation $L^2(\mu)$ is usually denoted by $\ell^2(A)$. Thus, $\ell^2(A)$ is the set of functions $f\colon A \to \mathbb{C}$ such that the sum $\sum_{\alpha \in A} |f(\alpha)|^2$ (as defined in Folland Section 0.5) is finite. The completeness of $\ell^2(A)$ is rather easy to prove directly (Folland Exercise 5.54). For the remainder of this section, $\mathcal{H}$ will denote a Hilbert space.

Proposition 5.107: 5.21.

If $x_n \to x$ and $y_n \to Y$, then $\langle x_n|y_n\rangle \to \langle x|y\rangle$.

Proof. By the Schwarz inequality,

$$\begin{aligned} |\langle x_n|y_n\rangle - \langle x|y\rangle| &= |\langle x_n - x|y_n\rangle + \langle x|y_n - y\rangle| \\ &\leqslant \|x_n - x\|\|y_n\| + \|x\|\|y_n - y\|, \end{aligned}$$

which tends to zero since $\|y_n\| \to \|y\|$. □

Proposition 5.108: 5.22: The Parallelogram Law.

For all $x, y \in \mathcal{H}$,

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

("The sum of the squares of the diagonals of a parallelogram is the sum of the squares of the four sides.")

Proof. Add the two formulas $\|x \pm y\|^2 = \|x\|^2 \pm 2\operatorname{Re}\langle x|y\rangle + \|y\|^2$. If $x, y \in X$, we say that x is orthogonal to y and write $x \perp y$ if $\langle x|y\rangle = 0$. If $E \subset \mathcal{H}$, we define

$$E^\perp = \{x \in \mathcal{H} \mid \langle x|y\rangle = 0 \text{ for all } y \in E\}.$$

□

It is immediate from Proposition 107 and the linearity of the inner product in its first argument that $E^\perp$ is a closed subspace of $\mathcal{H}$.

Theorem 5.109: 5.23: The Pythagorean Theorem.

If $x_1, \ldots, x_n \in \mathcal{H}$ and $x_j \perp x_k$ for $j \neq k$,
$$\left\|\sum_1^n x_j\right\|^2 = \sum_1^n \|x_j\|^2.$$

Proof. $\|\sum x_j\|^2 = \langle \sum x_j | \sum x_j \rangle = \sum_{j,k} \langle x_j | x_k \rangle$. The terms with $k \neq j$ are all zero, leaving only $\sum \langle x_j | x_j \rangle = \sum \|x_j\|^2$. □

Theorem 5.110: 5.24.

If $\mathcal{M}$ is a closed subspace of $\mathcal{H}$, then $\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^\perp$; that is, each $x \in \mathcal{H}$ can be expressed uniquely as $x = y + z$ where $y \in \mathcal{M}$ and $z \in \mathcal{M}\perp$. Moreover, y and z are the unique elements of $\mathcal{M}$ and $\mathcal{M}^\perp$ whose distance to x is minimal.

Proof. Given $x \in \mathcal{H}$, let $\delta = \inf\{\|x - y\| \mid y \in \mathcal{M}\}$, and let $\{y_n\}$ be a sequence in $\mathcal{M}$ such that $\|x - y_n\| \to \delta$. By the paralellogram law,
$$2(\|y_n - x\|^2 + \|y_m - x\|^2) = \|y_n - y_m\|^2 + \|y_n + y_m - 2x\|^2,$$
so since $\frac{1}{2}(y_n + y_m) \in \mathcal{M}$,
$$\begin{aligned}\|y_n - y_m\|^2 &= 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\left\|\frac{1}{2}(y_n + y_m) - x\right\|^2 \\ &\leqslant 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\delta^2.\end{aligned}$$
As $m, n \to \infty$ this last quantity tends to zero, so $\{y_n\}$ is a Cauchy sequence. Let $y = \lim y_n$ and $z = x - y$. Then $y \in \mathcal{M}$ since $\mathcal{M}$ is closed, and $\|x - y\| = \delta$.

We claim that $z \in \mathcal{M}^\perp$. Indeed, if $u \in \mathcal{M}$, after multiplying u by a nonzero scalar we may assume that $\langle z | u \rangle$ is real. Then the function
$$f(t) = \|z + tu\|^2 = \|z\|^2 + 2t\langle z|u\rangle + t^2\|u\|^2$$
is real for $t \in \mathbb{R}$, and is has a minimum (namely, δ^2) at $t = 0$ because $z + tu = x - (y - tu)$ and $y - tu \in \mathcal{M}$. Thus $2\langle z|u\rangle = f'(0) = 0$, so $z \in \mathcal{M}^\perp$. Moreover, if z' is another element of $\mathcal{M}^\perp$, by the Pythagorean theorem (since $x - z = y \in \mathcal{M}$) we have
$$\|x - z'\|^2 = \|x - z\|^2 + \|z - z'\|^2 \geqslant \|x - z\|^2,$$
with equality if and only if $z = z'$. The same reasoning shows that y is the unique element of $\mathcal{M}$ closest to x.

Finally, if $x = y' + z'$ with $y' \in \mathcal{M}$ and $z' \in \mathcal{M}^\perp$, then $y - y' = z' - z \in \mathcal{M} \cap \mathcal{M}\mathcal{M}^\perp$, so $y - y'$ and $z' - z$ are orthogonal to themselves and hence are zero. □

If $y \in \mathcal{H}$, the Schwarz inequality shows that the formula $f_y(x) = \langle x|y\rangle$ defines a bounded linear functional on $\mathcal{H}$ such that $\|f_y\| = \|y\|$. Thus, the map $y \to f_y$ is a conjugate-linear isometry of $\mathcal{H}$ into $\mathcal{K}^*$. It is a fundamental fact that this map is surjective:

Theorem 5.111: 5.25.

If $f \in \mathcal{H}^*$, there is a unique $y \in \mathcal{H}$ such that $f(x) = \langle x|y\rangle$ for all $x \in X$.

Proof. Uniqueness is easy: If $\langle x|y\rangle = \langle x|y'\rangle$ for all x, by taking $x = y - y'$ we conclude that $\|y - y'\|^2 = 0$ and hence $y = y'$. If f is the zero functional, then obviously $y = 0$. Otherwise, let $\mathcal{M} = \{x \in \mathcal{H} \mid f(x) = 0\}$. Then $\mathcal{M}$ is a proper closed subspace of X, so $\mathcal{M}^\perp \neq \{0\}$ by Theorem 110. Pick $z \in \mathcal{M}^\perp$ with $\|z\| = 1$. If $u = f(x)z - f(z)x$ then $u \in \mathcal{M}$, so

$$0 = \langle u|z\rangle = f(x)\|z\|^2 - f(z)\langle x|z\rangle = f(x) - \left\langle x \middle| \overline{f(z)}z \right\rangle.$$

Hence $f(x) = \langle x|y\rangle$ where $y = \overline{f(z)}z$. □

Thus, Hilbert spaces are reflexive in a very strong sense: Not only is $\mathcal{H}$ naturally isomorphic to $\mathcal{H}^{**}$, it is naturally isomorphic (via a conjugate-linear map) to $\mathcal{H}^*$.

A subset $\{u_\alpha\}_{\alpha\in A}$ of $\mathcal{H}$ is called orthonormal if $\|u_\alpha\| = 1$ for all α and $u_\alpha \perp u_\beta$ whenever $\alpha \neq \beta$. If $\{x_n\}_1^\infty$ is a linearly independent sequence in $\mathcal{H}$, there is a standard inductive procedure, called the Gram-Schmidt process, for converting $\{x_n\}$ into an orthonormal sequence $\{u_n\}$ such that the linear span of $\{x_n\}_1^N$ coincides with the linear span of $\{u_n\}_1^N$ for all N. Namely, the first step is to set $u_1 = x_1/\|x_1\|$. Having defined $u_1, \dots, u_{N-1}$, we set $v_N = x_N - \sum_1^{N-1}\langle x_N|u_n\rangle u_n$. Then v_N is nonzero because x_N is not in the linear span of $x_1, \dots, x_{N-1}$ and hence of $u_1, \dots, u_{N-1}$, and $\langle v_N|u_m\rangle = \langle x_N|u_m\rangle - \langle x_N|u_m\rangle = 0$ for all $m < N$. We can therefore take $u_N = v_N/\|v_N\|$.

Theorem 5.112: 5.26: Bessel's Inequality.

If $\{u_\alpha\}_{\alpha\in A}$ is an orthonormal set in $\mathcal{H}$, then for any $x \in \mathcal{H}$

$$\sum_{\alpha\in A}|\langle x|u_\alpha\rangle|^2 \leqslant \|x\|^2.$$

In particular, $\{\alpha \mid \langle x|u_\alpha\rangle \neq 0\}$ is countable.

Proof. It suffices to show that $\sum_{\alpha\in F}|\langle x|u_\alpha\rangle|^2 \leqslant \|x\|^2$ for any finite $F \subset A$. But

$$\begin{aligned}
0 &\leqslant \left\|x - \sum_{\alpha\in F}\langle x|u_\alpha\rangle u_\alpha\right\|^2 \\
&= \|x\|^2 - 2\,\mathrm{Re}\left\langle x \middle| \sum_{\alpha\in F}\langle x|u_\alpha\rangle u_\alpha \right\rangle + \left\|\sum_{\alpha\in F}\langle x|u_\alpha\rangle u_\alpha\right\|^2 \\
&= \|x\|^2 - 2\sum_{\alpha\in F}|\langle x|u_\alpha\rangle|^2 + \sum_{\alpha\in F}|\langle x|u_\alpha\rangle|^2 \\
&= \|x\|^2 - \sum_{\alpha\in F}|\langle x|u_\alpha\rangle|^2,
\end{aligned}$$

where the Pythagorean theorem was used in the third line. □

Theorem 5.113: 5.27.

If $\{u_\alpha\}_{\alpha \in A}$ is an orthonormal set in $\mathcal{H}$, the following are equivalent:

(a) (Completeness) If $\langle x|u_\alpha\rangle = 0$ for all α, then $x = 0$.

(b) (Parseval's Identity) $\|x\|^2 = \sum_{\alpha\in A}|\langle x|u_\alpha\rangle|^2$ for all $x \in \mathcal{H}$.

(c) For each $x \in \mathcal{H}, x = \sum_{\alpha\in A}\langle x|u_\alpha\rangle u_\alpha$, where the sum on the right has only countably many nonzero terms and converges in the norm topology no matter how these terms are ordered.

Proof. (a) implies (c): If $x \in \mathcal{H}$, let $\alpha_1, \alpha_2, \ldots$ be any enumeration of the αs for which $\langle x|u_\alpha\rangle \neq 0$. By Bessel's inequality the series $\sum|\langle x|u_{\alpha_j}\rangle|^2$ converges, so by the Pythagorean theorem,

$$\left\|\sum_n^m \langle x|u_{\alpha_j}\rangle u_{\alpha_j}\right\|^2 = \sum_n^m |\langle x|u_{\alpha_j}\rangle|^2 \to 0 \text{ as } m, n \to \infty.$$

The series $\sum\langle x|u_{\alpha_j}\rangle u_{\alpha'}$ therefore converges since $\mathcal{H}$ is complete. If $y = x - \sum\langle x|u_{\alpha_j}\rangle u_{\alpha'}$, then clearly $\langle y|u_\alpha\rangle = 0$ for all α, so by (a), $y = 0$.

(c) implies (b): With notation as above, as in the proof of Bessel's inequality we have

$$\|x\|^2 - \sum_1^n |\langle x|u_{\alpha'}\rangle|^2 = \left\|x - \sum_1^n \langle x|u_{\alpha_j}\rangle u_{\alpha_2}\right\|^2 \to 0 \text{ as } n \to \infty.$$

Finally, that (b) implies (a) is obvious. □

An orthonormal set having the properties $(a - c)$ in Theorem 113 is called an ***orthonormal basis*** for $\mathcal{H}$. For example, let $\mathcal{H} = \ell^2(A)$. For each $\alpha \in A$, define $e_\alpha \in \ell^2(A)$ by $e_\alpha(\beta) = 1$ if $\beta = \alpha, e_\alpha(\beta) = 0$ otherwise. The set $\{e_\alpha\}_{\alpha\in A}$ is clearly orthonormal, and for any $f \in \ell^2(A)$ we have $\langle f|e_\alpha\rangle = f(\alpha)$, from which it follows that $\{e_\alpha\}$ is an orthonormal basis.

Proposition 5.114: 5.28.

Every Hilbert space has an orthonormal basis.

Proof. A routine application of Zorn's lemma shows that the collection of orthonormal sets, ordered by inclusion, has a maximal element; and maximality is equivalent to property (a) in Theorem 113. □

Proposition 5.115: 5.29.

A Hilbert space $\mathcal{H}$ is separable if and only if it has a countable orthonormal basis, in which case every orthonormal basis for $\mathcal{H}$ is countable.

Proof. If $\{x_n\}$ is a countable dense set in $\mathcal{H}$, by discarding recursively any x_n that is in the linear span of $x_1, \ldots, x_{n-1}$ we obtain a linearly independent sequence $\{y_n\}$ whose linear span is dense in $\mathcal{H}$. Application of the Gram-Schmidt process to $\{y_n\}$ yields an orthonormal sequence $\{u_n\}$ whose linear span is dense in $\mathcal{H}$ and which is therefore a basis. Conversely, if $\{u_n\}$ is a countable orthonormal basis, the finite linear combinations of the u_ns with coefficients in a countable dense subset of $\mathbb{C}$ form a countable dense set in $\mathcal{H}$. Moreover, if $\{v_\alpha\}_{\alpha\in A}$ is another orthonormal basis, for each n the set $A_n = \{\alpha \in A \mid \langle u_n|v_\alpha\rangle \neq 0\}$ is countable. By completeness of $\{u_n\}$, $A = \bigcup_1^\infty A_n$, so A is countable. □

Most Hilbert spaces that arise in practice are separable. We discuss some examples in Folland Exercise 5.60,Folland Exercise 5.61,Folland Exercise 5.62.

Definition 116. *If $\mathcal{H}_1$ and $\mathcal{H}_2$ are Hilbert spaces with inner products $\langle\cdot|\rangle_1$ and $\langle\cdot|\rangle_2$, a* ***unitary map*** *from $\mathcal{H}_1$ to $\mathcal{H}_2$ is an invertible linear map $U\colon \mathcal{H}_1 \to \mathcal{H}_2$ that preserves inner products:*

$$\langle Ux|Uy\rangle_2 = \langle x|y\rangle_1 \text{ for all } x, y \in \mathcal{K}_1.$$

By taking $y = x$, we see that every unitary map is an isometry: $\|Ux\|_2 = \|x\|_1$. Conversely, every surjective isometry is unitary (Folland Exercise 5.55). Unitary maps are the true "isomorphisms" in the category of Hilbert spaces; they preserve not only the linear structure and the topology but also the norm and the inner product. From the point of view of this abstract structure, every Hilbert space looks like an ℓ^2 space:

Proposition 5.117: 5.30.

Let $\{e_\alpha\}_{\alpha\in A}$ be an orthonormal basis for x. Then the correspondence $x \mapsto \widehat{x}$ defined by $\widehat{x}(\alpha) = \langle x|u_\alpha\rangle$ is a unitary map from $\mathcal{H}$ to $^2(A)$.

Proof. The map $x \mapsto \widehat{x}$ is clearly linear, and it is an isometry from $\mathcal{H}$ to $\ell^2(A)$ by the Parietal identity $\|x\|^2 = \sum |\widehat{x}(\alpha)|^2$. If $f \in \ell^2(A)$ then $\sum |f(\alpha)|^2 < \infty$, so the Pythagorean theorem shows that the partial sums of the series $\sum f(\alpha)u_\alpha$ (of which only countably many terms are nonzero) are Cauchy; hence $x = \sum f(\alpha)u_\alpha$ exists in $\mathcal{H}$ and $\widehat{x} = f$. By Folland Exercise 5.55, b, $x \mapsto \widehat{x}$ is unitary. □

Exercise 5.118: Folland Exercise 5.54.

For any nonempty set A, $\ell^2(A)$ is complete.

Exercise 5.119: Folland Exercise 5.55.

Let $\mathcal{H}$ be a Hilbert space.

(a) (The polarization identity) For any $x, y \in \mathcal{H}$,

$$\langle x|y\rangle = \frac{1}{4}(\|x+y\|^2 + \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2).$$

(Completeness is not needed here.)

(b) If $\mathcal{H}'$ is another Hilbert space, a linear map from $\mathcal{H}$ to $\mathcal{H}'$ is unitary if and only if it is isometric and surjective.

Solution. For (a), we have:

$$\begin{aligned}&\frac{1}{4}(\|x+y\|^2+\|x-y\|^2+i\|x+iy\|^2-i\|x-iy\|^2)\\&=\frac{1}{4}(2\langle x|y\rangle+2\langle y|x\rangle)+\frac{i}{4}(2\langle x|iy\rangle+2\langle iy|x\rangle)\\&=\frac{1}{2}(\langle x|y\rangle+\overline{\langle x|y\rangle})+\frac{i}{2}(\langle x|iy\rangle+\overline{\langle x|iy\rangle})\\&=\operatorname{Re}(\langle x|y\rangle)+i\operatorname{Im}(\langle x|y\rangle)=\langle x|y\rangle.\end{aligned}$$

For (b), let $U\colon H\to H'$ be unitary. Then U is surjective, and for each $x\in H$, we have $\|Ux\|^2=\langle Ux|Ux\rangle=\langle x|x\rangle=\|x\|^2$, meaning U is an isometry. Then we have by (a):

$$\begin{aligned}\langle Ux|Uy\rangle&=\frac{1}{4}(\|Ux+Uy\|^2-\|Ux-Uy\|^2+i\|Ux+iUy\|^2-i\|Ux-iUy\|^2)\\&=\frac{1}{2}(\langle x|y\rangle+\overline{\langle x|y\rangle})+\frac{i}{2}(\langle x|iy\rangle+\overline{\langle x|iy\rangle})\\&=\operatorname{Re}(\langle x|y\rangle)+i\operatorname{Im}(\langle x|y\rangle)=\langle x|y\rangle,\end{aligned}$$

completing the proof. □

Exercise 5.120: Folland Exercise 5.56.

If E is a subset of a Hilbert space $\mathcal{H}$, $(E^\perp)^\perp$ is the smallest closed subspace of $\mathcal{H}$ containing E.

Exercise 5.121: Folland Exercise 5.57.

Suppose that $\mathcal{H}$ is a Hilbert space and $T\in L(\mathcal{H},\mathcal{H})$.

(a) There is a unique $T^*\in L(\mathcal{H},\mathcal{H})$, called the adjoint of T, such that $\langle Tx|y\rangle=\langle x|T^*y\rangle$ for all $x,y\in\mathcal{H}$. (See Folland Exercise 5.22). We have $T^*=V^{-1}T^\dagger V$ where V is the conjugate-linear isomorphism from $\mathcal{H}$ to $\mathcal{K}^*$ in Theorem 111, $(Vy)(x)=\langle x|y\rangle$.)

(b) $\|T^*\|=\|T\|$, $\|T^*T\|=\|T\|^2$, $(aS+bT)^*=\overline{a}S^*+\overline{b}T^*$, $(ST)^*=T^*S^*$, and $T^{**}=T$.

(c) Let $\mathcal{R}$ and $\mathcal{N}$ denote range and nullspace; then $\mathcal{R}(T)^\perp=\mathcal{N}(T^*)$ and $\mathcal{N}(T)^\perp=\overline{\mathcal{R}(T^*)}$.

(d) T is unitary if and only if T is invertible and $T^{-1}=T^*$.

Solution.

(a) Define $T^*\colon H^*\to H^*$ by $T^*\coloneqq oloneqoloneqV^{-1}\circ T^\dagger\circ V$, where $V\colon H\to H^*$ sends $y\in H$ to $\langle -|y\rangle\in H^*$ and $T^\dagger\colon H^*\to H^*$ sends f to $f\circ T$. And $V\in L(H,H^*)$ (since for all $\|x\|=1$, $\|V_y(x)\|=\|\langle x|y\rangle\|\leqslant\|x\|\|y\|=\|y\|$ by the Cauchy-Schwarz inequality.)

Thus $\|V\| \leqslant 1$, hence V is bounded. And V is invertible by Theorem 111 (and Folland's subsequent remark), and $T^\dagger$ is bounded by Folland Exercise 5.22(a) so the composition $T^* = V^{-1} \circ T^\dagger \circ V$ is bounded. Moreover,

$$T^* y = V^{-1} \circ T^* \circ V(y) = V^{-1} \circ T^\dagger(\langle - | y \rangle) = V^{-1}(\langle - | y \rangle \circ T) = V^{-1}(\langle T(-) | y \rangle).$$

By definition of V^{-1}, $V^{-1}(\langle T(-)|y\rangle)$ is the element $z \in H$ such that $\langle x|z\rangle = \langle Tx|y\rangle$ for all $x \in H$. Hence, for all $x \in H$, we have

$$\langle x|T^* y\rangle = \langle x|z\rangle = \langle Tx|y\rangle,$$

as claimed.

To see T^* is unique, note that if we also had some S such that

$$\langle x|Sy\rangle = \langle Tx|y\rangle = \langle x|Sy\rangle$$

for all $x, y \in H$, then $\langle x|(T^* - S)y\rangle = 0$ for all $x, y \in H$. Since V^{-1} is an isomorphism then we have $\|(T^* - S)y\| = 0$ for all $y \in H$, so by passing to the supremum we conclude $\|T^* - S\| = 0$. Since the operator norm is a norm, we conclude $T^* - S = 0$, that is, $T^* = S$. Thus T^* is unique.

(b) Since for any $x, y \in H$ we have

$$\langle T^* x|y\rangle = \overline{\langle y|T^* x\rangle} = \overline{\langle Ty|x\rangle} = \langle x|Ty\rangle,$$

so by uniqueness from part (a) we obtain $T = T^{**}$. Again by uniqueness and the fact

$$(ST)^* = V^{-1}(ST)^\dagger V = V^{-1} T^\dagger S^\dagger V = T^* S^*,$$

we conclude $(ST)^* = T^* S^*$.

For any $x \in H$,

$$\|Tx\|^2 = \langle Tx|Tx\rangle = \langle x|T^* Tx\rangle \leqslant \|x\|\|T^* Tx\| \leqslant \|x\|\|T^*\|\|T\|,$$

which implies $\|Tx\| \leqslant \|T^*\|\|x\|$. Since $\|Tx\| = \inf\{C \mid \|Tx\| \leqslant C\|x\| \text{ for all } x \in X\}$, it follows that $\|T\| \leqslant \|T^*\|$. This reasoning is symmetric in T and T^*, so we similarly obtain $\|T^*\| \leqslant \|T^{**}\| = \|T\|$. Thus $\|T^*\| = \|T\|$.

Next, to see $\|T^* T\| = \|T\|^2$, note that $\|T^* T\| \leqslant \|T^*\|\|T\| = \|T\|^2$, and conversely we have $\|Tx\|^2 = \langle Tx|Tx\rangle = \langle x|T^* Tx\rangle \leqslant \|T^* T\|\|x\|$, so $\|T^* T\| = \|T\|^2$.

Lastly, for $a, b \in \mathbb{C}$ and $S, T \in L(H, H)$, observe that

$$\begin{aligned}(aS + bT)^* &= V^{-1}(aS + bT)^\dagger V = V^{-1}(aS^\dagger + bT^\dagger)V \\ &= \bar{a} V^{-1} S^\dagger V + \bar{b} V^{-1} T^\dagger V = \bar{a} S^* + \bar{b} T^*.\end{aligned}$$

(c) Note that $x \in H$ satisfies $\langle y|x\rangle = 0$ for all $y \in R(T)$ if and only if $\langle Ty|x\rangle = 0$ for all $y \in H$ if and only if $\langle y|T^* x\rangle = 0$ for all $y \in H$ if and only if $T^* x = 0$. Therefore, $R(T)^\perp = \{x \in H \mid \langle y|x\rangle = 0 \text{ for all } y \in H\} = \{x \in H \mid T^* x = 0\} = N(T^*)$. From this, we deduce that $N(T)^\perp = N(T^{**})^\perp = (R(T^*)^\perp)^\perp$ is the smallest (closed) linear subspace of H containing $R(T^*)$, which means $R(T^*) \subset N(T)^\perp$, forcing $\overline{R(T^*)} = N(T)^\perp$ as $\overline{R(T^*)}$ is itself a subspace.

(d) Let T be unitary. Then for any $x, y \in H$, we have
$$\langle Tx|y\rangle = \langle Tx|TT^{-1}y\rangle = \langle x|T^{-1}y\rangle,$$
so by uniqueness of T^* from part (a) $T^{-1} = T^*$. Conversely, if $T^{-1} = T^*$, then for any $x, y \in H$,
$$\langle Tx|Ty\rangle = \langle x|T^*Ty\rangle = \langle x|T^{-1}Ty\rangle = \langle x|y\rangle,$$
so T is unitary. □

Exercise 5.122: Folland Exercise 5.58.

Let $\mathcal{M}$ be a closed subspace of the Hilbert space $\mathcal{H}$, and for $x \in \mathcal{H}$ let Px be the element of $\mathcal{M}$ such that $x - Px \in \mathcal{M}^\perp$ as in Theorem 110.

(a) $P \in L(\mathcal{H}, \mathcal{H})$, and in the notation of Folland Exercise 5.57 we have $P^* = P, P^2 = P$, $\mathcal{R}(P) = \mathcal{M}$, and $\mathcal{N}(P) = \mathcal{M}^\perp . P$ is called the orthogonal projection onto $\mathcal{M}$.
(b) Conversely, suppose that $P \in L(\mathcal{H}, \mathcal{H})$ satisfies $P^2 = P^* = P$. Then $\mathcal{R}(P)$ is closed and P is the orthogonal projection onto $\mathcal{R}(P)$.
(c) If $\{u_\alpha\}$ is an orthonormal basis for $\mathcal{M}$, then $Px = \sum\langle x|u_\alpha\rangle u_\alpha$.

Exercise 5.123: Folland Exercise 5.59.

Every closed convex set K in a Hilbert space has a unique element of minimal norm. (If $0 \in K$, the result is trivial; otherwise, adapt the proof of Theorem 110.)

Solution. Set $\delta \coloneqq \inf_{y\in K} \|y\|$. Pick a sequence $\{y_n\} \subset K$ such that $\|y_n\| \to \delta$ as $n \to \infty$. We want to show that the limit is in K and is the unique element with norm δ. For all n, m, we can use the parallelogram law to write:
$$2\|y_n^2 - x\|^2 + 2\|y_m - x\|^2 = \|y_n + y_m - 2x\|^2 + \|y_n - y_m\|^2.$$
Note that $\frac{1}{2}(y_n + y_m) \in K$ since K is convex and $\frac{1}{2}(y_n + y_m) = ty_n + (1-t)y_m$ for $t = \frac{1}{2}$. It follows that $\left\|\frac{y_n+y_m}{2} - x\right\|^2$ is the square of the distance from x to something in y, and since δ is the infimum over all such distances, we conclude that this is $\|y_n + y_m - 2x\|^2 = 4\left\|\frac{y_n+y_m}{2} - x\right\|^2 \geqslant 4\delta^2$. It follows that
$$\|y_n - y_m\|^2 \leqslant 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - - - 4\delta^2 \to 0$$
as $m, n \to \infty$. Thus, $\{y_n\}$ is Cauchy, so since K is a complete vector space (as it is a closed subspace of the complete vector space H), we have that y_n converges to some $y_0 \in K$. Then, by the continuity of the norm, it follows that:
$$\|y_0\| = \left\|\lim_{n\to\infty} y_n\right\| = \lim_{n\to\infty} \|y_n\| = \delta,$$
where the last equality follows from the definition of δ.

We now have that the limit y_0 is in K and that $\|y_0\| = \delta$. It remains to show that any element in K with norm δ is identical to y_0.

Let $\delta = \|x\| = \|y\|$, $x, y \in K$. Since K is convex, $\frac{1}{2}x + \frac{1}{2}y \in K$. Then, observe that:

$$\begin{aligned}\delta &\leqslant \left\|\frac{1}{2}x + \frac{1}{2}y\right\} && \text{(since } \delta \text{ is the minimum distance)}\\ &\leqslant \frac{1}{2}\|x\| + \frac{1}{2}\|y\| && \text{(triangle inequality and scaling)}\\ &= \frac{1}{2}\delta + \frac{1}{2}\delta = \delta,\end{aligned}$$

so all the inequalities are equalities. Then, $\|x + y\| = \|x\| + \|y\| = 2\delta$. Using this, we have, by the parallelogram law, that:

$$4\delta^2 + \|x - y\|^2 = 2\delta^2 + 2\delta^2,$$

so $\|x - y\| = 0$, forcing $x = y$ since $\| - \|$ is a norm. □

Exercise 5.124: Folland Exercise 5.60.

Let $(X, \mathcal{M}, \mu)$ be a measure space. If $E \in \mathcal{M}$, we identify $L^2(E, \mu)$ with the subspace of $L^2(X, \mu)$ consisting of functions that vanish outside E. If $\{E_n\}$ is a disjoint sequence in $\mathcal{M}$ with $X = \bigcup_1^\infty E_n$, then $\{L^2(E_n, \mu)\}$ is a sequence of mutually orthogonal subspaces of $L^2(X, \mu)$, and every $f \in L^2(X, \mu)$ can be written uniquely as $f = \sum_1^\infty f_n$ (the series converging in norm) where $f_n \in L^2(E_n, \mu)$. If $L^2(E_n, \mu)$ is separable for every n, so is $L^2(X, \mu)$.

Exercise 5.125: Folland Exercise 5.61.

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces such that $L^2(\mu)$ and $L^2(\nu)$ are separable. If $\{f_m\}$ and $\{g_n\}$ are orthonormal bases for $L^2(\mu)$ and $L^2(\nu)$ and $h_{mn}(x, y) = f_m(x)g_n(y)$, then $\{h_{mn}\}$ is an orthonormal basis for $L^2(\mu \times \nu)$.

Exercise 5.126: Folland Exercise 5.62.

In this exercise the measure defining the L^2 spaces is Lebesgue measure.

(a) $C([0, 1])$ is dense in $L^2([0, 1])$. (Adapt the proof of Theorem 49.)
(b) The set of polynomials is dense in $L^2([0, 1])$.
(c) $L^2([0, 1])$ is separable.
(d) $L^2(\mathbb{R})$ is separable. (Use Folland Exercise 5.60.)
(e) $L^2(\mathbb{R}^n)$ is separable. (Use Folland Exercise 5.60.)

Solution.

(a) Fix $f \in L^2([0, 1])$. Let $\{\phi_n\}$ be simple functions such that $|\phi_m| \leqslant |\phi_n| \leqslant f$ for $m \leqslant n$, and $\phi_n \nearrow f$ pointwise. We have $|\phi_n - f|^2 \leqslant (2|f|)^2 = 4|f|^2 \in L^1([0, 1])$. Then by the dominated convergence theoremm

$$\lim_{n\to\infty}\int|\phi_n-f|^2=\int\lim_{n\to\infty}|\phi_n-f|^2<\varepsilon^2 m([0,1])<+\infty$$

for large enough n, so $\int|\phi_n-f|^2=\|\phi_n-f\|_2^2<\varepsilon$ for sufficiently large n.

If $I_k=(a,b)$, then we can approximate χ_{I_k} in the $L^1([0,1])$ metric by continuous functions that vanish outside (a,b). Indeed, given $\varepsilon>0$, take g to be the continuous function with $g=0$ on $[-1,a)\cup[b,1]$ and $g=1$ on $[a-\varepsilon,b+\varepsilon]$, and is linear on $[a,a+\varepsilon]$ and $[b-\varepsilon,b]$.

(b) Fix $f\in L^1([0,1])$. By part (a), there exists $g\in C([0,1])$ with $|f-g|<\sqrt{\varepsilon}/2$.

By the classical Stone-Weierstrass theorem, polynomials are dense in $C([0,1])$. Therefore, pick a polynomial $p\in C([0,1])$ such that $|g-p|<\sqrt{\varepsilon}/2$. Then, $|f-p|\leqslant|f-g|+|g-p|<\sqrt{\varepsilon}/2+\sqrt{\varepsilon}/2=\sqrt{\varepsilon}$.

Thus, $\|f-g\|_2^2=\int_{[0,1]}|f-p|^2\,dm<(\sqrt{\varepsilon})^2m([0,1])=\varepsilon$, as desired.

(c) The set of polynomials in $[0,1]$ with rational coefficients is countable (by the proof of the classical Stone-Weierstrass where such polynomials are used), so let the countable dense subset Z be the set of all rational valued polynomials on $[0,1]$.

(d) Let (X,μ) be a measure space, and for any μ-measurable E, we identify $L^2(E,\mu)$ as the subspace of $L^2(X)$ consisting of functions that vanish outside of E.

From Folland Exercise 5.60, we know that if $X=\bigsqcup_{n=1}^{\infty}E_n$ and $L^2(E_n,\mu)$ is separable for each n, then so is $L^2(X,\mu)$. Taking $(X,\mu)=(\mathbb{R},m)$ and $E_n=[n-1,n]$, we obtain the desired result.

(e) The Hilbert space $L^2(\mathbb{R}^n)$ with the inner product $\langle f|g\rangle=\int_{\mathbb{R}^n}f\overline{g}\,dm$ has an orthonormal basis. Thus, $L^2(\mathbb{R}^n)$ is a direct sum of pairwise orthogonal spaces $L^2(\mathbb{R})$, each of which is separable, and each is over a σ-finite measure space.

It follows that the union of the countable dense subsets from each of these spaces is itself a countable subset. We can decompose any $f\in L^2(\mathbb{R}^n)$ into its mutually orthogonal components and choose the element $\phi_i<\varepsilon/n$ from the corresponding dense subset of the ith direct summand, invoking Folland Exercise 5.61. □

Exercise 5.127: Folland Exercise 5.63.

Let $\mathcal{H}$ be an infinite-dimensional Hilbert space.

(a) Every orthonormal sequence in $\mathcal{H}$ converges weakly to 0.

(b) The unit sphere $S=\{x\mid\|x\|=1\}$ is weakly dense in the unit ball $B=\{x\mid\|x\|\leqslant 1\}$. (In fact, every $x\in B$ is the weak limit of a sequence in S.)

Exercise 5.128: Folland Exercise 5.64.

(a) For $k\in\mathbb{N}$, define $L_k\in L(\mathcal{H},\mathcal{H})$ by $L_k(\sum_1^\infty a_nu_n)=\sum_k^\infty a_nu_{n-k}$. Then $L_k\to 0$ in the strong operator topology but not in the norm topology.

(b) For $k \in \mathbb{Z}_{\geqslant 1}$, define $R_k \in L(\mathcal{H}, \mathcal{H})$ by $R_k(\sum_1^\infty a_n u_n) = \sum_1^\infty a_n u_{n+k}$. Then $R_k \to 0$ in the weak operator topology but not in the strong operator topology. c. $R_k L_k \to 0$ in the strong operator topology, but $L_k R_k = I$ for all k. (Use Folland Exercise 5.53(b).)

Exercise 5.129: Folland Exercise 5.65.

$\ell^2(A)$ is unitarily isomorphic to $\ell^2(B)$ if and only if $\text{card}(A) = \text{card}(B)$.

Exercise 5.130: Folland Exercise 5.66.

Let $\mathcal{M}$ be a closed subspace of $L^2([0,1], m)$ that is contained in $C([0,1])$.

(a) There exists $C > 0$ such that $\|f\|_u \leqslant C\|f\|_{L^2}$ for all $f \in \mathcal{M}$. (Use the closed graph theorem.)

(b) For each $x \in [0,1]$ there exists $g_x \in \mathcal{M}$ such that $f(x) = \langle f | g_x \rangle$ for all $f \in \mathcal{M}$, and $\|g_x\|_{L^2} \leqslant C$.

(c) The dimension of $\mathcal{M}$ is at most C^2. (Hint: If $\{f_j\}$ is an orthonormal sequence in $\mathcal{M}$, $\sum |f_j(x)|^2 \leqslant C^2$ for all $x \in [0,1]$.)

Exercise 5.131: Folland Exercise 5.67: The Mean Ergodic Theorem.

Let U be a unitary operator on the Hilbert space $\mathcal{H}$, $\mathcal{M} = \{x \mid Ux = x\}$, P the orthogonal projection onto $\mathcal{M}$ (Folland Exercise 5.58), and $S_n = n^{-1} \sum_0^{n-1} U^j$. Then $S_n \to P$ in the strong operator topology. (If $x \in \mathcal{M}$, then $S_n x = x$; if $x = y - Uy$ for some y, then $S_n x \to 0$. By Folland Exercise 5.57(d), $\mathcal{M} = \{x \mid U^* x = x\}$. Apply Folland Exercise 5.57(c) with $T = I - U$.)

6 L^p Spaces

L^p spaces are a class of Banach spaces of functions whose norms are defined in terms of integrals and which generalize the L^1 spaces discussed in Chapter 2. They furnish interesting examples of the general theory of Chapter 5 and play a central role in modern analysis.

In this chapter we shall be working on a fixed measure space $(X, \mathcal{M}, \mu)$.

6.1 Basic Theory of L^p Spaces

Definition 1. *If f is a measurable function on X and $0 < p < \infty$, we define*

$$\|f\|_p \coloneqq \left(\int |f|^p d\mu\right)^{1/p},$$

allowing the possibility that $\|f\|_p = \infty$, and we set

$$L^p(X, \mathcal{M}, \mu) \coloneqq \{f : X \to \mathbb{C} \mid f \text{ is measurable and } \|f\|_p < \infty\}$$

We abbreviate $L^p(X, \mathcal{M}, \mu)$ by $L^p(\mu)$, $L^p(X)$, or simply L^p when this will cause no confusion. As we have done with L^1, we consider two functions to define the same element of L^p when they are equal almost everywhere. We sometimes denote by $L^0(X, \mathcal{M}, \mu)$ the set of equivalence classes of $\mathcal{M}$-measurable functions that are equal a.e., and this may also be denoted by $L^0(\mu)$ or even L^0.

Notation 2. *If A is any nonempty set, we define $\ell^p(A)$ to be $L^p(\mu)$ where μ is counting measure on $(A, \mathcal{P}(A))$, and we denote $\ell^p(\mathbb{Z}_{\geqslant 1})$ simply by ℓ^p.*

Lemma 6.3.

L^p is a vector space for any $p \in (0, \infty)$.

Proof. If $f, g \in L^p$, then

$$|f + g|^p \leqslant [2 \max(|f|, |g|)]^p \leqslant 2^p(|f|^p + |g|^p)$$

so that $f + g \in L^p$. □

Our notation suggests $\|-\|_p$ is a norm on L^p. Indeed, it is obvious that $\|f\|_p = 0$ if and only if $f = 0$ a.e. and $\|cf\|_p = |c|\|f\|_p$, so the only question is the triangle inequality. It turns out that the latter is valid precisely when $p \geqslant 1$, so our attention will be focused almost exclusively on this case.

Warning 6.4.

Before proceeding further, however, let us see why the triangle inequality fails for $p < 1$. Suppose $a > 0, b > 0$, and $0 < p < 1$. For $t > 0$ we have $t^{p-1} > (a + t)^{p-1}$, and by integrating from 0 to b we obtain $a^p + b^p > (a + b)^p$. Thus, if E and F are disjoint sets of positive finite measure in X and we set $a = \mu(E)^{1/p}$ and $b = \mu(F)^{1/p}$, we see that

$$\|\chi_E + \chi_F\|_p = (a^p + b^p)^{1/p} > a + b = \|\chi_E\|_p + \|\chi_F\|_p.$$

The cornerstone of the theory of L^p spaces is Hölder's inequality, which we now derive.

Lemma 6.5: 6.1.

If $a \geqslant 0, b \geqslant 0$, and $0 < \lambda < 1$, then

$$a^\lambda b^{1-\lambda} \leqslant \lambda a + (1-\lambda)b$$

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

Proof. The result is obvious if $b = 0$; otherwise, dividing both sides by b and setting $t = a/b$, we are reduced to showing that $t^\lambda \leqslant \lambda t + (1-\lambda)$ with equality if and only if $t = 1$. But by elementary calculus, $t^\lambda - \lambda t$ is strictly increasing for $t < 1$ and strictly decreasing for $t > 1$, so its maximum value, namely $1 - \lambda$, occurs at $t = 1$. □

Theorem 6.6: 6.2: Hölder's Inequality.

Suppose $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$ (that is, $q = p/(p-1)$). If f and g are measurable functions on X, then

$$\|fg\|_1 \leqslant \|f\|_p \|g\|_q \tag{6.6.1}$$

In particular, if $f \in L^p$ and $g \in L^q$, then $fg \in L^1$, and in this case equality holds in Equation (6.6.1) if and only if $\alpha|f|^p = \beta|g|^q$ a.e. for some constants α, β with $\alpha\beta \neq 0$.

Proof. The result is trivial if $\|f\|_p = 0$ or $\|g\|_q = 0$ (since then $f = 0$ or $g = 0$ a.e.), or if $\|f\|_p = \infty$ or $\|g\|_q = \infty$. Moreover, we observe that if Equation (6.6.1) holds for a particular f and g, then it also holds for all scalar multiples of f and g, for if f and g are replaced by af and bg, both sides of Equation (6.6.1) change by a factor of $|ab|$. It therefore suffices to prove that Equation (6.6.1) holds when $\|f\|_p = \|g\|_q = 1$ with equality if and only if $|f|^p = |g|^q$ a.e. To this end, we apply Lemma 5 with $a = |f(x)|^p, b = |g(x)|^q$, and $\lambda = p^{-1}$ to obtain

$$|f(x)g(x)| \leqslant p^{-1}|f(x)|^p + q^{-1}|g(x)|^q$$

Integration of both sides yields

$$\|fg\|_1 \leqslant p^{-1} \int |f|^p + q^{-1} \int |g|^q = p^{-1} + q^{-1} = 1 = \|f\|_p \|g\|_q$$

Equality holds here if and only if it holds a.e. in (6.4), and by Lemma 5 this happens precisely when $|f|^p = |g|^q$ a.e. □

The condition $p^{-1} + q^{-1} = 1$ occurring in Hölder's inequality turns up frequently in L^p theory. If $1 < p < \infty$, the number $q = p/(p-1)$ such that $p^{-1} + q^{-1} = 1$ is called the ***conjugate exponent*** to p.

Theorem 6.7: 6.5: Minkowski's Inequality.

If $1 \leqslant p < \infty$ and $f, g \in L^p$, then

$$\|f+g\|_p \leqslant \|f\|_p + \|g\|_p$$

Proof. The result is obvious if $p = 1$ or if $f + g = 0$ a.e. Otherwise, we observe that

$$|f+g|^p \leqslant (|f|+|g|)|f+g|^{p-1}$$

and apply Hölder's inequality, noting that $(p-1)q = p$ when q is the conjugate exponent to p:

$$\begin{aligned}\int |f+g|^p &\leqslant \|f\|_p \left\||f+g|^{p-1}\right\|_q + \|g\|_p \left\||f+g|^{p-1}\right\|_q \\ &= (\|f\|_p + \|g\|_p)\left(\int |f+g|^p\right)^{1/q}.\end{aligned}$$

Therefore,

$$\|f+g\|_p = \left[\int |f+g|^p\right]^{1-(1/q)} \leqslant \|f\|_p + \|g\|_p.$$

□

This result shows that, for $p \geqslant 1$, L^p is a normed vector space. The following theorem shows that even more is true.

Theorem 6.8: 6.6.

For $1 \leqslant p < \infty$, L^p is a Banach space.

Proof. We use Theorem 8. Suppose $\{f_k\} \subset L^p$ and $\sum_1^\infty \|f_k\|_p = B < \infty$. Let $G_n = \sum_1^n |f_k|$ and $G = \sum_1^\infty |f_k|$. Then $\|G_n\|_p \leqslant \sum_1^n \|f_k\|_p \leqslant B$ for all n, so by the monotone convergence theorem, $\int G^p = \lim \int G_n^p \leqslant B^p$. Hence $G \in L^p$, and in particular $G(x) < \infty$ a.e., which implies that the series $\sum_1^\infty f_k$ converges a.e. Denoting its sum by F, we have $|F| \leqslant G$ and hence $F \in L^p$; moreover, $|F - \sum_1^n f_k|^p \leqslant (2G)^p \in L^1$, so by the dominated convergence theorem,

$$\left\|F - \sum_1^n f_k\right\|_p^p = \int \left|F - \sum_1^n f_k\right|^p \to 0$$

Thus the series $\sum_1^\infty f_k$ converges in the L^p norm. □

Proposition 6.9: 6.7.

For $1 \leqslant p < \infty$, the set of simple functions $f = \sum_1^n a_j \chi_{E_j}$, where $\mu(E_j) < \infty$ for all j, is dense in L^p.

Proof. Clearly such functions are in L^p. If $f \in L^p$, choose a sequence $\{f_n\}$ of simple functions such that $f_n \to f$ a.e. and $|f_n| \leqslant |f|$, according to Theorem 18. Then $f_n \in L^p$

and $|f_n - f|^p \leqslant 2^p|f|^p \in L^1$, so by the dominated convergence theorem, $\|f_n - f\|_p \to 0$. Moreover, if $f_n = \sum a_j \chi_{E_j}$ where the E_j are disjoint and the a_j are nonzero, we must have $\mu(E_j) < \infty$ since $\sum |a_j|^p \mu(E_j) = \int |f_n|^p < \infty$. □

To complete the picture of L^p spaces, we introduce a space corresponding to the limiting value $p = \infty$.

Definition 10. *If f is a measurable function on X, we define*

$$\|f\|_\infty = \inf\{a \geqslant 0 \mid \mu(\{x \mid |f(x)| > a\}) = 0\}$$

with the convention that $\inf \varnothing = \infty$. $\|f\|_\infty$ is called the ***essential supremum*** *of $|f|$ and is sometimes written*

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

We now define

$$L^\infty(X, \mathcal{M}, \mu) = \{f \colon X \to \mathbb{C} \mid f \text{ is measurable and } \|f\|_\infty < \infty\}$$

with the same notational conventions of L^p and the usual convention that functions that are equal a.e. define the same element of L^∞. define

Note 11. *We observe that the infimum in Definition 10 is actually attained, for*

$$\{x \mid |f(x)| > a\} = \bigcup_1^\infty \{x \mid |f(x)| > a + n^{-1}\}$$

and if the sets on the right are null, so is the one on the left.

Thus $f \in L^\infty$ if and only if there is a bounded measurable function g such that $f = g$ a.e.; we can take $g = f\chi_E$ where $E = \{x \mid |f(x)| \leqslant \|f\|_\infty\}$.

Two remarks: First, for fixed X and $\mathcal{M}$, $L^\infty(X, \mathcal{M}, \mu)$ depends on μ only insofar as μ determines which sets have measure zero; if μ and ν are mutually absolutely continuous, then $L^\infty(\mu) = L^\infty(\nu)$. Second, if μ is not semifinite, for some purposes it is appropriate to adopt a slightly different definition of L^∞. This point will be explored in Folland Exercise 6.23, Folland Exercise 6.24, and Folland Exercise 6.25.

The results we have proved for $1 \leqslant p < \infty$ extend easily to the case $p = \infty$, as follows:

Theorem 6.12: 6.8.

- (a) If f and g are measurable functions on X, then $\|fg\|_1 \leqslant \|f\|_1 \|g\|_\infty$. If $f \in L^1$ and $g \in L^\infty$, $\|fg\|_1 = \|f\|_1 \|g\|_\infty$ if and only if $|g(x)| = \|g\|_\infty$ a.e. on the set where $f(x) \neq 0$.
- (b) $\|\cdot\|_\infty$ is a norm on L^∞.
- (c) $\|f_n - f\|_\infty \to 0$ if and only if there exists $E \in \mathcal{M}$ such that $\mu(E^c) = 0$ and $f_n \to f$ uniformly on E.
- (d) L^∞ is a Banach space.
- (e) The simple functions are dense in L^∞.

The proof is left to the reader (Folland Exercise 6.2).

Convention 6.13.

In view of Theorem 12(a) and the formal equality $1^{-1} + \infty^{-1} = 1$, it is natural to regard 1 and ∞ as conjugate exponents of each other, and we do so henceforth.

Theorem 12(c) shows that $\|\cdot\|_\infty$ is closely related to, but usually not identical with, the uniform norm $\|\cdot\|_u$. However, if we are dealing with Lebesgue measure, or more generally any Borel measure that assigns positive values to all open sets, then $\|f\|_\infty = \|f\|_u$ whenever f is continuous, since $\{x \mid |f(x)| > a\}$ is open. In this situation we may use the notations $\|f\|_\infty$ and $\|f\|_u$ interchangeably, and we may regard the space of bounded continuous functions as a (closed!) subspace of L^∞.

Note 14 (Very Important Note)**.** *In general we have $L^p \not\subset L^q$ for all $p \neq q$; to see what is at issue, it is instructive to consider the following simple examples on $(0, \infty)$ with Lebesgue measure. Let $f_a(x) = x^{-a}$, where $a > 0$. Elementary calculus shows that $f_a\chi_{(0,1)} \in L^p$ if and only if $p < a^{-1}$, and $f_a\chi_{(1,\infty)} \in L^p$ if and only if $p > a^{-1}$. Thus we see two reasons why a function f may fail to be in L^p: either $|f|^p$ blows up too rapidly near some point, or it fails to decay sufficiently rapidly at infinity. In the first situation the behavior of $|f|^p$ becomes worse as p increases, while in the second it becomes better. In other words, if $p < q$, functions in L^p can be locally more singular than functions in L^q, whereas functions in L^q can be globally more spread out than functions in L^p. These somewhat imprecisely expressed ideas are actually a rather accurate guide to the general situation, concerning which we now give four precise results. The last two show that inclusions $L^p \subset L^q$ can be obtained under conditions on the measure space that disallow one of the types of bad behavior described above; for a more general result, see Folland Exercise 6.5.*

Proposition 6.15: 6.9.

If $0 < p < q < r \leqslant \infty$, then $L^q \subset L^p + L^r$. That is, each $f \in L^q$ is the sum of a function in L^p and a function in L^r.

Proof. If $f \in L^q$, let $E = \{x \mid |f(x)| > 1\}$ and set $g = f\chi_E$ and $h = f\chi_{E^c}$. Then $|g|^p = |f|^p\chi_E \leqslant |f|^q\chi_E$, so $g \in L^p$, and $|h|^r = |f|^r\chi_{E^c} \leqslant |f|^q\chi_{E^c}$, so $h \in L^r$. (For $r = \infty$, obviously $\|h\|_\infty \leqslant 1$.) □

Proposition 6.16: 6.10.

If $0 < p < q < r \leqslant \infty$, then $L^p \cap L^r \subset L^q$ and $\|f\|_q \leqslant \|f\|_p^\lambda \|f\|_r^{1-\lambda}$, where $\lambda \in (0, 1)$ is defined by

$$q^{-1} = \lambda p^{-1} + (1 - \lambda)r^{-1}, \text{ that is, } \lambda = \frac{q^{-1} - r^{-1}}{p^{-1} - r^{-1}}$$

Proof. If $r = \infty$, we have $|f|^q \leqslant \|f\|_\infty^{q-p}|f|^p$ and $\lambda = p/q$, so

$$\|f\|_q \leqslant \|f\|_p^{p/q}\|f\|_\infty^{1-(p/q)} = \|f\|_p^\lambda\|f\|_\infty^{1-\lambda}.$$

If $r < \infty$, we use Hölder's inequality, taking the pair of conjugate exponents to be $p/\lambda q$ and $r/(1-\lambda)q$:

$$\int |f|^q = \int |f|^{\lambda q}|f|^{(1-\lambda)q} \leqslant \||f|^{\lambda q}\|_{p/\lambda q}\||f|^{(1-\lambda)q}\|_{r/(1-\lambda)q}$$
$$= \left[\int |f|^p\right]^{\lambda q/p}\left[\int |f|^r\right]^{(1-\lambda)q/r} = \|f\|_p^{\lambda q}\|f\|_r^{(1-\lambda)q}.$$

Taking qth roots, we are done. □

Proposition 6.17: 6.11.

If A is any set and $0 < p < q \leqslant \infty$, then $\ell^p(A) \subset \ell^q(A)$ and $\|f\|_q \leqslant \|f\|_p$.

Proof. Obviously $\|f\|_\infty^p = \sup_\alpha |f(\alpha)|^p \leqslant \sum_\alpha |f(\alpha)|^p$, so that $\|f\|_\infty \leqslant \|f\|_p$. The case $q < \infty$ then follows from Proposition 16: if $\lambda = p/q$,

$$\|f\|_q \leqslant \|f\|_p^\lambda \|f\|_\infty^{1-\lambda} \leqslant \|f\|_p.$$ □

Proposition 6.18: 6.12.

If $\mu(X) < \infty$ and $0 < p < q \leqslant \infty$, then $L^p(\mu) \supset L^q(\mu)$ and $\|f\|_p \leqslant \|f\|_q \mu(X)^{(1/p)-(1/q)}$.

Proof. If $q = \infty$, this is obvious:

$$\|f\|_p^p = \int |f|^p \leqslant \|f\|_\infty^p \int 1 = \|f\|_\infty^p \mu(X)$$

If $q < \infty$, we use Hölder's inequality with the conjugate exponents q/p and $q/(q-p)$:

$$\|f\|_p^p = \int |f|^p \cdot 1 \leqslant \||f|^p\|_{q/p}\|1\|_{q/(q-p)} = \|f\|_q^p \mu(X)^{(q-p)/q}.$$ □

We conclude this section with a few remarks about the significance of the L^p spaces. The three most obviously important ones are L^1, L^2, and L^∞. With L^1 we are already familiar; L^2 is special because it is a Hilbert space; and the topology on L^∞ is closely related to the topology of uniform convergence. Unfortunately, L^1 and L^∞ are pathological in many respects, and it is more fruitful to deal with the intermediate L^p spaces. One manifestation of this is the duality theory in Folland Section 6.2; another is the fact that many operators of interest in Fourier analysis and differential equations are bounded on L^p for $1 < p < \infty$ but not on L^1 or L^∞. (Some examples are mentioned in Folland Section 9.4.)

Exercise 6.19: Folland Exercise 6.1.

When does equality hold in Minkowski's inequality? (The answer is different for $p = 1$ and for $1 < p < \infty$. What about $p = \infty$?)

Exercise 6.20: Folland Exercise 6.2.

Prove Theorem 12.

Exercise 6.21: Folland Exercise 6.3.

If $1 \leqslant p < r \leqslant \infty, L^p \cap L^r$ is a Banach space with norm $\|f\| = \|f\|_p + \|f\|_r$, and if $p < q < r$, the inclusion map $L^p \cap L^r \to L^q$ is continuous.

Exercise 6.22: Folland Exercise 6.4.

If $1 \leqslant p < r \leqslant \infty, L^p + L^r$ is a Banach space with norm $\|f\| = \inf\{\|g\|_p + \|h\|_r \mid f = g + h\}$, and if $p < q < r$, the inclusion map $L^q \to L^p + L^r$ is continuous.

Exercise 6.23: Folland Exercise 6.5.

Suppose $0 < p < q < \infty$. Then $L^p \not\subset L^q$ if and only if X contains sets of arbitrarily small positive measure, and $L^q \not\subset L^p$ if and only if X contains sets of arbitrarily large finite measure.

(For the "if" implication: In the first case there is a disjoint sequence $\{E_n\}$ with $0 < \mu(E_n) < 2^{-n}$, and in the second case there is a disjoint sequence $\{E_n\}$ with $1 \leqslant \mu(E_n) < \infty$. Consider $f = \sum a_n \chi_{E_n}$ for suitable constants a_n.) What about the case $q = \infty$?

Exercise 6.24: Folland Exercise 6.6.

Suppose $0 < p_0 < p_1 \leqslant \infty$. Find examples of functions f on $(0, \infty)$ (with Lebesgue measure), such that $f \in L^p$ if and only if (a) $p_0 < p < p_1$, (b) $p_0 \leqslant p \leqslant p_1$, (c) $p = p_0$. (Consider functions of the form $f(x) = x^{-a}|\log x|^b$.)

Exercise 6.25: Folland Exercise 6.7.

If $f \in L^p \cap L^\infty$ for some $p < \infty$, so that $f \in L^q$ for all $q > p$, then $\|f\|_\infty = \lim_{q\to\infty} \|f\|_q$.

Solution. First suppose $\|f\|_p = 0$. Then $0 = \|f\|_p^p = \int |f|^p$, so $|f| = 0$ a.e. This means

$\|f\|_\infty = 0$ and $\|f\|_q = 0$ for all q, so

$$\|f\|_\infty = 0 = \lim_{q\to\infty} 0 = \lim_{q\to\infty} \|f\|_q,$$

which affirms the claim.

Now suppose $\|f\|_p > 0$. By Folland Proposition 6.10 with $r = \infty$, for all $q > 0$ and all $p \in (1, q)$ we have

$$\|f\|_q \leqslant \|f\|_p^{p/q} \|f\|_\infty^{1-p/q}.$$

Taking the limit at $q \to \infty$, we obtain

$$\lim_{q\to\infty} \|f\|_q \leqslant \|f\|_p^{p/q} \|f\|_\infty^{1-p/q} = \|f\|_p^0 \|f\|_\infty^{1-0} = \infty,$$

where we used that the map $q \mapsto \|f\|_p^q$ is continuous as a function of $q \in (0, \infty)$ (since $\|f\|_p$ is nonnegative).

To show the reverse inequality, it suffices to show $\liminf_{q\to\infty} \|f\|_q \leqslant \|f\|_\infty$. We can prove this as follows: Fix $n \in \mathbb{Z}_{\geqslant 1}$ and let

$$E_n := \{x \in X \mid |f| \geqslant \|f\|_\infty - 1/n\}.$$

Since $\mu(E_n) > 0$ (by definition of $\|-\|_\infty$), we have

$$\|f\|_q^q = \int |f|^q \geqslant \int_{E_n} |f|^q \geqslant \int_{E_n} (\|f\|_\infty - 1/n)^q = \mu(E_n)(\|f\|_\infty - 1/n)^q.$$

Taking the qth root of both sides, we obtain

$$\|f\|_q \geqslant \mu(E_n)^{1/q}(\|f\|_\infty - 1/n). \tag{6.25.1}$$

And $\mu(E_n) < \infty$, since otherwise $\infty = \mu(E_n)^{1/q}(\|f\|_\infty - 1/n) \leqslant \|f\|_q^q$, contradicting $f \in L^q$. Also $\mu(E_n) > 0$ (by definition of $\|-\|_\infty$), so by taking $q \to \infty$ we have by Equation (6.25.1) that

$$\lim_{q\to\infty} \|f\|_q \geqslant \mu(E_n)^0(\|f\|_\infty - 1/n) = \|f\|_\infty - 1/n.$$

Since n was arbitrary, we conclude $\lim_{q\to\infty} \|f\|_q \geqslant \|f\|_\infty$, which completes the proof. □

Exercise 6.26: Folland Exercise 6.8.

Suppose $\mu(X) = 1$ and $f \in L^p$ for some $p > 0$, so that $f \in L^q$ for $0 < q < p$.

(a) $\log \|f\|_q \geqslant \int \log|f|$. (Use Exercise 42 d in Folland Section 3.5, with $F(t) = e^t$.)
(b) $\left(\int |f|^q - 1\right)/q \geqslant \log \|f\|_q$, and $\left(\int |f|^q - 1\right)/q \to \int \log|f|$ as $q \to 0$.
(c) $\lim_{q\to 0} \|f\|_q = \exp\left(\int \log|f|\right)$.

Solution.

(a) Here we use the convention $\log(0) = -\infty$ and $\log\infty = \infty$. We may assume $\int \log|f| \neq -\infty$, since otherwise the desired inequality is

$$\log|f|^q = q \int \log|f| = -\infty \leqslant \log\|f\|_q,$$

which holds irregardless of the value of $\|f\|_q$. The exponential is convex and $\mu(X) = 1$, so by Jensen's inequality (Folland Exercise 3.42(d)), we obtain

$$\exp\left(\int \log|f|^q\right) \leqslant \int \exp(\log|f|^q) = \int |f|^q.$$

Taking the logarithm of both sides, we deduce

$$q\int \log|f| = \int \log|f|^q \leqslant \log\int |f|^q = \log\|f\|_q^q = q\log\|f\|_q.$$

By dividing through by $q > 0$, we conclude $\int \log|f| \leqslant \log\|f\|_q$

(b) Since $\log x \leqslant x - 1$ for all $x \in [0,\infty]$, we have

$$q\log\|f\|_q = \log\int |f|^q \leqslant \int |f|^q - 1.$$

Then divide through by $q > 0$ to obtain the desired inequality.

It remains to show $\left(\int |f|^q - 1\right)/q \to \int \log|f|$ as $q \searrow 0$. We have $\chi_{\{|f|\geqslant 1\}}\frac{|f|^q-1}{q} \leqslant \chi_{\{|f|\geqslant 1\}}\frac{|f|^p-1}{p} \in L^1$, so by the dominated convergence theorem

$$\lim_{q\searrow 0}\int \chi_{\{|f|\geqslant 1\}}\frac{|f|^q - 1}{q} = \int \lim_{q\searrow 0}\chi_{\{|f|\geqslant 1\}}\frac{|f(x)|^q - 1}{q} = \int \chi_{\{|f|\geqslant 1\}}\log|f|, \tag{6.26.1}$$

where for the second equality we used the limit definition of the logarithm on $[0,\infty]$. On the other hand, by the fundamental theorem of calculus, we have

$$\chi_{\{|f|<1\}}\frac{|f|^q - 1}{q} = \int_1^{|f|}\chi_{\{|f|<1\}}t^{q-1} = \int_{|f|}^1 \chi_{\{|f|<1\}}t^{q-1},$$

which increases as q decreases. As everything here is measurable, by the monotone convergence theorem

$$\lim_{q\searrow 0}\int \chi_{\{|f|<1\}}\frac{|f|^q - 1}{q} = \int \chi_{\{|f|<1\}}\log|f|. \tag{6.26.2}$$

Now by Equations (6.26.1) and (6.26.2), we conclude

$$\begin{aligned}\lim_{q\searrow 0}\int \frac{|f|^q - 1}{q} &= \lim_{q\searrow 0}\int (\chi_{\{|f|<1\}} + \chi_{\{|f|\geqslant 1\}})\frac{|f|^q - 1}{q}\\ &= \int \chi_{\{|f|<1\}}\log|f| + \int \chi_{\{|f|\geqslant 1\}}\log|f| = \int \log|f|,\end{aligned}$$

as claimed.

(c) We have

$$\exp\left(\int \log|f|\right) \leqslant \exp(\log\|f\|_q) \leqslant \exp\left(\int |f|^q - 1\right)/q,$$

where the first and second inequalities are by parts (a) and (b), respectively. By part (b) and continuity of the exponential,

$$\exp\left(\int |f|^q - 1\right)/q \to \int \log|f|$$

as $q \to 0$. Now by the squeeze theorem for limits, we conclude

$$\lim_{q\to 0}\|f\|_q = \exp\left(\int \log|f|\right). \qquad \square$$

Exercise 6.27: Folland Exercise 6.9.

Suppose $1 \leqslant p < \infty$. If $\|f_n - f\|_p \to 0$, then $f_n \to f$ in measure, and hence some subsequence converges to f a.e. On the other hand, if $f_n \to f$ in measure and $|f_n| \leqslant g \in L^p$ for all n, then $\|f_n - f\|_p \to 0$.

Exercise 6.28: Folland Exercise 6.10.

Suppose $1 \leqslant p < \infty$. If $f_n, f \in L^p$ and $f_n \to f$ a.e., then $\|f_n - f\|_p \to 0$ if and only if $\|f_n\|_p \to \|f\|_p$. (Use Folland Exercise 2.20.)

Solution. We also prove or disprove the assertion in the case $p = \infty$.

($\Rightarrow$) If $\varepsilon > 0$ and $\|f_n - f\|_p \to 0$, then by the triangle inequality $\|f_n\|_p - \|f\|_p \leqslant \|f_n - f\|_\infty < \varepsilon$ for all sufficiently large $n \in \mathbb{Z}_{\geqslant 1}$, so the forward implication holds. Note that this argument works for all $p \in [1, \infty]$.

($\Leftarrow$) Since $\|f_n\|_p \to \|f\|_p$, we have $\|f_n\|_p^p \to \|f\|_p^p$. Setting $g_n \coloneqq 2^p \max\{|f_n|^p, |f|^p\}$, $g \coloneqq 2^p|f|^p \geqslant 0$, $h_n \coloneqq 2^p|f_n - f|^p$, and $h \coloneqq 0$, we observe that

- $h_n \to h$ a.e.,
- $g_n \to g$ a.e.,
- $g_n \in L^1$ since $f_n, f \in L^p$ implies $|f_n|^p, |f|^p \in L^1$ (hence also $\max\{|f_n|^p, |f|^p\} \in L^p$),
- $h_n \in L^1$ since by the triangle inequality $h_n \leqslant 2^p \max\{|f|_n^p, |f|^p\} = g_n \in L^1$ and g_n,
- $|h_n| = |f_n - f|^p \leqslant (|f_n| + |f|)^p \leqslant 2\max\{|f_n|^p, |f|^p\} \leqslant 2^p \max\{|f_n|^p, |f|^p\} = g_n \in L^1$ (since $f_n, f \in L^p$, hence $|f_n|^p, |f|^p \in L^1$), and
- $\int g_n = 2^p \int \max\{|f_n|^p, |f|^p\} \to 2^p \int |f|^p = \int g$ by hypothesis.

We can therefore apply the generalized dominated convergence theorem (Folland Exercise 2.20) to obtain

$$2^p \int |f_n - f|^p = \int h_n \to \int h = \int 0 = 0.$$

By dividing through by $2^p > 0$, we obtain

$$\|f_n - f\|_p^p \to 0,$$

which implies $\|f_n - f\|_p \to 0$.

The above argument fails in the case $p = \infty$: if $p = \infty$, then when the measure space is $(R, \mathcal{L}, m)$, we have

$$|\|\chi_{(-n,n)}\|_\infty - \|\chi_{\mathbb{R}}\|_\infty| = 0 \to 0 \text{ as } n \to \infty,$$

but

$$\|\chi_{(-n,n)} - \chi_{\mathbb{R}}\|_\infty = 1 \not\to 0 \text{ as } n \to \infty.$$ □

Exercise 6.29: Folland Exercise 6.11.

If f is a measurable function on X, define the essential range R_f of f to be the set of all $z \in \mathbb{C}$ such that $\{x \mid |f(x) - z| < \varepsilon\}$ has positive measure for all $\varepsilon > 0$.

(a) R_f is closed.

(b) If $f \in L^\infty$, then R_f is compact and $\|f\|_\infty = \max\{|z| \mid z \in R_f\}$.

Exercise 6.30: Folland Exercise 6.12.

If $p \neq 2$, the L^p norm does not arise from an inner product on L^p, except in trivial cases when $\dim(L^p) \leqslant 1$. (Show that the parallelogram law fails.)

Solution. Let $(X, \mathcal{M}, \mu)$ be a measure space. Recall that since $\dim L^p \geqslant 2$, there exist disjoint sets $A, B \in \mathcal{M}$ of positive finite measure. Then for all $p \in [1,\infty) \setminus \{2\}$,

$$\begin{aligned}
2\left\|\frac{\chi_A}{\mu(A)^{1/p}}\right\|_p + 2\left\|\frac{\chi_B}{\mu(B)^{1/p}}\right\|_p &= 4 \neq 4^{1/p} + 4^{1/p} = (1+1)^{2/p} + (1+1)^{2/p} \qquad \text{(since } p \neq 2\text{)}\\
&= \left(\frac{1}{\mu(A)}\int|\chi_A|^p + \frac{1}{\mu(B)}\int|\chi_B|^p\right)^{2/p} + \left(\frac{1}{\mu(A)}\int|\chi_A|^p - \frac{1}{\mu(B)}\int|\chi_B|^p\right)^{2/p}\\
&= \left(\int\left|\frac{\chi_A}{\mu(A)^{1/p}}\right|^p + \int\left|\frac{\chi_B}{\mu(B)^{1/p}}\right|^p\right)^{2/p} + \left(\int\left|\frac{\chi_A}{\mu(A)^{1/p}}\right|^p - \int\left|\frac{\chi_B}{\mu(B)^{1/p}}\right|^p\right)^{2/p}\\
&= \left(\int\left|\frac{\chi_A}{\mu(A)^{1/p}} + \frac{\chi_B}{\mu(B)^{1/p}}\right|^p\right)^{2/p} + \left(\int\left|\frac{\chi_A}{\mu(A)^{1/p}} - \frac{\chi_B}{\mu(B)^{1/p}}\right|^p\right)^{2/p} \qquad \text{(since } A \cap B = \varnothing\text{)}\\
&= \left\|\frac{\chi_A}{\mu(A)^{1/p}} + \frac{\chi_B}{\mu(B)^{1/p}}\right\|_p^2 + \left\|\frac{\chi_A}{\mu(A)^{1/p}} - \frac{\chi_B}{\mu(B)^{1/p}}\right\|_p^2
\end{aligned}$$

Hence the parallelogram law fails. And if $p = \infty$, then with A and B as above we have

$$2 = \|\chi_A + \chi_B\|_\infty + \|\chi_A - \chi_B\|_\infty \neq 4 = 2\|\chi_A\|_\infty + 2\|\chi_B\|_\infty.$$

Thus for all $p \in [1,\infty] \setminus \{2\}$, $\|-\|_p$ does not arise from an inner product. □

Exercise 6.31: Folland Exercise 6.13.

$L^p(\mathbb{R}^n, m)$ is separable for $1 \leqslant p < \infty$. However, $L^\infty(\mathbb{R}^n, m)$ is not separable. (There is an uncountable set $\mathcal{F} \subset L^\infty$ such that $\|f - g\|_\infty \geqslant 1$ for all $f, g \in \mathcal{F}$ with $f \neq g$.)

Exercise 6.32: Folland Exercise 6.14.

If $g \in L^\infty$, the operator T defined by $Tf = fg$ is bounded on L^p for $1 \leqslant p \leqslant \infty$. Its operator norm is at most $\|g\|_\infty$, with equality if μ is semifinite.

Exercise 6.33: Folland Exercise 6.15: The Vitali Convergence Theorem.

Suppose $1 \leqslant p < \infty$ and $\{f_n\}_1^\infty \subset L^p$. In order for $\{f_n\}$ to be Cauchy in the L^p norm it is necessary and sufficient for the following three conditions to hold: (i) $\{f_n\}$ is Cauchy in measure; (ii) the sequence $\{|f_n|^p\}$ is uniformly integrable (see Folland Exercise 6.11 in Folland Section 3.2); and (iii) for every $\varepsilon > 0$ there exists $E \subset X$ such that $\mu(E) < \infty$ and $\int_{E^c} |f_n|^p < \varepsilon$ for all n. (To prove the sufficiency: Given $\varepsilon > 0$, let E be as in (iii), and let $A_{mn} = \{x \in E \mid |f_m(x) - f_n(x)| \geqslant \varepsilon\}$. Then the integrals of $|f_n - f_m|^p$ over $E \smallsetminus A_{mn}$, A_{mn}, and E^c are small when m and n are large—for three different reasons.)

Exercise 6.34: Folland Exercise 6.16.

If $0 < p < 1$, the formula $\rho(f,g) = \int |f-g|^p$ defines a metric on L^p that makes L^p into a complete topological vector space. (The proof of Theorem 8 still works for $p < 1$ if $\|f\|_p$ is replaced by $\int |f|^p$, as it uses only the triangle inequality and not the homogeneity of the norm.)

Exercise 6.35.

Determine precisely the set of triples $(p,q,r) \in \overline{\mathbb{R}}^3$ with $1 \leqslant r \leqslant p, q \leqslant \infty$ such that the following holds: if $f \in L^p(\mathbb{R}^n)$ and $g \in L^q(\mathbb{R}^n)$, then $fg \in L^r(\mathbb{R}^n)$ and $\|fg\|_r \leqslant \|f\|_p \|g\|_q$. (Here the underlying measure is Lebesgue measure.) Prove your answer.

Solution. We claim the set of triples for which this holds is given by

$$\mathcal{R} := \left\{(p,q,r) \in \overline{\mathbb{R}}^3 \;\middle|\; 1/p + 1/q = 1/r\right\}.$$

Proof. First suppose $(p,q,r) \in \mathcal{R}$, $f \in L^p(\mathbb{R}^n)$, and $g \in L^q(\mathbb{R}^n)$.

- Case 1: $1 \leqslant r \leqslant p, q < \infty$. Then $|f|^r \in L^{p/r}(\mathbb{R}^n)$ and $|g|^r \in L^{q/r}(\mathbb{R}^n)$, so by Hölder's inequality $|fg|^r = |f|^r |g|^r \in L^1(\mathbb{R}^n)$, hence $fg \in L^r$, and
$$\||fg|^r\|_1 \leqslant \||f|^r\|_{p/r} \||g|^r\|_{q/r}.$$
By raising both sides to the power of $1/r$, we obtain
$$\||fg|^r\|_1^{1/r} \leqslant \||f|^r\|_{p/r}^{1/r} \||g|^r\|_{q/r}^{1/r}, \tag{6.35.1}$$
so
$$\|fg\|_r = \left(\int |fg|^r\right)^{1/r} = \||fg|^r\|_1^{1/r} \overset{(6.35.1)}{\leqslant} \||f|^r\|_{p/r}^{1/r} \||g|^r\|_{q/r}^{1/r}$$
$$= \left(\int (|f|^r)^{p/r}\right)^{\frac{1}{\not r}\cdot\frac{\not r}{p}} \left(\int (|g|^r)^{q/r}\right)^{\frac{1}{\not r}\cdot\frac{\not r}{q}} = \left(\int |f|^p\right)^{1/p} \left(\int |g|^q\right)^{1/q} = \|f\|_p \|g\|_q.$$

- Case 2: $1 \leqslant r \leqslant p < q = \infty$ or $1 \leqslant r \leqslant q < p = \infty$. (Without loss of generality take $1 \leqslant r \leqslant p < q = \infty$.) Then $1/r = 1/p$, and since $g \in L^\infty$, there exists a bounded function g' such that $g' = g$ a.e.; thus $|fg'|^p = |fg|^p$ a.e., so
$$\|fg\|_p^p = \|fg'\|_p^p = \int |fg'|^p \leqslant \|g'\|_\infty^p \int |f|^p = \|g'\|_\infty^p \|f\|_p^p < \infty.$$
Hence $fg \in L^r$ $(= L^p)$, and by taking the pth root of both sides (and noting that the right-hand side is just $\|g\|_\infty^p \|f\|_p^p$ since $g = g'$ a.e.), we recover the desired inequality.
- Case 3: $p = q = r = \infty$. Then the claim holds, since if $E \in \mathcal{L}^n$ is an arbitrary set of positive measure then our assumptions imply $|f|_E|, |g|_E| < \infty$, hence $|f|_E| \cdot |g|_E| = |fg|_E| < \infty$, so fg is bounded on E. But E was an arbitrary set of positive measure, so $\|fg\|_\infty < \infty$. Thus $fg \in L^\infty$. And the inequality holds, since for a.e. x we have
$$|f(x)g(x)| \leqslant \|f\|_\infty |g(x)| \leqslant \|f\|_\infty \|g\|_\infty,$$
so $\|fg\|_\infty \leqslant \|f\|_\infty \|g\|_\infty$.

Now suppose $(p, q, r) \in \overline{\mathbb{R}}^3 \setminus \mathcal{R}$.

- Case 1: $1 \leqslant r \leqslant p, q < \infty$. If $1/r > 1/p + 1/q$, but the desired conclusion fails, since otherwise
$$2^{1/r} = \left\| \left(\frac{2\chi_{B_1(0)}}{\mu(B_1(0))} \right)^2 \right\|_r \leqslant \left\| \frac{2\chi_{B_1(0)}}{\mu(B_1(0))} \right\|_p \left\| \frac{2\chi_{B_1(0)}}{\mu(B_1(0))} \right\|_q = 2^{1/p} \cdot 2^{1/q},$$
so $1/r \leqslant 1/p + 1/q$, a contradiction. It fails similarly if $1/r < 1/p + 1/q$, since otherwise
$$\frac{1}{2^r} = \left\| \left(\frac{\chi_{B_1(0)}}{2\mu(B_1(0))} \right)^2 \right\|_r \leqslant \left\| \frac{\chi_{B_1(0)}}{2\mu(B_1(0))} \right\|_p \left\| \frac{\chi_{B_1(0)}}{2\mu(B_1(0))} \right\|_q = \frac{1}{2^p} \cdot \frac{1}{2^q},$$
so $2^{1/p+1/q} \leqslant 2^{1/r}$, and hence $1/r \geqslant 1/p + 1/q$, a contradiction.
- Case 2: $1 \leqslant r \leqslant p < q = \infty$ or $1 \leqslant r \leqslant q < p = \infty$. (Without loss of generality take $1 \leqslant r \leqslant p < q = \infty$.) If $1/p < 1/r$, then the desired conclusion fails, since otherwise
$$\mu(B_1(0))^{1/r} = \|\chi_{B_1(0)}^2\|_r \leqslant \|\chi_{B_1(0)}\|_\infty \|\chi_{B_1(0)}\|_p = 1 \cdot \mu(B_1(0))^{1/p},$$
so $1/r \leqslant 1/p$, a contradiction.

  Similarly, if $1/p > 1/r$, then the desired conclusion fails, since otherwise
$$\mu(B_1(0))^{-1/r} = \left\| \left(\frac{\chi_{B_1(0)}}{\mu(B_1(0))} \right)^2 \right\|_r \leqslant \left\| \frac{\chi_{B_1(0)}}{\mu(B_1(0))} \right\|_p \left\| \frac{\chi_{B_1(0)}}{\mu(B_1(0))} \right\|_\infty = \mu(B_1(0))^{-1/p},$$
so $1/p \leqslant 1/r$, a contradiction.
- Case 3: $p = q = r = \infty$. Then the desired conclusion fails, since otherwise
$$\mu(B_1(0)) = \|(\chi_{B_1(0)})^2\|_r \leqslant \|\chi_{B_1(0)}\|_\infty \|\chi_{B_1(0)}\|_\infty = 1 \cdot 1 = 1,$$
which fails for all $n \in \mathbb{Z}_{\geqslant 1}$.

We conclude $\mathcal{R}$ is precisely the set of triples such that the given statement is true. □

6.2 The Dual of L^p

Suppose that p and q are conjugate exponents. Hölder's inequality shows that each $g \in L^q$ defines a bounded linear functional ϕ_g on L^p by

$$\phi_g(f) = \int fg$$

and the operator norm of ϕ_g is at most $\|g\|_q$. (If $p = 2$ and we are thinking of L^2 as a Hilbert space, it is more appropriate to define $\phi_g(f) = \int f\overline{g}$. The same convention can be used for $p \neq 2$ without changing the results below in an essential way.) In fact, the map $g \to \phi_g$ is almost always an isometry from L^q into $(L^p)^*$.

Proposition 6.36: 6.13.

Suppose that p and q are conjugate exponents and $1 \leqslant q < \infty$. If $g \in L^q$, then

$$\|g\|_q = \|\phi_g\| = \sup\left\{\left|\int fg\right| \,\middle|\, \|f\|_p = 1\right\}$$

If μ is semifinite, this result holds also for $q = \infty$.

Proof. Hölder's inequality says that $\|\phi_g\| \leqslant \|g\|_q$, and equality is trivial if $g = 0$ (a.e.). If $g \neq 0$ and $q < \infty$, let

$$f = \frac{|g|^{q-1}\overline{\operatorname{sgn} g}}{\|g\|_q^{q-1}}$$

Then

$$\|f\|_p^p = \frac{\int |g|^{(q-1)p}}{\|g\|_q^{(q-1)p}} = \frac{\int |g|^q}{\int |g|^q} = 1$$

so

$$\|\phi_g\| \geqslant \int fg = \frac{\int |g|^q}{\|g\|_q^{q-1}} = \|g\|_q$$

(If $q = 1$, then $f = \overline{\operatorname{sgn} g}, \|f\|_\infty = 1$, and $\int fg = \|g\|_1$.) If $q = \infty$, for $\varepsilon > 0$ let $A = \{x \mid |g(x)| > \|g\|_\infty - \varepsilon\}$. Then $\mu(A) > 0$, so if μ is semifinite there exists $B \subset A$ with $0 < \mu(B) < \infty$. Let $f = \mu(B)^{-1}\chi_B\overline{\operatorname{sgn} g}$; then $\|f\|_1 = 1$, so

$$\|\phi_g\| \geqslant \int fg = \frac{1}{\mu(B)}\int_B |g| \geqslant \|g\|_\infty - \varepsilon$$

Since ε is arbitrary, $\|\phi_g\| = \|g\|_\infty$. □

Conversely, if $f \to \int fg$ is a bounded linear functional on L^p, then $g \in L^q$ in almost all cases. In fact, we have the following stronger result.

Theorem 6.37: 6.14.

Let p and q be conjugate exponents. Suppose that g is a measurable function on X such that $fg \in L^1$ for all f in the space Σ of simple functions that vanish outside a set of finite measure, and the quantity

$$M_q(g) = \sup\left\{ \left| \int fg \right| \;\middle|\; f \in \Sigma \text{ and } \|f\|_p = 1 \right\}$$

is finite. Also, suppose either that $S_g = \{x \mid g(x) \neq 0\}$ is σ-finite or that μ is semifinite. Then $g \in L^q$ and $M_q(g) = \|g\|_q$.

Proof. First, we remark that if f is a bounded measurable function that vanishes outside a set E of finite measure and $\|f\|_p = 1$, then $\left|\int fg\right| \leqslant M_q(g)$. Indeed, by Theorem 18 there is a sequence $\{f_n\}$ of simple functions such that $|f_n| \leqslant |f|$ (in particular, f_n vanishes outside E) and $f_n \to f$ a.e. Since $|f_n| \leqslant \|f\|_\infty \chi_E$ and $\chi_E g \in L^1$, by the dominated convergence theorem we have $\left|\int fg\right| = \lim\left|\int f_n g\right| \leqslant M_q(g)$.

Now suppose that $q < \infty$. We may assume that S_g is σ-finite, as this condition automatically holds when μ is semifinite; see Folland Exercise 6.17. Let $\{E_n\}$ be an increasing sequence of sets of finite measure such that $S_g = \bigcup_1^\infty E_n$. Let $\{\phi_n\}$ be a sequence of simple functions such that $\phi_n \to g$ pointwise and $|\phi_n| \leqslant |g|$, and let $g_n = \phi_n \chi_{E_n}$. Then $g_n \to g$ pointwise, $|g_n| \leqslant |g|$, and g_n vanishes outside E_n. Let

$$f_n = \frac{|g_n|^{q-1}\overline{\operatorname{sgn} g}}{\|g_n\|_q^{q-1}}$$

Then as in the proof of Proposition 36 we have $\|f_n\|_p = 1$, and by Fatou's lemma,

$$\begin{aligned} \|g\|_q \leqslant \liminf \|g_n\|_q &= \liminf \int |f_n g_n| \\ &\leqslant \liminf \int |f_n g| = \liminf \int f_n g \leqslant M_q(g) \end{aligned}$$

(For the last estimate we used the remark at the beginning of the proof.) On the other hand, Hölder's inequality gives $M_q(g) \leqslant \|g\|_q$, so the proof is complete for the case $q < \infty$.

Now suppose $q = \infty$. Given $\varepsilon > 0$, let $A = \{x \mid |g(x)| \geqslant M_\infty(g) + \varepsilon\}$. If $\mu(A)$ were positive, we could choose $B \subset A$ with $0 < \mu(B) < \infty$ (either because μ is semifinite or because $A \subset S_g$). Setting $f = \mu(B)^{-1}\chi_B \overline{\operatorname{sgn} g}$, we would then have $\|f\|_1 = 1$, and $\int fg = \mu(B)^{-1}\int_B |g| \geqslant M_\infty(g) + \varepsilon$. But this is impossible by the remark at the beginning of the proof. Hence $\|g\|_\infty \leqslant M_\infty(g)$, and the reverse inequality is obvious.

The last and deepest part of the description of $(L^p)^*$ is the fact that the map $g \to \phi_g$ is, in almost all cases, a surjection.

Theorem 6.38: 6.15.

Let p and q be conjugate exponents. If $1 < p < \infty$, for each $\phi \in (L^p)^*$ there exists $g \in L^q$ such that $\phi(f) = \int fg$ for all $f \in L^p$, and hence L^q is isometrically isomorphic to $(L^p)^*$. The same conclusion holds for $p = 1$ provided μ is σ-finite.

Proof. First let us suppose that μ is finite, so that all simple functions are in L^p. If $\phi \in (L^p)^*$ and E is a measurable set, let $\nu(E) = \phi(\chi_E)$. For any disjoint sequence $\{E_j\}$, if $E = \bigcup_1^\infty E_j$ we have $\chi_E = \sum_1^\infty \chi_{E_j}$ where the series converges in the L^p norm:

$$\left\|\chi_E - \sum\nolimits_1^n \chi_{E_j}\right\|_p = \left\|\sum\nolimits_{n+1}^\infty \chi_{E_j}\right\|_p = \mu\Big(\bigcup\nolimits_{n+1}^\infty E_j\Big)^{1/p} \to 0 \text{ as } n \to \infty$$

(It is at this point that we need the assumption that $p < \infty$.) Hence, since ϕ is linear and continuous,

$$\nu(E) = \sum\nolimits_1^\infty \phi(\chi_{E_j}) = \sum\nolimits_1^\infty \nu(E_j)$$

so that ν is a complex measure. Also, if $\mu(E) = 0$, then $\chi_E = 0$ as an element of L^p, so $\nu(E) = 0$; that is, $\nu \ll \mu$. By the Radon-Nikodym theorem there exists $g \in L^1(\mu)$ such that $\phi(\chi_E) = \nu(E) = \int_E g d\mu$ for all E and hence $\phi(f) = \int fg d\mu$ for all simple functions f. Moreover, $\left|\int fg\right| \leqslant \|\phi\| \|f\|_p$, so $g \in L^q$ by Theorem 37. Once we know this, it follows from Proposition 9 that $\phi(f) = \int fg$ for all $f \in L^p$.

Now suppose that μ is σ-finite. Let $\{E_n\}$ be an increasing sequence of sets such that $0 < \mu(E_n) < \infty$ and $X = \bigcup_1^\infty E_n$, and let us agree to identify $L^p(E_n)$ and $L^q(E_n)$ with the subspaces of $L^p(X)$ and $L^q(X)$ consisting of functions that vanish outside E_n. The preceding argument shows that for each n there exists $g_n \in L^q(E_n)$ such that $\phi(f) = \int f g_n$ for all $f \in L^p(E_n)$, and $\|g_n\|_q = \|\phi | L^p(E_n)\| \leqslant \|\phi\|$. The function g_n is unique modulo alterations on nullsets, so $g_n = g_m$ a.e. on E_n for $n < m$, and we can define g a.e. on X by setting $g = g_n$ on E_n. By the monotone convergence theorem, $\|g\|_q = \lim \|g_n\|_q \leqslant \|\phi\|$, so $g \in L^q$. Moreover, if $f \in L^p$, then by the dominated convergence theorem, $f\chi_{E_n} \to f$ in the L^p norm and hence $\phi(f) = \lim \phi(f\chi_{E_n}) = \lim \int_{E_n} fg = \int fg$.

Finally, suppose that μ is arbitrary and $p > 1$, so that $q < \infty$. As above, for each σ-finite set $E \subset X$ there is an a.e.-unique $g_E \in L^q(E)$ such that $\phi(f) = \int f g_E$ for all $f \in L^p(E)$ and $\|g_E\|_q \leqslant \|\phi\|$. If F is σ-finite and $F \supset E$, then $g_F = g_E$ a.e. on E, so $\|g_F\|_q \geqslant \|g_E\|_q$. Let M be the supremum of $\|g_E\|_q$ as E ranges over all σ-finite sets, noting that $M \leqslant \|\phi\|$. Choose a sequence $\{E_n\}$ so that $\|g_{E_n}\|_q \to M$, and set $F = \bigcup_1^\infty E_n$. Then F is σ-finite and $\|g_F\|_q \geqslant \|g_{E_n}\|_q$ for all n, whence $\|g_F\|_q = M$. Now, if A is a σ-finite set containing F, we have

$$\int |g_F|^q + \int |g_{A \setminus F}|^q = \int |g_A|^q \leqslant M^q = \int |g_F|^q$$

and thus $g_{A\setminus F} = 0$ and $g_A = g_F$ a.e. (Here we use the fact that $q < \infty$.) But if $f \in L^p$, then $A = F \cup \{x \mid f(x) \neq 0\}$ is σ-finite, so $\phi(f) = \int f g_A = \int f g_F$. Thus we may take $g = g_F$, and the proof is complete. □

Corollary 6.39: 6.16.

If $1 < p < \infty$, L^p is reflexive.

We conclude with some remarks on the exceptional cases $p = 1$ and $p = \infty$. For any measure μ, the correspondence $g \mapsto \phi_g$ maps L^∞ into $(L^1)^*$, but in general it is neither injective nor surjective. Injectivity fails when μ is not semifinite. Indeed, if $E \subset X$ is a set of infinite measure that contains no subsets of positive finite measure, and $f \in L^1$, then $\{x \mid f(x) \neq 0\}$ is σ-finite and hence intersects E in a null set. It follows that $\phi_{\chi_E} = 0$ although $\chi_E \neq 0$ in L^∞. This problem, however, can be remedied by redefining L^∞; see Exercises 23-24. The failure of surjectivity is more subtle and is best illustrated by an example; see also Folland Exercise 6.25.

Let X be an uncountable set, μ = counting measure on $(X, \mathcal{P}(X))$, $\mathcal{P}$ = the σ algebra of countable or co-countable sets, and μ_0 = the restriction of μ to $\mathcal{P}$. Every $f \in L^1(\mu)$ vanishes outside a countable set, and it follows that $L^1(\mu) = L^1(\mu_0)$. On the other hand, $L^\infty(\mu)$ consists of all bounded functions on X, whereas $L^\infty(\mu_0)$ consists of those bounded functions that are constant except on a countable set. With this in mind, it is easy to see that the dual of $L^1(\mu_0)$ is $L^\infty(\mu)$ and not the smaller space $L^\infty(\mu_0)$.

As for the case $p = \infty$: the map $g \to \phi_g$ is always an isometric injection of L^1 into $(L^\infty)^*$ by Proposition 36, but it is almost never a surjection. We shall say more about this in Folland Section 6.6; for the present, we give a specific example. (Another example can be found in Folland Exercise 6.19.)

Let $X = [0, 1], \mu$ = Lebesgue measure. The map $f \mapsto f(0)$ is a bounded linear functional on $C(X)$, which we regard as a subspace of L^∞. By the Hahn-Banach theorem there exists $\phi \in (L^\infty)^*$ such that $\phi(f) = f(0)$ for all $f \in C(X)$. To see that ϕ cannot be given by integration against an L^1 function, consider the functions $f_n \in C(X)$ defined by $f_n(x) = \max(1 - nx, 0)$. Then $\phi(f_n) = f_n(0) = 1$ for all n, but $f_n(x) \to 0$ for all $x > 0$, so by the dominated convergence theorem, $\int f_n g \to 0$ for all $g \in L^1$.

Exercise 6.40: Folland Exercise 6.17.

With notation as in Theorem 37, if μ is semifinite, $q < \infty$, and $M_q(g) < \infty$, then $\{x \mid |g(x)| > \varepsilon\}$ has finite measure for all $\varepsilon > 0$ and hence S_g is σ-finite.

Exercise 6.41: Folland Exercise 6.18.

The self-duality of L^2 follows from Hilbert space theory (Theorem 111), and this fact can be used to prove the Lebesgue-Radon-Nikodym theorem by the following argument due to von Neumann. Suppose that μ, ν are positive finite measures on $(X, \mathcal{M})$ (the σ-finite case follows easily as in §3.2), and let $\lambda = \mu + \nu$.

(a) The map $f \mapsto \int f d\nu$ is a bounded linear functional on $L^2(\lambda)$, so $\int f d\nu = \int f g d\lambda$

for some $g \in L^2(\lambda)$. Equivalently, $\int f(1-g)d\nu = \int fg d\mu$ for $f \in L^2(\lambda)$.
(b) $0 \leqslant g \leqslant 1 \lambda$-a.e., so we may assume $0 \leqslant g \leqslant 1$ everywhere.
(c) Let $A = \{x \mid g(x) < 1\}, B = \{x \mid g(x) = 1\}$, and set $\nu_a(E) = \nu(A \cap E)$, $\nu_s(E) = \nu(B \cap E)$. Then $\nu_s \perp \mu$ and $\nu_a \ll \mu$; in fact, $d\nu_a = g(1-g)^{-1}\chi_A d\mu$.

Exercise 6.42: Folland Exercise 6.19.

Define $\phi_n \in (l^\infty)^*$ by $\phi_n(f) = n^{-1}\sum_1^n f(j)$. Then the sequence $\{\phi_n\}$ has a weak* cluster point ϕ, and ϕ is an element of $(l^\infty)^*$ that does not arise from an element of ℓ^1.

Exercise 6.43: Folland Exercise 6.20.

Suppose $\sup_n \|f_n\|_p < \infty$ and $f_n \to f$ a.e.
(a) If $1 < p < \infty$, then $f_n \to f$ weakly in L^p. (Given $g \in L^q$, where q is conjugate to p, and $\varepsilon > 0$, there exist (i) $\delta > 0$ such that $\int_E |g|^q < \varepsilon$ whenever $\mu(E) < \delta$, (ii) $A \subset X$ such that $\mu(A) < \infty$ and $\int_{X \setminus A} |g|^q < \varepsilon$, and (iii) $B \subset A$ such that $\mu(A \setminus B) < \delta$ and $f_n \to f$ uniformly on B.)
(b) The result of (a) is false in general for $p = 1$. (Find counterexamples in $L^1(\mathbb{R}, m)$ and ℓ^1.) It is, however, true for $p = \infty$ if μ is σ-finite and weak convergence is replaced by weak* convergence.

Exercise 6.44: Folland Exercise 6.21.

If $1 < p < \infty$, $f_n \to f$ weakly in $\ell^p(A)$ if and only if $\sup_n \|f_n\|_p < \infty$ and $f_n \to f$ pointwise.

Solution. Let $1 < p < \infty$, let $f \in \ell^p(A)$ (we may assume this as mentioned on canvas), and let $q' = p$.

- Suppose $f_n \to f$ weakly in ℓ^p and $q = p'$. Then in particular the ℓ^q function $\chi_{\{a\}}$ has
$$\sum\nolimits_{a \in A} f_n(a)\chi_{\{a\}} = f_n(a) \to f(a) \text{ as } n \to \infty,$$
so $f_n \to f$ pointwise. For each n, define $\widehat{f}_n(g) = \int g f_n$. Since $f_n \to f$ weakly, the sequence $\{z_n\}_{n=1}^\infty \subset \mathbb{C}$ given by $z_n := \int g f_n$ converges, and hence is bounded in $\mathbb{C}$. Then for all $g \in \ell^q$,
$$\sup_n |\widehat{f}_n(g)| = \sup_n |z_n| < \infty,$$
so
$$\sup_n \|f_n\|_p = \sup_n \|\widehat{f}\| < \infty,$$
where the final inequality is by the uniform boundedness theorem.
- Conversely, suppose that $f_n \to f$ pointwise and $\sup_n \|f_n\|_p < \infty$. Fix $g \in \ell^q = \ell^{p'}$ and $\varepsilon > 0$. We claim $|\langle g, f_n\rangle - \langle g, f\rangle| < \varepsilon$, where $\langle -, * \rangle := \int |(-) \cdot (*)|$. Let $M =$

$\|f\|_p + \sup_n \|f_n\|_p$. Then $M < \infty$ by hypothesis, and we may assume $M > 0$ (since otherwise f_n, and hence f are 0). Since $\|g\|_q^q = \sum_{a\in A}|g(a)|^q < \infty$, we must have $g(a) = 0$ for all but countably many $a \in A$. Thus we may assume $A = \mathbb{Z}_{\geqslant 1}$.

For all $k \in \{1, \ldots, K-1\}$, there exists $N_K \in \mathbb{Z}_{\geqslant 1}$ such that for all $n \geqslant N_k$, $|f_n(k) - f(k)| < \varepsilon/(2(K-1)|g(k)|)$. (If $|g(k)| = 0$, then we may ignore the term $|g(k)||f_n(k) - f(k)| = 0$ in the sum, so this is valid.) Thus, for all $n \geqslant \max\{N_1, \ldots, N_k\}$,

$$\sum_{k=1}^{K-1} |g(k)||f_n(k) - f(k)| \leqslant \sum_{k=1}^{K-1} \frac{\varepsilon \cancel{|g(k)|}}{2(K-1)\cancel{|g(k)|}} = \frac{\varepsilon}{2}. \tag{6.44.1}$$

On the other hand, since $\|g\|_q^q < \infty$, there exists $K \geqslant 2$ such that for all sufficiently large n,

$$\|\chi_{A'} g\|_q^q = \sum_{k=K}^{\infty} |g(k)|^q < \left(\frac{\varepsilon}{2M}\right)^q.$$

Then, respectively, by Hölder's inequality and the triangle inequality, for all sufficiently large n,

$$\sum_{k=K}^{\infty} |g(k)||f_n(k) - f(k)| \leqslant \|f_n - f\|_p \|\chi_{A'} g\|_q \leqslant M \frac{\varepsilon}{2M} = \frac{\varepsilon}{2}. \tag{6.44.2}$$

Thus

$$\begin{aligned} |\langle g, f_n - f\rangle| &= \sum_{k=1}^{\infty} |g(k)||f_n(k) - f(k)| \\ &= \underbrace{\sum_{k=1}^{K-1} |g(k)||f_n(k) - f(k)|}_{<\varepsilon/2 \text{ by } (6.44.1)} + \underbrace{\sum_{k=K}^{\infty} |g(k)||f_n(k) - f(k)|}_{<\varepsilon/2 \text{ by } (6.44.2)} < \varepsilon, \end{aligned}$$

so $f_n \to f$ weakly. □

Exercise 6.45: Folland Exercise 6.22.

Let $X = [0, 1]$, with Lebesgue measure.

(a) Let $f_n(x) = \cos 2\pi n x$. Then $f_n \to 0$ weakly in L^2 (see Folland Exercise 5.63), but $f_n \not\to 0$ a.e. or in measure.

(b) Let $f_n(x) = n\chi_{(0,1/n)}$. Then $f_n \to 0$ a.e. and in measure, but $f_n \not\to 0$ weakly in L^p for any p.

Exercise 6.46: Folland Exercise 6.23.

Let $(X, \mathcal{M}, \mu)$ be a measure space. A set $E \in \mathcal{M}$ is called locally null if $\mu(E \cap F) = 0$ for every $F \in \mathcal{M}$ such that $\mu(F) < \infty$. If $f\colon X \to \mathbb{C}$ is a measurable function, define

$$\|f\|_* = \inf\{a \mid \{x \mid |f(x)| > a\} \text{ is locally null}\}$$

and let $\mathcal{L}^\infty = \mathcal{L}^\infty(X, \mathcal{L}, \mu)$ be the space of all measurable f such that $\|f\|_* < \infty$. We consider $f, g \in \mathcal{L}^\infty$ to be identical if $\{x \mid f(x) \neq g(x)\}$ is locally null.

(a) If E is locally null, then $\mu(E)$ is either 0 or ∞. If μ is semifinite, then every locally null set is null.
(b) $\|\cdot\|_*$ is a norm on $\mathcal{L}^\infty$ that makes $\mathcal{L}^\infty$ into a Banach space. If μ is semifinite, then $\mathcal{L}^\infty = L^\infty$.

Exercise 6.47: Folland Exercise 6.24.

If $g \in \mathcal{L}^\infty$ (see Folland Exercise 6.23), then $\|g\|_* = \sup\{|\int fg| \mid \|f\|_1 = 1\}$, so the map $g \mapsto \phi_g$ is an isometry from $\mathcal{L}^\infty$ into $(L^1)^*$. Conversely, if $M_\infty(g) < \infty$ as in Theorem 37, then $g \in \mathcal{L}^\infty$ and $M_\infty(g) = \|g\|_*$.

Exercise 6.48: Folland Exercise 6.25.

Suppose μ is decomposable (see Folland Exercise 3.15). Then every $\phi \in (L^1)^*$ is of the form $\phi(f) = \int fg$ for some $g \in \mathcal{L}^\infty$, and hence $(L^1)^* \cong \mathcal{L}^\infty$ (see Exercises 46 and 47). (If $\mathcal{L}$ is a decomposition of μ and $f \in L^1$, there exists $\{E_j\} \subset \mathcal{L}$ such that $f = \sum_1^\infty f\chi_{E_j}$ where the series converges in L^1.)

6.3 Some Useful Inequalities

Estimates and inequalities lie at the heart of the applications of L^p spaces in analysis. The most basic of these are the Hölder and Minkowski inequalities. In this section we present a few additional important results in this area. The first one is almost a triviality, but it is sufficiently useful to warrant special mention.

Theorem 6.49: 6.17: Chebyshev's Inequality.

If $f \in L^p(0 < p < \infty)$, then for any $\alpha > 0$,

$$\mu(\{x \mid |f(x)| > \alpha\}) \leqslant \left[\frac{\|f\|_p}{\alpha}\right]^p.$$

Proof. Let $E_\alpha = \{x \mid |f(x)| > \alpha\}$. Then

$$\|f\|_p^p = \int |f|^p \geqslant \int_{E_\alpha} |f|^p \geqslant \alpha^p \int_{E_\alpha} 1 = \alpha^p \mu(E_\alpha). \qquad \square$$

The next result is a rather general theorem about boundedness of integral operators on L^p spaces.

Theorem 6.50: 6.18: Schur's Test (or Generalized Young's Inequality).

Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ be σ-finite measure spaces, and let $K(x,y)$ be a measurable function on $X \times Y$.

(i) $\int_X |K(x,y)|\,\mathrm{d}\mu(x) \leqslant C$ for a.e. $y \in Y$,
(ii) $\int_Y |K(x,y)|\,\mathrm{d}\nu(y) \leqslant C$ for a.e. $x \in X$.

Then for all $p \in [1,\infty]$ and all $f \in L^p$, the integral

$$Tf(x) \coloneqq \int_Y K(x,y)f(y)\,\mathrm{d}\nu(y)$$

converges absolutely for a.e. $x \in X$, $Tf \in L^p$, and $\|Tf\|_p \leqslant C\|f\|_p$ (and, in particular, T is bounded).

Proof. Suppose that $1 < p < \infty$. Let q be the conjugate exponent to p. By applying Hölder's inequality to the product

$$|K(x,y)f(y)| = |K(x,y)|^{1/q}(|K(x,y)|^{1/p}|f(y)|)$$

we have

$$\begin{aligned}\int |K(x,y)f(y)|d\nu(y) &\leqslant \left[\int |K(x,y)|d\nu(y)\right]^{1/q}\left[\int |K(x,y)||f(y)|^p d\nu(y)\right]^{1/p}\\ &\leqslant C^{1/q}\left[\int |K(x,y)||f(y)|^p d\nu(y)\right]^{1/p}\end{aligned}$$

for a.e. $x \in X$. Hence, by Tonelli's theorem,

$$\begin{aligned}\int\left[\int |K(x,y)f(y)|d\nu(y)\right]^p d\mu(x) &\leqslant C^{p/q}\iint |K(x,y)||f(y)|^p d\nu(y)d\mu(x)\\ &\leqslant C^{(p/q)+1}\int |f(y)|^p d\nu(y).\end{aligned}$$

Since the last integral is finite, Fubini's theorem implies that $K(x,\cdot)f \in L^1(\nu)$ for a.e. x, so that Tf is well defined a.e., and

$$\int |Tf(x)|^p d\mu(x) \leqslant C^{(p/q)+1}\|f\|_p^p$$

Taking pth roots, we are done.

For $p = 1$ the proof is similar but easier and requires only the hypothesis $\int |K(x,y)|\,\mathrm{d}\mu(x) \leqslant C$; for $p = \infty$ the proof is trivial and requires only the hypothesis $\int |K(x,y)|d\nu(y) \leqslant C$. Details are left to the reader (Folland Exercise 6.26). □

Minkowski's inequality states that the L^p norm of a sum is at most the sum of the L^p norms. There is a generalization of this result in which sums are replaced by integrals:

Theorem 6.51: 6.19: Minkowski's Inequality for Integrals.

Suppose that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{M}, \nu)$ are σ-finite measure spaces, and let f be an $(\mathcal{M} \otimes \mathcal{M})$-measurable function on $X \times Y$.

(a) If $f \geqslant 0$ and $1 \leqslant p < \infty$, then

$$\left[\int\left(\int f(x,y)d\nu(y)\right)^p d\mu(x)\right]^{1/p} \leqslant \int\left[\int f(x,y)^p d\mu(x)\right]^{1/p} d\nu(y).$$

(b) If $1 \leqslant p \leqslant \infty$, $f(\cdot, y) \in L^p(\mu)$ for a.e. y, and the function $y \mapsto \|f(\cdot,y)\|_p$ is in $L^1(\nu)$, then $f(x,\cdot) \in L^1(\nu)$ for a.e. x, the function $x \mapsto \int f(x,y)d\nu(y)$ is in $L^p(\mu)$, and

$$\left\|\int f(\cdot,y)d\nu(y)\right\|_p \leqslant \int \|f(\cdot,y)\|_p d\nu(y).$$

Proof. If $p = 1$, (a) is merely Tonelli's theorem. If $1 < p < \infty$, let q be the conjugate exponent to p and suppose $g \in L^q(\mu)$. Then by Tonelli's theorem and Hölder's inequality,

$$\begin{aligned}\int\left[\int f(x,y)d\nu(y)\right]|g(x)|d\mu(x) &= \iint f(x,y)|g(x)|d\mu(x)d\nu(y)\\ &\leqslant \|g\|_q \int\left[\int f(x,y)^p d\mu(x)\right]^{1/p} d\nu(y).\end{aligned}$$

Assertion (a) therefore follows from Theorem 37. When $p < \infty$, (b) follows from (a) (with f replaced by $|f|$) and Fubini's theorem; when $p = \infty$, it is a simple consequence of the monotonicity of the integral. □

Our final result is a theorem concerning integral operators on $(0,\infty)$ with Lebesgue measure.

Theorem 6.52: 6.20.

Let K be a Lebesgue measurable function on $(0,\infty) \times (0,\infty)$ such that $K(\lambda x, \lambda y) = \lambda^{-1}K(x,y)$ for all $\lambda > 0$ and $\int_0^\infty |K(x,1)|x^{-1/p}dx = C < \infty$ for some $p \in [1,\infty]$, and let q be the conjugate exponent to p. For $f \in L^p$ and $g \in L^q$, let

$$Tf(y) = \int_0^\infty K(x,y)f(x)dx, \quad Sg(x) = \int_0^\infty K(x,y)g(y)dy$$

Then Tf and Sg are defined a.e., and $\|Tf\|_p \leqslant C\|f\|_p$ and $\|Sg\|_q \leqslant C\|g\|_q$.

Proof. Setting $z = x/y$, we have

$$\int_0^\infty |K(x,y)f(x)|dx = \int_0^\infty |K(yz,y)f(yz)|ydz = \int_0^\infty |K(z,1)f_z(y)|dz$$

where $f_z(y) = f(yz)$; moreover,

$$\|f_z\|_p = \left[\int_0^\infty |f(yz)|^p dy\right]^{1/p} = \left[\int_0^\infty |f(x)|^p z^{-1} dx\right]^{1/p} = z^{-1/p}\|f\|_p$$

Therefore, by Minkowski's inequality for integrals, Tf exists a.e. and

$$\|Tf\|_p \leqslant \int_0^\infty |K(z,1)| \|f_z\|_p dz = \|f\|_p \int_0^\infty |K(z,1)| z^{-1/p} dz = C\|f\|_p$$

Finally, setting $u = y^{-1}$, we have

$$\begin{aligned} \int_0^\infty |K(1,y)| y^{-1/q} dy &= \int_0^\infty |K(y^{-1},1)| y^{-1-(1/q)} dy \\ &= \int_0^\infty |K(u,1)| u^{-1/p} du = C \end{aligned}$$

so the same reasoning shows that Sg is defined a.e. and that $\|Sg\|_q \leqslant C\|g\|_q$. □

Corollary 6.53: 6.21.

Let

$$Tf(y) = y^{-1} \int_0^y f(x) dx, \quad Sg(x) = \int_x^\infty y^{-1} g(y) dy$$

Then for $1 < p \leqslant \infty$ and $1 \leqslant q < \infty$,

$$\|Tf\|_p \leqslant \frac{p}{p-1}\|f\|_p, \quad \|Sg\|_q \leqslant q\|g\|_q$$

Proof. Let $K(x,y) = y^{-1}\chi_E(x,y)$ where $E = \{(x,y) \mid x < y\}$. Then $\int_0^\infty |K(x,1)| x^{-1/p} dx = \int_0^1 x^{-1/p} dx = p/(p-1) = q$, where q is the conjugate exponent to p, so Theorem 52 yields the result. □

Corollary 53 is a special case of Hardy's inequalities; the general result is in Folland Exercise 6.29.

Exercise 6.54: Folland Exercise 6.26.

Complete the proof of Theorem 50 for the cases $p = 1$ and $p = \infty$.

Exercise 6.55: Folland Exercise 6.27.

(Hilbert's Inequality) The operator $Tf(x) = \int_0^\infty (x+y)^{-1} f(y) dy$ satisfies $\|Tf\|_p \leqslant C_p\|f\|_p$ for $1 < p < \infty$, where $C_p = \int_0^\infty x^{-1/p}(x+1)^{-1} dx$. (For those who know about contour integrals: Show that $C_p = \pi \csc(\pi/p)$.)

Exercise 6.56: Folland Exercise 6.28.

Let I_α be the αth fractional integral operator as in Folland Exercise 2.61 and let $J_\alpha f(x) = x^{-\alpha} I_\alpha f(x)$.

(a) J_α is bounded on $L^p(0,\infty)$ for $1 < p \leqslant \infty$; more precisely,

$$\|J_\alpha f\|_p \leqslant \frac{\Gamma(1-p^{-1})}{\Gamma(\alpha+1-p^{-1})}\|f\|_p$$

(b) There exists $f \in L^1(0,\infty)$ such that $J_1 f \notin L^1(0,\infty)$.

Exercise 6.57: Folland Exercise 6.29.

Suppose that $1 \leqslant p < \infty, r > 0$, and h is a nonnegative measurable function on $(0,\infty)$. Then:

$$\int_0^\infty x^{-r-1}\left[\int_0^x h(y)dy\right]^p dx \leqslant \left(\frac{p}{r}\right)^p \int_0^\infty x^{p-r-1}h(x)^p dx$$
$$\int_0^\infty x^{r-1}\left[\int_x^\infty h(y)dy\right]^p dx \leqslant \left(\frac{p}{r}\right)^p \int_0^\infty x^{p+r-1}h(x)^p dx.$$

(Apply Theorem 52 with $K(x,y) = x^{\beta-1}y^{-\beta}\chi_{(0,\infty)}(y-x)$, $f(x) = x^\gamma h(x)$, and $g(x) = x^\delta h(x)$ for suitable β,γ,δ.)

Exercise 6.58: Folland Exercise 6.30.

Suppose that K is a nonnegative measurable function on $(0,\infty)$ such that $\int_0^\infty K(x)x^{s-1}dx = \phi(s) < \infty$ for $0 < s < 1$.

(a) If $1 < p < \infty, p^{-1}+q^{-1} = 1$, and f,g are nonnegative measurable functions on $(0,\infty)$, then (with $\int = \int_0^\infty$)

$$\iint K(xy)f(x)g(y)dxdy \leqslant \phi(p^{-1})\left[\int x^{p-2}f(x)^p dx\right]^{1/p}\left[\int g(x)^q dx\right]^{1/q}.$$

(b) The operator $Tf(x) = \int_0^\infty K(xy)f(y)dy$ is bounded on $L^2((0,\infty))$ with norm $\leqslant \phi(\frac{1}{2})$. (Interesting special case: If $K(x) = e^{-x}$, then T is the Laplace transform and $\phi(s) = \Gamma(s)$.)

Solution.

(a) The integrand of the left-hand side is a nonnegative measurable function (since f, g, and K are), so we can apply Tonelli's theorem below:

$$\int_0^\infty\int_0^\infty K(xy)f(x)g(y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^\infty\int_0^\infty K(z)\frac{f(z/y)}{y}g(y)\,\mathrm{d}z\,\mathrm{d}y \;\; (z \coloneqq xy,\ \mathrm{d}x = \mathrm{d}z/y)$$
$$= \int_0^\infty K(z)\int_0^\infty \frac{f(z/y)}{y}g(y)\,\mathrm{d}y\,\mathrm{d}z$$

$$\begin{aligned}
&= \int_0^\infty K(z)\phi_g\left(y \mapsto \frac{f(z/y)}{y}\right) \mathrm{d}z \\
&\leqslant \int_0^\infty K(z)\left\| y \mapsto \frac{f(z/y)}{y}\right\|_p \|g\|_q \,\mathrm{d}y\,\mathrm{d}z \\
&= \int_0^\infty K(z)\left(\int_0^\infty \frac{f(w)^p}{(z/w)^p}\frac{-z}{w^2}\,\mathrm{d}w\right)^{1/p}\left(\int_0^\infty g(y)^p\,\mathrm{d}y\right)^{1/q}\mathrm{d}z \\
&\qquad\qquad (\text{substituting } w \coloneqq z/y,\ \mathrm{d}y = -z\,\mathrm{d}w/w^2) \\
&= \int_0^\infty K(z)z^{-1+1/p}\left(\int_0^\infty f(w)^p w^{p-2}\,\mathrm{d}w\right)^{1/p}\left(\int_0^\infty g(y)^p\,\mathrm{d}y\right)^{1/q}\mathrm{d}z.
\end{aligned}$$

Since $\int_0^\infty K(z)z^{1/p-1} = \phi(1/p)$ by definition, the desired inequality follows.

(b) Now consider $p = q = 2$ and define $T\colon L^2((0,\infty)) \to L^2((0,\infty))$ by $f(x) \mapsto \int_0^\infty K(xy)f(y)\,\mathrm{d}y$. Then T is linear, and T is bounded since for all $f \in L^2((0,\infty))$,

$$\begin{aligned}
\|Tf\|_2^2 &= \int |Tf(y)|^2\,\mathrm{d}y \\
&= \int\left|\int K(xy)f(x)\,\mathrm{d}x\right|^2 dy \\
&\leqslant \int\left(\int K(xy)|f(x)|\,\mathrm{d}x\right)^2 \mathrm{d}y \leqslant \phi\left(\frac{1}{2}\right)^2 \int x^0|f(x)|^2dx = \phi\left(\frac{1}{2}\right)^2\|f\|_2^2.
\end{aligned}$$

where the last inequality is by part (a). Since $f \in L^2((0,\infty))$, this shows $Tf \in L^2((0,\infty))$, so T is indeed a linear map $L^2((0,\infty)) \to L^2((0,\infty))$, and moreover that T is bounded and $\|Tf\|_2 \leqslant \phi\left(\frac{1}{2}\right)\|f\|_2$, which implies $\|T\| \leqslant \phi(1/2)$, as claimed. □

Exercise 6.59: Folland Exercise 6.31.

(A Generalized Hölder Inequality) Suppose that $1 \leqslant p_j \leqslant \infty$ and $\sum_1^n p_j^{-1} = r^{-1} \leqslant 1$. If $f_j \in L^{p_j}$ for $j = 1, \ldots, n$, then $\prod_1^n f_j \in L^r$ and $\left\|\prod_1^n f_j\right\|_r \leqslant \prod_1^n \|f_j\|_{p_j}$. (First do the case $n = 2$.)

Exercise 6.60: Folland Exercise 6.32.

Suppose that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{M}, \nu)$ are σ-finite measure spaces and $K \in L^2(\mu \times \nu)$. If $f \in L^2(\nu)$, the integral $Tf(x) = \int K(x,y)f(y)d\nu(y)$ converges absolutely for a.e. $x \in X$; moreover, $Tf \in L^2(\mu)$ and $\|Tf\|_2 \leqslant \|K\|_2\|f\|_2$.

Exercise 6.61: Folland Exercise 6.33.

Given $1 < p < \infty$, let $Tf(x) = x^{-1/p}\int_0^x f(t)dt$. If $p^{-1} + q^{-1} = 1$, then T is a bounded linear map from $L^q((0,\infty))$ to $C_0((0,\infty))$.

Exercise 6.62: Folland Exercise 6.34.

If f is absolutely continuous on $[\varepsilon, 1]$ for $0 < \varepsilon < 1$ and $\int_0^1 x|f'(x)|^p dx < \infty$, then $\lim_{x\to 0} f(x)$ exists (and is finite) if $p > 2$, $|f(x)|/|\log x|^{1/2} \to 0$ as $x \to 0$ if $p = 2$, and $|f(x)|/x^{1-(2/p)} \to 0$ as $x \to 0$ if $p < 2$.

Exercise 6.63.

The "uncentered" maximal function $\widetilde{M}f$ is defined by $(\widetilde{M}f)(x) = \sup_{x\in B} \frac{1}{m(B)} \int_B |f(y)|dy$ where the supremum is taken over all balls containing x (not only those balls centered at x). Here m denotes Lebesgue measure on $\mathbb{R}^n$.

(a) Obviously $(Mf)(x) \leqslant (\widetilde{M}f)(x)$. Show that there exists a constant c (depending only on the dimension) such that $(\widetilde{M}f)(x) \leqslant c(Mf)(x)$.

(b) Determine explicitly the function $\widetilde{M}(\chi_{[0,1]})$.

(c) It will be shown in class that M and $\widetilde{M}$ are bounded operators on $L^p(\mathbb{R}^n)$ for $1 < p \leqslant \infty$. Does there exist a pair (p, q) with $1 < p, q < \infty$ and $p \neq q$ such that M or $\widetilde{M}$ is a bounded operator from $L^p(\mathbb{R}^n)$ to $L^q(\mathbb{R}^n)$?

Solution.

(a) Fix $x \in \mathbb{R}^n$, let S be the collection of open balls containing x, let T be the collection of open balls centered at x, and for all Lebesgue measurable subsets E of $\mathbb{R}^n$ define

$$A_E|f| := \frac{1}{m(E)} \int_E |f(y)|\, dy.$$

Since $T \subset S$,

$$Mf(x) = \sup\nolimits_{E\in T} A_E|f| \leqslant \sup\nolimits_{E\in S} A_E|f| = \widetilde{M}f(x).$$

For the other inequality, let B_r be any ball containing x of radius r. Then $B \subset B_{2r}(x)$, so

$$\frac{1}{m(B_r)} \int_{B_r} |f(y)|\, dy \leqslant \frac{m(B_{2r}(x))}{m(B_r)} \frac{1}{m(B_{2r}(x))} \int_{B_{2r}(x)} |f(y)|dy \leqslant 2^n Mf(x)$$

Since B was any ball containing x, by taking the supremum over all such balls of all radii we obtain

$$\widetilde{M}f(x) \leqslant 2^n Mf(x).$$

(b) If $B \in S$, then $B = (a, b)$ for some $a, b \in \mathbb{R}$ such that $a < x < b$, so

$$A_B\chi_{[0,1]}(x) = \frac{1}{b-a} \int_{(a,b)} \chi_{[0,1]}(y)\, dy = \begin{cases} 1 & \text{if } (a,b) \subset [0,1], \\ \frac{m((a,b)\cap[0,1])}{b-a} & \text{if } (a,b) \cap [0,1] \neq \varnothing \text{ and } (a,b) \not\subset [0,1], \\ 0 & \text{if } (a,b) \cap [0,1] = \varnothing. \end{cases}$$

We now break into cases:

- If $x \in (0,1)$ then we can choose a, b such that $0 < a < x < b < 1$, in which case $\widetilde{M}\chi_{[0,1]}(x) = 1$.
- If $x = 0$ (resp. $x = 1$) then by considering the sequence of open intervals $\{E_n = (-1/n, 1)\}_{n=1}^{\infty}$ (resp. $\{E_n = (0, 1+1/n)\}_{n=1}^{\infty}$), we see $\widetilde{M}\chi_{[0,1]}(x) = \lim_{n\to\infty} A_{E_n}\chi_{[0,1]}(x) = 1$, so $\widetilde{M}\chi_{[0,1]}(x) = 1$ if $x \in \{0\} \cup \{1\}$.
- If $x < 0$, then for a fixed point $q \in [0,1]$ and the sequence $\{E_n = (x - 1/n, q)\}_{n=1}^{\infty}$, we have
$$A_{E_n}\chi_{[0,1]}(x) = \frac{m((x-1/n, q) \cap [0,1])}{q - x + 1/n} = \frac{q}{q - x + 1/n},$$
which tends to $q/(q-x)$ as $n \to \infty$. As a function of $q \in [0,1]$, $q/(q-x)$ is increasing to 1. Thus by taking $q = 1$ and the open sets $\{E_n = (x - 1/n, q + 1/n)\}_{n=1}^{\infty}$, we conclude that when $x < 0$, $\widetilde{M}\chi_{[0,1]}(x) = \lim_{q\nearrow 1} A_{E_n}\chi_{[0,1]}(x) = \lim_{q\nearrow 1} q/(q-x) = 1/(1-x)$.
- If $x > 1$, then by arguing similarly we obtain $\widetilde{M}\chi_{[0,1]}(x) = 1/x$ if $x > 1$.

We conclude
$$\widetilde{M}\chi_{[0,1]}(x) = \begin{cases} 1 & \text{if } 0 \leqslant x \leqslant 1, \\ 1/(1-x) & \text{if } x < 0, \\ 1/x & \text{if } x > 1. \end{cases} \qquad \square$$

(c) No. By part (a) M is bounded if and only if $\widetilde{M}$ is, so it suffices to prove M is not bounded as a map $L^p(\mathbb{R}^n) \to L^q(\mathbb{R}^n)$. Consider an arbitrary $t \in (0,\infty)$ and consider the open cube $(0,t)^n \subset \mathbb{R}^n$. For any $x \in \mathbb{R}^n$, we have
$$\begin{aligned} \|M\chi_{(0,t)^n}\|_q^q &= \int_{\mathbb{R}^n} |M\chi_{(0,t)^n}(x)|^q \, dx = \int_{\mathbb{R}^n} \left| \sup_{r>0} \frac{1}{m(B_r(x))} \int_{B_r(x)} \chi_{(0,t)^n}(y)\, dy \right|^q dx \\ &= \int_{\mathbb{R}^n} \left| \sup_{r>0} \frac{m(B_r(x) \cap (0,t)^n)}{m(B_r(x))} \right|^q dx = \int \chi_{(0,t)^n}(x)\, dx = m((0,t)^n), \end{aligned}$$
so $\|M\chi_{(0,t)^n}\|_q = m((0,t)^n)^{1/q} = t^{n/q}$. On the other hand, for an arbitrary constant C,
$$C\|\chi_{(0,t)^n}\|_p = Cm((0,t)^n)^{1/p} = Ct^{n/p}.$$
If M were bounded as an operator $L^p(\mathbb{R}^n) \to L^q(\mathbb{R}^n)$, then there exists a constant C such that for all $t \in (0,\infty)$, $t^{n/q} \leqslant Ct^{n/p}$, or equivalently, such that
$$t^{n(\frac{1}{q} - \frac{1}{p})} \leqslant C.$$
But this cannot be true at all $t \in (0,\infty)$ since p, q, n are fixed; by choosing sufficiently small t when $1/p > 1/q$ or sufficiently large t (when $1/p < 1/q$), this fails. Thus M, hence also $\widetilde{M}$, is unbounded as an operator $L^p(\mathbb{R}^n) \to L^q(\mathbb{R}^n)$.

6.4 Distribution Functions and Weak L^p

If f is a measurable function on $(X, \mathcal{M}, \mu)$, we define its distribution function $\lambda_f : (0, \infty) \to [0, \infty]$ by

$$\lambda_f(\alpha) = \mu(\{|f| > \alpha\})$$

(This is closely related, but not identical, to the "distribution functions" discussed in Folland Section 1.5 and Folland Section 10.1.) We compile the basic properties of λ_f in a proposition.

Proposition 6.64: 6.22.

(a) λ_f is decreasing and right continuous.
(b) If $|f| \leqslant |g|$, then $\lambda_f \leqslant \lambda_g$.
(c) If $|f_n|$ increases to $|f|$, then λ_{f_n} increases to λ_f.
(d) If $f = g + h$, then $\lambda_f(\alpha) \leqslant \lambda_g\left(\frac{1}{2}\alpha\right) + \lambda_h\left(\frac{1}{2}\alpha\right)$.

Proof. Let $E(\alpha, f) = \{x \mid |f(x)| > \alpha\}$. The function λ_f is decreasing since $E(\alpha, f) \supset E(\beta, f)$ if $\alpha < \beta$, and it is right continuous since $E(\alpha, f)$ is the increasing union of $\{E(\alpha + n^{-1}, f)\}_1^\infty$. If $|f| \leqslant |g|$, then $E(\alpha, f) \subset E(\alpha, g)$, so $\lambda_f \leqslant \lambda_g$. If $|f_n|$ increases to $|f|$, then $E(\alpha, f)$ is the increasing union of $\{E(\alpha, f_n)\}$, so λ_{f_n} increases to λ_f. Finally, if $f = g + h$, then $E(\alpha, f) \subset E\left(\frac{1}{2}\alpha, g\right) \cup E\left(\frac{1}{2}\alpha, h\right)$, which implies that $\lambda_f(\alpha) \leqslant \lambda_g\left(\frac{1}{2}\alpha\right) + \lambda_h\left(\frac{1}{2}\alpha\right)$.

Suppose that $\lambda_f(\alpha) < \infty$ for all $\alpha > 0$. In view of Proposition 64a, λ_f defines a negative Borel measure ν on $(0, \infty)$ such that $\nu((a, b]) = \lambda_f(b) - \lambda_f(a)$ whenever $0 < a < b$. (Our construction of Borel measures on $\mathbb{R}$ in Folland Section 1.5 works equally well on $(0, \infty)$.) We can therefore consider the Lebesgue-Stieltjes integrals $\int \phi d\lambda_f = \int \phi d\nu$ of functions ϕ on $(0, \infty)$. The following result shows that the integrals of functions of $|f|$ on X can be reduced to such Lebesgue-Stieltjes integrals. □

Proposition 6.65: 6.23.

If $\lambda_f(\alpha) < \infty$ for all $\alpha > 0$ and ϕ is a nonnegative Borel measurable function on $(0, \infty)$, then

$$\int_X \phi \circ |f| d\mu = -\int_0^\infty \phi(\alpha) d\lambda_f(\alpha)$$

Proof. If ν is the negative measure determined by λ_f, we have

$$\nu((a, b]) = \lambda_f(b) - \lambda_f(a) = -\mu(\{x \mid a < |f(x)| \leqslant b\}) = -\mu(|f|^{-1}((a, b]))$$

It follows that $\nu(E) = -\mu(|f|^{-1}(E))$ for all Borel sets $E \subset (0, \infty)$, by the uniqueness of extensions (Theorem 33). But this means that $\int_X \phi \circ |f| d\mu = -\int_0^\infty \phi(\alpha) d\lambda_f(\alpha)$ when ϕ is the characteristic function of a Borel set, and hence when ϕ is simple. The general case then follows by virtue of Theorem 18 and the monotone convergence theorem.

The case of this result in which we are most interested is $\phi(\alpha) = \alpha^p$, which gives

$$\int |f|^p d\mu = -\int_0^\infty \alpha^p d\lambda_f(\alpha)$$

A more useful form of this equation is obtained by integrating the right side by parts (Theorem 79) to obtain $\int |f|^p d\mu = p\int_0^\infty \alpha^{p-1}\lambda_f(\alpha)d\alpha$. The validity of this calculation is not clear unless we know that $\alpha^p\lambda_f(\alpha) \to 0$ as $\alpha \to 0$ and $\alpha \to \infty$; nonetheless, the conclusion is correct.

Proposition 6.66: 6.24.

If $0 < p < \infty$, then

$$\int |f|^p d\mu = p\int_0^\infty \alpha^{p-1}\lambda_f(\alpha)d\alpha$$

Proof. If $\lambda_f(\alpha) = \infty$ for some $\alpha > 0$, then both integrals are infinite. If not, and f is simple, then λ_f is bounded as $\alpha \to 0$ and vanishes for α sufficiently large, so the integration by parts described above works. (It is also easy to verify the formula directly in this case.) For the general case, let $\{g_n\}$ be a sequence of simple functions that increases to $|f|$; then the desired result is true for g_n, and it follows for f by Proposition 6.22c and the monotone convergence theorem.

A variant of the L^p spaces that turns up rather often is the following. If f is a measurable function on X and $0 < p < \infty$, we define

$$[f]_p = (\sup_{\alpha>0}\alpha^p\lambda_f(\alpha))^{1/p}$$

and we define weak L^p to be the set of all f such that $[f]_p < \infty$. $[\cdot]_p$ is not a norm; it is easily checked that $[cf]_p = |c|[f]_p$, but the triangle inequality fails. However, weak L^p is a topological vector space; see Folland Exercise 6.35.

The relationship between L^p and weak L^p is as follows. On the one hand,

$$L^p \subset \text{ weak } L^p, \quad \text{and} \quad [f]_p \leqslant \|f\|_p$$

(This is just a restatement of Chebyshev's inequality.) On the other hand, if we replace $\lambda_f(\alpha)$ by $([f]_p/\alpha)^p$ in the integral $p\int_0^\infty \alpha^{p-1}\lambda_f(\alpha)d\alpha$, which equals $\|f\|_p^p$, we obtain a constant times $\int_0^\infty \alpha^{-1}d\alpha$, which is divergent at both 0 and ∞– but just barely. One needs only slightly stronger estimates on λ_f near 0 and ∞ to obtain $f \in L^p$. (See also Folland Exercise 6.36.) The standard example of a function that is in weak L^p but not in L^p is $f(x) = x^{-1/p}$ on $(0, \infty)$ (with Lebesgue measure).

Frequently it is convenient to express a function as the sum of a "small" part and a "big" part. The following is a way of doing this that gives a simple formula for the distribution functions.

Proposition 6.67: 6.25.

If f is a measurable function and $A > 0$, let $E(A) = \{x \mid |f(x)| > A\}$, and set

$$h_A = f\chi_{X\smallsetminus E(A)} + A(\operatorname{sgn} f)\chi_{E(A)}, \quad g_A = f - h_A = (\operatorname{sgn} f)(|f| - A)\chi_{E(A)}$$

Then

$$\lambda_{g_A}(\alpha) = \lambda_f(\alpha + A), \quad \lambda_{h_A}(\alpha) = \begin{cases} \lambda_f(\alpha) & \text{if } \alpha < A \\ 0, & \text{if } \alpha \geqslant A. \end{cases}$$

The proof is left to the reader (Folland Exercise 6.37).

Exercise 6.68: Folland Exercise 6.35.

For any measurable f and g we have $[cf]_p = |c|[f]_p$ and $[f + g]_p \leqslant 2([f]_p^p + [g]_p^p)^{1/p}$; hence weak L^p is a vector space. Moreover, the "balls" $\{g \mid [g - f]_p < r\}$ $(r > 0, f \in$ weak $L^p)$ generate a topology on weak L^p that makes weak L^p into a topological vector space.

Exercise 6.69: Folland Exercise 6.36.

If $f \in$ weak L^p and $\mu(\{x \mid f(x) \neq 0\}) < \infty$, then $f \in L^q$ for all $q < p$. On the other hand, if $f \in$ (weak $L^p) \cap L^\infty$, then $f \in L^q$ for all $q > p$.

Solution. Suppose $f \in$ weak L^p, $0 < q < p$, and $\mu(\{|f| \neq 0\}) < \infty$. Define

$$E_n := \begin{cases} \{0 < |f| \leqslant 1\} & \text{if } n = 0, \\ \{2^{n-1} < |f| \leqslant 2^n\} & \text{if } n \in \mathbb{Z}_{\geqslant 1}. \end{cases}$$

Then $|f| = \sum_{n=0}^\infty \chi_{E_n}|f|$, so

$$\begin{aligned}
\|f\|_q^q = \int |f|^q \leqslant \int \Big|\sum_{n=0}^\infty 2^n \chi_{E_n}\Big|^q \leqslant \int \sum_{n=0}^\infty 2^{nq}\chi_{E_n} & \quad \text{(by the triangle inequality)} \\
= \sum_{n=0}^\infty 2^{nq}\mu(E_n) & \quad \text{(by the monotone convergence theorem for series)} \\
= \mu(E_0) + \sum_{n=1}^\infty 2^{nq}\lambda_f(2^{n-1}) & \quad \text{(since } E_n \subset \{|f| > 2^{n-1}\} \text{ and isolating } \mu(E_0)) \\
= \mu(E_0) + \sum_{n=1}^\infty 2^{nq}\lambda_f(2^{n-1}) & \quad \text{(since } [f]_p^p \geqslant 2^{(n-1)p}\lambda_f(2^{n-1}) \text{ by definition of } [f]_p) \\
= \mu(E_0) + \sum_{n=1}^\infty 2^{nq-(np-p)}[f]_p^p, & \\
= \mu(E_0) + \left(\frac{[f]_p}{2}\right)^p \sum_{n=1}^\infty (2^n)^{q-p}, &
\end{aligned}$$

which is finite since $E_0 \subset \{|f| \neq 0\}$—which by hypothesis has finite measure—and the infinite sum is a geometric series with ratio $2^{q-p} \in (-1, 1)$ since $q < p$, and thus converges.

Now instead suppose $f \in (\text{weak } L^p) \cap L^\infty$ and $p < q < \infty$. Since f is already L^∞, we can assume $q < \infty$. Define

$$E_n := \begin{cases} \{|f| > 1\} & \text{if } n = 0, \\ \{\frac{1}{2^n} < |f| \leqslant \frac{1}{2^{n-1}}\} & \text{if } n \in \mathbb{Z}_{\geqslant 1}. \end{cases}$$

Computing similarly to before, we have

$$\begin{aligned} \int |f|^q &\leqslant \int \Big(\sum_{n=0}^{\infty} (2^{1-n})^q \chi_{E_n}\Big) \\ &= \|f\|_\infty^q \mu(E_0) + \sum_{n=1}^{\infty} 2^{q-nq} \mu(E_n) \\ &\leqslant \|f\|_\infty^q \lambda_f(1) + \sum_{n=1}^{\infty} 2^{q-nq} \lambda_f(2^{-n}) \\ &\leqslant \|f\|_\infty^q [f]_p^p + \sum_{n=1}^{\infty} 2^{q-nq+np} [f]_p^p, \end{aligned}$$

which again is finite for the same reasons as before. Thus $f \in L^q$ for all $p < q \leqslant \infty$. □

Exercise 6.70: Folland Exercise 6.37.

Prove Proposition 67.

Exercise 6.71: Folland Exercise 6.38.

$f \in L^p$ if and only if $\sum_{-\infty}^{\infty} 2^{kp} \lambda_f(2^k) < \infty$.

Exercise 6.72: Folland Exercise 6.39.

If $f \in L^p$, then $\lim_{\alpha \to 0} \alpha^p \lambda_f(\alpha) = \lim_{\alpha \to \infty} \alpha^p \lambda_f(\alpha) = 0$. (First suppose f is simple.)

Exercise 6.73: Folland Exercise 6.40.

If f is a measurable function on X, its decreasing rearrangement is the function $f^* \colon (0, \infty) \to [0, \infty]$ defined by

$$f^*(t) = \inf\{\alpha \mid \lambda_f(\alpha) \leqslant t\} \quad (\text{where } \inf \varnothing = \infty)$$

(a) f^* is decreasing. If $f^*(t) < \infty$ then $\lambda_f(f^*(t)) \leqslant t$, and if $\lambda_f(\alpha) < \infty$ then $f^*(\lambda_f(\alpha)) \leqslant \alpha$.
(b) $\lambda_f = \lambda_{f^*}$, where λ_{f^*} is defined with respect to Lebesgue measure on $(0, \infty)$.
(c) If $\lambda_f(\alpha) < \infty$ for all $\alpha > 0$ and $\lim_{\alpha \to \infty} \lambda_f(\alpha) = 0$ (so that $f^*(t) < \infty$ for all $t > 0$), and ϕ is a nonnegative measurable function on $(0, \infty)$, then $\int_X \phi \circ |f| d\mu = \int_0^\infty \phi \circ f^*(t) dt$. In particular, $\|f\|_p = \|f^*\|_p$ for $0 < p < \infty$.
(d) If $0 < p < \infty$, $[f]_p = \sup_{t > 0} t^{1/p} f^*(t)$.
(e) The name "rearrangement" for f^* comes from the case where f is a nonnegative function on $(0, \infty)$. To see why it is appropriate, pick a step function on $(0, \infty)$

assuming four or five different values and draw the graphs of f and f^*.

6.5 Interpolation of L^p Spaces

If $1 \leqslant p < q < r \leqslant \infty$, then $(L^p \cap L^r) \subset L^q \subset (L^p + L^r)$, and it is natural to ask whether a linear operator T on $L^p + L^r$ that is bounded on both L^p and L^r is also bounded on L^q. The answer is affirmative, and this result can be generalized in various ways. The two fundamental theorems on this question are the Riesz-Thorin and Marcinkiewicz interpolation theorems, which we present in this section. We begin with the Riesz-Thorin theorem, whose proof is based on the following result from complex function theory.

Lemma 6.74: 6.26: The Three Lines Lemma.

Let ϕ be a bounded continuous function on the strip $0 \leqslant \operatorname{Re} z \leqslant 1$ that is holomorphic on the interior of the strip. If $|\phi(z)| \leqslant M_0$ for $\operatorname{Re} z = 0$ and $|\phi(z)| \leqslant M_1$ for $\operatorname{Re} z = 1$, then $|\phi(z)| \leqslant M_0^{1-t} M_1^t$ for $\operatorname{Re} z = t$, $0 < t < 1$.

Proof. For $\varepsilon > 0$ let $\phi_\varepsilon(z) = \phi(z) M_0^{z-1} M_1^{-z} \exp(\varepsilon z(z-1))$. Then ϕ_ε satisfies the hypotheses of the lemma with M_0 and M_1 replaced by 1, and also $|\phi_\varepsilon(z)| \to 0$ as $|\operatorname{Im} z| \to \infty$. Thus $|\phi_\varepsilon(z)| \leqslant 1$ on the boundary of the rectangle $0 \leqslant \operatorname{Re} z \leqslant 1$, $-A \leqslant \operatorname{Im} z \leqslant A$ provided that A is large, and the maximum modulus principle therefore implies that $|\phi_\varepsilon(z)| \leqslant 1$ on the strip $0 \leqslant \operatorname{Re} z \leqslant 1$. Letting $\varepsilon \to 0$, we obtain the desired result:

$$|\phi(z)| M_0^{t-1} M_1^{-t} = \lim_{\varepsilon \to 0} |\phi_\varepsilon(z)| \leqslant 1 \text{ for } \operatorname{Re} z = t.$$

□

Theorem 6.75: 6.27: The Riesz-Thorin Interpolation Theorem.

Suppose that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{M}, \nu)$ are measure spaces and $p_0, p_1, q_0, q_1 \in [1, \infty]$. If $q_0 = q_1 = \infty$, suppose also that ν is semifinite. For $0 < t < 1$, define p_t and q_t by

$$\frac{1}{p_t} = \frac{1-t}{p_0} + \frac{t}{p_1}, \qquad \frac{1}{q_t} = \frac{1-t}{q_0} + \frac{t}{q_1}$$

If T is a linear map from $L^{p_0}(\mu) + L^{p_1}(\mu)$ into $L^{q_0}(\nu) + L^{q_1}(\nu)$ such that $\|Tf\|_{q_0} \leqslant M_0 \|f\|_{p_0}$ for $f \in L^{p_0}(\mu)$ and $\|Tf\|_{q_1} \leqslant M_1 \|f\|_{p_1}$ for $f \in L^{p_1}(\mu)$, then $\|Tf\|_{q_t} \leqslant M_0^{1-t} M_1^t \|f\|_{p_t}$ for $f \in L^{p_t}(\mu), 0 < t < 1$.

Proof. To begin with, we observe that the case $p_0 = p_1$ follows from Proposition 16: If $p = p_0 = p_1$, then

$$\|Tf\|_{q_t} \leqslant \|Tf\|_{q_0}^{1-t} \|Tf\|_{q_1}^t \leqslant M_0^{1-t} M_1^t \|f\|_p$$

Thus we may assume that $p_0 \neq p_1$, and in particular that $p_t < \infty$ for $0 < t < 1$.

Let Σ_X (resp. Σ_Y) be the space of all simple functions on X (resp. Y) that vanish outside sets of finite measure. Then $\Sigma_X \subset L^p(\mu)$ for all p and Σ_X is dense in $L^p(\mu)$ for

$p < \infty$, by Proposition 9; similarly for Σ_Y. The main part of the proof consists of showing that $\|Tf\|_{q_t} \leqslant M_0^{1-t} M_1^t \|f\|_{p_t}$ for all $f \in \Sigma_X$. However, by Theorem 37,

$$\|Tf\|_{q_t} = \sup\left\{ \left| \int (Tf) g d\nu \right| \,\middle|\, g \in \Sigma_Y \text{ and } \|g\|_{q_t'} = 1 \right\}$$

where q_t' is the conjugate exponent to q_t. (Note that $Tf \in L^{q_0} \cap L^{q_1}$, so $\{y \mid Tf(y) \neq 0\}$ must be σ-finite unless $q_0 = q_1 = \infty$; hence the hypotheses of Theorem 37 are satisfied.) Moreover, we may assume that $f \neq 0$ and rescale f so that $\|f\|_{p_t} = 1$. We therefore wish to establish the following claim: - If $f \in \Sigma_X$ and $\|f\|_{p_t} = 1$, then $\left|\int (Tf) g d\nu\right| \leqslant M_0^{1-t} M_1^t$ for all $g \in \Sigma_Y$ such that $\|g\|_{q_t'} = 1$.

Let $f = \sum_1^m c_j \chi_{E_j}$ and $g = \sum_1^n d_k \chi_{F_k}$ where the E_js and the F_ks are disjoint in X and Y and the c_js and d_ks are nonzero. Write c_j and d_k in polar form: $c_j = |c_j| e^{i\theta_j}$, $d_k = |d_k| e^{i\psi_k}$. Also, let

$$\alpha(z) = (1-z) p_0^{-1} + z p_1^{-1}, \quad \beta(z) = (1-z) q_0^{-1} + z q_1^{-1}$$

thus $\alpha(t) = p_t^{-1}$ and $\beta(t) = q_t^{-1}$ for $0 < t < 1$. Fix $t \in (0,1)$; we have assumed that $p_t < \infty$ and hence $\alpha(t) > 0$, so we may define

$$f_z = \sum\nolimits_1^m |c_j|^{\alpha(z)/\alpha(t)} e^{i\theta_j} \chi_{E_j}$$

If $\beta(t) < 1$, we define

$$g_z = \sum\nolimits_1^n |d_k|^{(1-\beta(z))/(1-\beta(t))} e^{i\psi_k} \chi_{F_k}$$

while if $\beta(t) = 1$ we define $g_z = g$ for all z. (We henceforth assume that $\beta(t) < 1$ and leave the easy modification for $\beta(t) = 1$ to the reader.) Finally, we set

$$\phi(z) = \int (Tf_z) g_z d\nu$$

Thus,

$$\phi(z) = \sum\nolimits_{j,k} A_{jk} |c_j|^{\alpha(z)/\alpha(t)} |d_k|^{(1-\beta(z))/(1-\beta(t))}$$

where

$$A_{jk} = e^{i(\theta_j + \psi_k)} \int (T\chi_{E_j}) \chi_{F_k} d\nu$$

so that ϕ is an entire holomorphic function of z that is bounded in the strip $0 \leqslant \operatorname{Re} z \leqslant 1$. Since $\int (Tf) g d\nu = \phi(t)$, by the three lines lemma it will suffice to show that $|\phi(z)| \leqslant M_0$ for $\operatorname{Re} z = 0$ and $|\phi(z)| \leqslant M_1$ for $\operatorname{Re} z = 1$. However, since

$$\alpha(is) = p_0^{-1} + is(p_1^{-1} - p_0^{-1}), \quad 1 - \beta(is) = (1 - q_0^{-1}) - is(q_1^{-1} - q_0^{-1})$$

for $s \in \mathbb{R}$, we have

$$|f_{is}| = |f|^{\operatorname{Re}[\alpha(is)/\alpha(t)]} = |f|^{p_t/p_0}, \quad |g_{is}| = |g|^{\operatorname{Re}[(1-\beta(is))/(1-\beta(t))]} = |g|^{q_t'/q_0'}$$

Therefore, by Hölder's inequality,

$$|\phi(is)| \leqslant \|Tf_{is}\|_{q_0} \|g_{is}\|_{q_0'} \leqslant M_0 \|f_{is}\|_{p_0} \|g_{is}\|_{q_0'} = M_0 \|f\|_{p_t} \|g\|_{q_t'} = M_0$$

A similar calculation shows that $|\phi(1+is)| \leqslant M_1$, so the claim is proved.

We have now shown that $\|Tf\|_{q_t} \leqslant M_0^{1-t}M_1^t\|f\|_{p_t}$ for $f \in \Sigma_X$, so in view of Proposition 9, $T|\Sigma_X$ has a unique extension to $L^{p_t}(\mu)$ satisfying the same estimate there. It remains to show that this extension is T itself, that is, that T satisfies this estimate for all $f \in L^{p_t}(\mu)$. Given such an f, choose a sequence $\{f_n\}$ in Σ_X such that $|f_n| \leqslant |f|$ and $f_n \to f$ pointwise. Also, let $E = \{x : |f(x)| > 1\}, g = f\chi_E$, $g_n = f_n\chi_E, h = f - g$, and $h_n = f_n - g_n$. Then if $p_0 < p_1$ (which we may assume, by relabeling the ps), we have $g \in L^{p_0}(\mu), h \in L^{p_1}(\mu)$, and by the dominated convergence theorem, $\|f_n - f\|_{p_t} \to 0, \|g_n - g\|_{p_0} \to 0$, and $\|h_n - h\|_{p_1} \to 0$. Hence $\|Tg_n - Tg\|_{q_0} \to 0$ and $\|Th_n - Th\|_{q_1} \to 0$, so by passing to a suitable subsequence we may assume that $Tg_n \to Tg$ a.e. and $Th_n \to Th$ a.e. (Folland Exercise 6.9). But then $Tf_n \to Tf$ a.e., so by Fatou's lemma,

$$\|Tf\|_{q_t} \leqslant \liminf\|Tf_n\|_{q_t} \leqslant \liminf M_0^{1-t}M_1^t\|f_n\|_{p_t} = M_0^{1-t}M_1^t\|f\|_{p_t},$$

and we are done. □

The conclusion of the Riesz-Thorin theorem can be restated in a slightly stronger form. Let $M(t)$ be the operator norm of T as a map from $L^{p_t}(\mu)$ to $L^{q_t}(\nu)$. We have shown that $M(t) \leqslant M_0^{1-t}M_1^t$. It is possible for strict inequality to hold; however, if $0 < s < t < u < 1$ and $t = (1-\tau)s + \tau u$, the theorem may be applied again to show that $M(t) \leqslant M(s)^{1-\tau}M(u)^\tau$. In short, the conclusion is that $\log M(t)$ is a convex function of t.

We now turn to the Marcinkiewicz theorem, for which we need some more terminology. Let T be a map from some vector space $\mathcal{D}$ of measurable functions on $(X, \mathcal{D}, \mu)$ to the space of all measurable functions on $(Y, \mathcal{D}, \nu)$. T is called ***sublinear*** if $|T(f+g)| \leqslant |Tf| + |Tg|$ and $|T(cf)| = c|Tf|$ for all $f, g \in \mathcal{D}$ and $c > 0$. Now let $p, q \in [1, \infty]$.

- A sublinear map T is ***strong type*** $(\boldsymbol{p}, \boldsymbol{q})$ if $L^p(\mu) \subset \mathcal{D}$ and T maps $L^p(\mu)$ into $L^q(\nu)$, and there exists $C > 0$ such that $\|Tf\|_q \leqslant C\|f\|_p$ for all $f \in L^p(\mu)$.
- A sublinear map T is ***weak type*** $(\boldsymbol{p}, \boldsymbol{q})$ if $L^p(\mu) \subset \mathcal{D}, T$ maps $L^p(\mu)$ into weak $L^q(\nu)$, and there exists $C > 0$ such that $[Tf]_q \leqslant C\|f\|_p$ for all $f \in L^p(\mu)$. Also, we shall say that T is weak type (p, ∞) if and only if T is strong type (p, ∞).

Theorem 6.76: 6.28: The Marcinkiewicz Interpolation Theorem.

Suppose that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{M}, \nu)$ are measure spaces; p_0, p_1, q_0, q_1 are elements of $[1, \infty]$ such that $p_0 \leqslant q_0$, $p_1 \leqslant q_1$, and $q_0 \neq q_1$; and

$$\frac{1}{p} = \frac{1-t}{p_0} + \frac{t}{p_1} \text{ and } \frac{1}{q} = \frac{1-t}{q_0} + \frac{t}{q_1}, \qquad \text{where } 0 < t < 1$$

If T is a sublinear map from $L^{p_0}(\mu) + L^{p_1}(\mu)$ to the space of measurable functions on Y that is weak types (p_0, q_0) and (p_1, q_1), then T is strong type (p, q). More precisely, if $[Tf]_{q_j} \leqslant C_j\|f\|_{p_j}$ for $j = 0, 1$, then $\|Tf\|_q \leqslant B_p\|f\|_p$ where B_p depends only on p_j, q_j, C_j in addition to p; and for $j = 0, 1, B_p|p - p_j|$ (resp. B_p) remains bounded as

$p \to p_j$ if $p_j < \infty$ (resp. $p_j = \infty$).

Proof. The case $p_0 = p_1$ is easy and is left to the reader (Folland Exercise 6.42). Without loss of generality we may therefore assume that $p_0 < p_1$, and for the time being we also assume that $q_0 < \infty$ and $q_1 < \infty$ (whence also $p_0 < p_1 < \infty$). Given $f \in L^p(\mu)$ and $A > 0$, let g_A and h_A be as in Proposition 67. Then by Propositions 6.24 and 6.25,

$$\begin{aligned}\int |g_A|^{p_0} d\mu &= p_0 \int_0^\infty \beta^{p_0-1}\lambda_{g_A}(\beta)d\beta = p_0 \int_0^\infty \beta^{p_0-1}\lambda_f(\beta+A)d\beta \\ &= p_0 \int_A^\infty (\beta-A)^{p_0-1}\lambda_f(\beta)d\beta \leqslant p_0 \int_A^\infty \beta^{p_0-1}\lambda_f(\beta)d\beta \\ \int |h_A|^{p_1} d\mu &= p_1 \int_0^\infty \beta^{p_1-1}\lambda_{h_A}(\beta)d\beta = p_1 \int_0^A \beta^{p_1-1}\lambda_f(\beta)d\beta\end{aligned}$$

Likewise,

$$\int |Tf|^q d\nu = q\int_0^\infty \alpha^{q-1}\lambda_{Tf}(\alpha)d\alpha = 2^q q \int_0^\infty \alpha^{q-1}\lambda_{Tf}(2\alpha)d\alpha. \tag{6.30}$$

Since T is sublinear, by Proposition 64(d) we have

$$\lambda_{Tf}(2\alpha) \leqslant \lambda_{Tg_A}(\alpha) + \lambda_{Th_A}(\alpha)$$

This is true for all $\alpha > 0$ and $A > 0$, so we may take A to depend on α. We now make a specific choice of A. Namely, it follows from the equations defining p and q that

$$\frac{p_0(q_0-q)}{q_0(p_0-p)} = \frac{p^{-1}(q^{-1}-q_0^{-1})}{q^{-1}(p^{-1}-p_0^{-1})} = \frac{p^{-1}(q^{-1}-q_1^{-1})}{q^{-1}(p^{-1}-p_1^{-1})} = \frac{p_1(q_1-q)}{q_1(p_1-p)}$$

we denote the common value of these quantities by σ, and we take $A = \alpha^\sigma$. Then by (6.29), (6.30), (6.31), and the weak type estimates on T,

$$\begin{aligned}\|Tf\|_q^q &\leqslant 2^q q\int_0^\infty \alpha^{q-1}[(C_0\|g_A\|_{p_0}/\alpha)^{q_0} + (C_1\|h_A\|_{p_1}/\alpha)^{q_1}]d\alpha \\ &\leqslant 2^q q C_0^{q_0} p_0^{q_0/p_0}\int_0^\infty \alpha^{q-q_0-1}\left[\int_{\alpha^\sigma}^\infty \beta^{p_0-1}\lambda_f(\beta)d\beta\right]^{q_0/p_0} d\alpha \\ &\quad + 2^q q C_1^{q_1} p_1^{q_1/p_1}\int_0^\infty \alpha^{q-q_1-1}\left[\int_0^{\alpha^\sigma} \beta^{p_1-1}\lambda_f(\beta)d\beta\right]^{q_1/p_1} d\alpha \\ &= \sum_{j=0}^1 2^q q C_j^{p_j} p_j^{q_j/p_j}\int_0^\infty\left[\int_0^\infty \phi_j(\alpha,\beta)d\beta\right]^{q_j/p_j} d\alpha\end{aligned}$$

where, denoting by χ_0 and χ_1 the characteristic functions of $\{(\alpha,\beta) \mid \beta > \alpha^\sigma\}$ and $\{(\alpha,\beta) \mid \beta < \alpha^\sigma\}$,

$$\phi_j(\alpha,\beta) = \chi_j(\alpha,\beta)\alpha^{(q-q_j-1)p_j/q_j}\beta^{p_j-1}\lambda_f(\beta)$$

Since $q_0/p_0 \geqslant 1$ and $q_1/p_1 \geqslant 1$, we may apply Minkowski's inequality for integrals to obtain

$$\int_0^\infty \left[\int_0^\infty \phi_j(\alpha,\beta)d\beta\right]^{q_j/p_j} d\alpha \leqslant \left[\int_0^\infty \left[\int_0^\infty \phi_j(\alpha,\beta)^{q_j/p_j} d\alpha\right]^{p_j/q_j} d\beta\right]^{q_j/p_j}.$$

Let $\tau = 1/\sigma$. If $q_1 > q_0$, then $q - q_0$ and σ are positive and the inequality $\beta > \alpha^\sigma$ is equivalent to $\alpha < \beta^\tau$, so

$$\begin{aligned}\int_0^\infty \left[\int_0^\infty \phi_0(\alpha,\beta)^{q_0/p_0} d\alpha^{p_0/q_0} d\beta\right. &= \int_0^\infty \left[\int_0^{\beta^\tau} \alpha^{q-q_0-1} d\alpha\right]^{p_0/q_0} \beta^{p_0-1}\lambda_f(\beta)d\beta \\ &= (q-q_0)^{-p_0/q_0} \int_0^\infty \beta^{p_0-1+p_0(q-q_0)/q_0\sigma}\lambda_f(\beta)d\beta \\ &= (q-q_0)^{-p_0/q_0} \int_0^\infty \beta^{p-1}\lambda_f(\beta)d\beta \\ &= |q-q_0|^{-p_0/q_0} p^{-1}\|f\|_p^p\end{aligned}$$

where we have used (6.32) to simplify the exponent of β. On the other hand, if $q_1 < q_0$, then $q - q_0$ and σ are negative and the inequality $\beta > \alpha^\sigma$ is equivalent to $\alpha > \beta^\tau$, so as above,

$$\begin{aligned}\int_0^\infty \left[\int_0^\infty \phi_0(\alpha,\beta)^{q_0/p_0} d\alpha\right]^{p_0/q_0} d\beta &= \int_0^\infty \left[\int_{\beta^\tau}^\infty \alpha^{q-q_0-1} d\alpha\right]^{p_0/q_0} \beta^{p_0-1}\lambda_f(\beta)d\beta \\ &= (q_0-q)^{-p_0/q_0} \int_0^\infty \beta^{p-1}\lambda_f(\beta)d\beta \\ &= |q-q_0|^{-p_0/q_0} p^{-1}\|f\|_p^p\end{aligned}$$

A similar calculation shows that

$$\int_0^\infty \left[\int_0^\infty \phi_1(\alpha,\beta)^{q_1/p_1} d\alpha\right]^{p_1/q_1} d\beta = |q-q_1|^{-p_1/q_1} p^{-1}\|f\|_p^p$$

Combining these results with (6.33) and (6.34), we see that

$$\sup\{\|Tf\|_q \mid \|f\|_p = 1\} \leqslant B_p = 2q^{1/q}\left[\sum_{j=0}^1 C_j^{q_j}(p_j/p)^{q_j/p_j}|q-q_j|^{-1}\right]^{1/q}.$$

But since $|T(cf)| = c|Tf|$ for $c > 0$, this implies that $\|Tf\|_q \leqslant B_p\|f\|_p$ for all $f \in L^p(\mu)$, and we are done. (The verification of the asserted properties of B_p is left as an easy exercise.)

It remains to show how to modify this argument to deal with the exceptional cases $q_0 = \infty$ or $q_1 = \infty$. We distinguish three cases.

Case I: $p_1 = q_1 = \infty$ (so $p_0 \leqslant q_0 < \infty$). Instead of taking $A = \alpha^\sigma$ in the decomposition of f, we take $A = \alpha/C_1$. Then $\|Th_A\|_\infty \leqslant C_1\|h_A\|_\infty \leqslant \alpha$, so $\lambda_{Th_A}(\alpha) = 0$, and we obtain

(6.33) with $\phi_1 = 0$ and α^σ replaced by α/C_1 in the definition of ϕ_0. The same argument as above then gives

$$\|Tf\|_q \leqslant 2[qC_0^{q_0}C_1^{q-q_0}(p_0/p)^{q_0/p_0}|q-q_0|^{-1}]^{1/q}\|f\|_p$$

Case II: $p_0 < p_1 < \infty, q_0 < q_1 = \infty$. Again the idea is to choose A so that $\lambda_{Th_A}(\alpha) = 0$, and the proper choice is $A = (\alpha/d)^\sigma$ where $d = C_1[p_1\|f\|_p^p/p]^{1/p_1}$ and $\sigma = p_1/(p_1-p)$ (the limiting value of the σ defined by (6.32) as $q_1 \to \infty$). Indeed, since $p_1 > p$, we have

$$\begin{aligned}\|Th_A\|_\infty^{p_1} &\leqslant C_1^{p_1}\|h_A\|_{p_1}^{p_1} = C_1^{p_1}p_1\int_0^A \alpha^{p_1-1}\lambda_f(\alpha)d\alpha \\ &\leqslant C_1^{p_1}p_1A^{p_1-p}\int_0^A \alpha^{p-1}\lambda_f(\alpha)d\alpha = C_1^{p_1}\frac{p_1}{p}\left[\frac{\alpha}{d}\right]^{p_1}\|f\|_p^p = \alpha^{p_1}.\end{aligned}$$

As in Case I, then, we find that $\phi_1 = 0$ in (6.33) and the integral involving ϕ_0 is majorized by a constant B_p when $\|f\|_p = 1$, which yields the desired result.

Case III: $p_0 < p_1 < \infty, q_1 < q_0 = \infty$. The argument is essentially the same as in Case II, except that we take $A = (\alpha/d)^\sigma$ with d chosen so that $\lambda_{Tg_A}(\alpha) = 0$. □

The lengthy formulas in this proof may seem daunting, but the ideas are reasonably simple. To elucidate them, we recommend the exercise of writing out the proof for two special (but important) cases: (i) $p_0 = q_0 = 1, p_1 = q_1 = 2$, and (ii) $p_0 = q_0 = 1$, $p_1 = q_1 = \infty$.

Let us compare our two interpolation theorems. The Marcinkiewicz theorem requires some restrictions on p_j and q_j that are not present in the Riesz-Thorin theorem; these restrictions, however, are satisfied in all the interesting applications. Apart from this, the hypotheses of the Marcinkiewicz theorem are weaker: T is allowed to be sublinear rather than linear, and it needs only to satisfy weak-type estimates at the endpoints. The conclusion in both cases is that T is bounded from $L^p(\mu)$ to $L^q(\nu)$, but the Riesz-Thorin theorem produces a much sharper estimate for the operator norm of T. Thus neither theorem includes the other.

We conclude with two applications of the Marcinkiewicz theorem. The first one concerns the Hardy-Littlewood maximal operator H discussed in Folland Section 3.4,

$$Hf(x) = \sup_{r>0}\frac{1}{m(B(r,x))}\int_{B(r,x)}|f(y)|dy \quad (f \in L^1_{\text{loc}}(\mathbb{R}^n))$$

H is obviously sublinear and satisfies $\|Hf\|_\infty \leqslant \|f\|_\infty$ for all $f \in L^\infty$. Moreover, Theorem 44 says precisely that H is weak type $(1,1)$. We conclude:

Corollary 6.77: 6.35.

There is a constant $C > 0$ such that if $1 < p < \infty$ and $f \in L^p(\mathbb{R}^n)$, then

$$\|Hf\|_p \leqslant C\frac{p}{p-1}\|f\|_p$$

Our second application is a theorem on integral operators related to Theorem 50.

Theorem 6.78: 6.36.

Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{M}, \nu)$ are σ-finite measure spaces, and $1 < q < \infty$. Let K be a measurable function on $X \times Y$ such that, for some $C > 0$, we have $[K(x, \cdot)]_q \leqslant C$ for a.e. $x \in X$ and $[K(\cdot, y)]_q \leqslant C$ for a.e. $y \in Y$. If $1 \leqslant p < \infty$ and $f \in L^p(\nu)$, the integral

$$Tf(x) = \int K(x, y) f(y) d\nu(y)$$

converges absolutely for a.e. $x \in X$, and the operator T thus defined is weak type $(1, q)$ and strong type (p, r) for all p, r such that $1 < p < r < \infty$ and $p^{-1} + q^{-1} = r^{-1} + 1$. More precisely, there exist constants B_p independent of K such that

$$[Tf]_q \leqslant B_1 C \|f\|_1, \quad \|Tf\|_r \leqslant B_p C \|f\|_p \quad (p > 1, r^{-1} = p^{-1} + q^{-1} - 1 > 0)$$

Proof. Let p', q' be the conjugate exponents to p, q; then

$$r^{-1} = p^{-1} + q^{-1} - 1 = p^{-1} - (q')^{-1} = q^{-1} - (p')^{-1}$$

so $p < q'$ and $q < p'$. Suppose $0 \neq f \in L^p(1 \leqslant p < q')$; by multiplying f and K by constants, we may assume that $\|f\|_p = C = 1$. Given a positive number A whose value will be fixed later, define

$$E = \{(x, y) \mid |K(x, y)| > A\}, \quad K_1 = (\operatorname{sgn} K)(|K| - A)\chi_E, \quad K_2 = K - K_1,$$

and let T_1, T_2 be the operators corresponding to K_1, K_2. Then by Propositions 66 and 67, since $q > 1$ we have

$$\int |K_1(x, y)| d\nu(y) = \int_0^\infty \lambda_{K(x,\cdot)}(\alpha + A) d\alpha \leqslant \int_A^\infty \alpha^{-q} d\alpha = \frac{A^{1-q}}{q - 1}$$

and likewise

$$\int |K_1(x, y)| d\mu(x) \leqslant \frac{A^{1-q}}{q - 1}$$

Hence, by Theorem 50, the integral defining $T_1 f(x)$ converges for a.e. x and

$$\|T_1 f\|_p \leqslant \frac{A^{1-q}}{q - 1} \|f\|_p = \frac{A^{1-q}}{q - 1}$$

Similarly, since $q < p'$,

$$\begin{aligned} \int |K_2(x, y)|^{p'} d\nu(y) &= p' \int_0^A \alpha^{p'-1} \lambda_{K(x,\cdot)}(\alpha) d\alpha \\ &\leqslant p' \int_0^A \alpha^{p'-1-q} d\alpha = \frac{p' A^{p'-q}}{p' - q} \end{aligned}$$

Therefore, by Hölder's inequality, the integral defining $T_2f(x)$ converges for every x, and

$$\|T_2f\|_\infty \leqslant \left[\frac{p'A^{p'-q}}{p'-q}\right]^{1/p'}\|f\|_p = \left[\frac{r}{q}\right]^{1/p'}A^{q/r}$$

We have thus established that $Tf = T_1f + T_2f$ is well defined a.e.

Next, given $\alpha > 0$, we wish to estimate $\lambda_{Tf}(\alpha)$. But by Proposition 64(d),

$$\lambda_{Tf}(\alpha) \leqslant \lambda_{T_1f}\left(\frac{1}{2}\alpha\right) + \lambda_{T_2f}\left(\frac{1}{2}\alpha\right)$$

and by (6.38), if we choose

$$A = \left[\frac{\alpha}{2}\right]^{r/q}\left[\frac{q}{r}\right]^{r/qp'}$$

we will have $\|T_2f\|_\infty \leqslant \frac{1}{2}\alpha$, so that $\lambda_{T_2f}\left(\frac{1}{2}\alpha\right) = 0$. With this choice of A, then, by (6.37) and Chebyshev's inequality we obtain

$$\begin{aligned}\lambda_{Tf}(\alpha) \leqslant \lambda_{T_1f}\left(\frac{1}{2}\alpha\right) &\leqslant \left[\frac{2\|T_1f\|_p}{\alpha}\right]^p \leqslant \left[\frac{2A^{1-q}}{(q-1)\alpha}\right]^p \\ &= \frac{2^{p-(1-q)pr/q}}{(q-1)^p}\left[\frac{q}{r}\right]^{(1-q)pr/qp'}\alpha^{-p+(1-q)pr/q} = C_p\left[\frac{\|f\|_p}{\alpha}\right]^r\end{aligned}$$

because $\|f\|_p = 1$ and

$$\frac{(1-q)pr}{q} - p = p\left(\frac{-r}{q'} - 1\right) = -p \cdot \frac{r}{p} = -r$$

A simple homogeneity argument now yields the estimate $\lambda_{Tf}(\alpha) \leqslant C_p(\|f\|_p/\alpha)^r$ with no restriction on $\|f\|_p$, so we have shown that T is weak type (p, r), and in particular (for $p = 1$) weak type $(1, q)$.

Finally, given $p \in (1, q')$, choose $\tilde{p} \in (p, q')$ and define $\tilde{r}$ by $\tilde{r}^{-1} = \tilde{p}^{-1} - (q')^{-1}$. Then T is weak types $(1, q)$ and $(\tilde{p}, \tilde{r})$, so it follows from the Marcinkiewicz theorem that T is strong type (p, r). □

Exercise 6.79: Folland Exercise 6.41.

Suppose $1 < p \leqslant \infty$ and $p^{-1} + q^{-1} = 1$. If T is a bounded operator on L^p such that $\int(Tf)g = \int f(Tg)$ for all $f, g \in L^p \cap L^q$, then T extends uniquely to a bounded operator on L^r for all r in $[p, q]$ (if $p < q$) or $[q, p]$ (if $q < p$). If $p = \infty$, further assume that μ is semifinite.

Solution. Let $p \in (1, \infty]$, let $q = (p-1)/p$, let Σ be the set of simple functions that vanish outside a set of finite measure, and let r lie in the closed interval between p and q.

Claim 80. *T maps $L^p \cap L^q$ into L^q and is bounded as a map $L^p \cap L^q \to L^q$.*

Proof. Let $f \in L^p \cap L^q$. Then $Tf \in L^p$ by hypothesis. Thus if $p < \infty$ then $|Tf|^p \in L^1$ (since $Tf \in L^p$), so $\{|Tf|^p \neq 0\} = \{Tf \neq 0\}$ is σ-finite by Folland Proposition 2.23(a). On the other hand, if $p = \infty$ then μ is semifinite by hypothesis. In either case, it follows from Folland Theorem 6.14 that

$$\|Tf\|_q = \sup\left\{\left|\int g(Tf)\right| \;\middle|\; g \in \Sigma \text{ and } \|g\|_p = 1\right\}, \tag{6.80.1}$$

so it suffices to show the right-hand side is finite. To that end, suppose $g \in \Sigma$ and $\|g\|_p = 1$. We have $g \in L^q$ since $g \in \Sigma$, so in particular $g \in L^p \cap L^q$. Then

$$\begin{aligned}\left|\int g(Tf)\right| &= \left|\int f(Tg)\right| && \text{(by our hypothesis on } T\text{)}\\ &\leqslant \|f\|_p \|Tg\|_q && \text{(by Hölder's inequality)}\\ &\leqslant \|f\|_p \|T\|_{L^p \to L^p} \|g\|_p \\ &\leqslant \|f\|_p \|T\|_{L^p \to L^p} && \text{(since } \|g\|_p = 1\text{)}.\end{aligned}$$

Our above estimate is independent of our choice of g, so by Equation (6.80.1)

$$\|Tf\|_q \leqslant \|T\|_{L^p \to L^p} \|f\|_p.$$

Thus T maps $L^p \cap L^q$ into L^q and is bounded as a map $(L^p \cap L^q, \|-\|_p) \to (L^q, \|-\|_q)$. □

Claim 81. *The map*

$$\begin{aligned}\widetilde{T} \colon L^p + L^q &\longrightarrow L^p + L^q,\\ f + g = h &\longmapsto \widetilde{T}g := Tf + \lim_{n\to\infty} Tg_n,\end{aligned}$$

where $\{g_n\}_{n=1}^\infty \subset L^p \cap L^q$ *and* $g_n \to g$ *in* L^q*, is a well-defined bounded linear operator.*

Proof.

- $\widetilde{T}$ *is well-defined*: Let $g \in L^p + L^q$. Since $L^p \cap L^q$ is dense in $L^p + L^q$ (because $L^p \cap L^q$ contains Σ, which is a dense subset in both L^p and L^q), such an approximating sequence $\{g_n\}_{n=1}^\infty$ as in the claim exists in $L^p \cap L^q$.

  Next we show $\widetilde{T}$ is independent of the choice of sequence $\{g_n\}_{n=1}^\infty \subset L^p \cap L^q$. Since $\{g_n\}_{n=1}^\infty$ is Cauchy in L^q and T is bounded as a map $L^p \cap L^q \to L^q$ by the first claim,

  $$\|Tg_n - Tg_m\|_q = \|T(g_n - g_m)\|_q \leqslant \|T\|_{L^p \to L^q} \|g_n - g_m\|_q \to 0$$

  as $n, m \to \infty$. By uniqueness of the limit (as L^q is a Banach space), we conclude $\widetilde{T}g$ is independent of the choice of approximating sequence sequence.
- $\widetilde{T}$ *is linear*: We are given $\widetilde{T}$ is linear on L^q, so it suffices to show linearity on L^p. Suppose $g, g' \in L^p \cap L^q$, $\alpha \in \mathbb{C}$, $\{g_n\}_{n=1}^\infty, \{g'_n\}_{n=1}^\infty \subset L^p \cap L^q$, and $g_n \to g, g'_n \to g'$ in L^q. Then

  $$\begin{aligned}\widetilde{T}(\alpha g + g') &= \lim_{n\to\infty} T(\alpha g + g')\\ &= \lim_{n\to\infty} (\alpha T g_n + T g'_n) && \text{(by linearity of } T\text{)}\\ &= \alpha \lim_{n\to\infty} Tg_n + \lim_{n\to\infty} Tg'_n && \text{(by linearity of limits that exist)}\end{aligned}$$

$$= \alpha\widetilde{T}g + \widetilde{T}g' \qquad \text{(by definition of } \widetilde{T}\text{)}.$$

Hence $\widetilde{T}$ is linear.

- *$\widetilde{T}$ is bounded as a map $L^q \to L^q$*: Let $g \in L^q$ and let $\{g_n\}_{n=1}^\infty \subset L^p \cap L^q$ such that $g_n \to g$ in L^q. Since $q < \infty$ by hypothesis, we can write

$$\begin{aligned}
\|\widetilde{T}g\|_q^q = \int |\widetilde{T}g|^q &= \int \Big|\lim_{n\to\infty} Tg_n\Big|^q \\
&= \int \lim_{n\to\infty} |Tg_n|^q && \text{(by continuity of } \mathbb{R} \ni x \mapsto |x|^q \in \mathbb{R}\text{)} \\
&\leqslant \liminf_{n\to\infty} \|Tg_n\|_q^q && \text{(by Fatou's lemma)} \\
&\leqslant \|T\|_{L^p\to L^q}^q \liminf_{n\to\infty} \|g_n\|_q^q \\
& \qquad \text{(since } T \text{ is bounded as an operator } L^p \cap L^q \to L^q\text{)} \\
&= \|T\|_{L^p\to L^q}^q \lim_{n\to\infty} \|g_n\|_q^q \\
& \qquad \text{(since } \lim_{n\to\infty} \|g_n\|_q^q \text{ exists, hence equals the liminf; see below)} \\
&= \|T\|_{L^p\to L^q}^q \|g\|_q^q.
\end{aligned}$$

The penultimate equality here follows from the fact $g_n \to g$ in L^q, since for all $\varepsilon > 0$ and all sufficiently large n,

$$\|g_n\|_q \leqslant \|g\|_q + \|g_n - g\|_q < \|Tg\|_q + \varepsilon < \infty;$$

taking the qth power, we obtain $\|g_n\|_q^q \leqslant (\|g\|_q + \varepsilon)^q < \infty$, so $\lim_{n\to\infty} \|g_n\|_q^q = \|g\|_q^q$. ☐

Claim 82. *$\widetilde{T}$ is the unique bounded operator on L^r for all r in the interval $[p, q]$ (if $p < q$) or $[q, p]$ (if $q < p$) that extends T.*

Proof. Since $\widetilde{T}$ is strong type (p, p) and strong type (q, q), by the Riesz–Thorin theorem $\widetilde{T}$ is strong-type (r, r) for all r in the interval $[p, q]$ (if $p < q$) or $[q, p]$ (if $q < p$). To see $\widetilde{T}$ is the unique such extension, suppose S is another such extension of T. We can write each $h \in L^r$ as a sum $h = f + g$ for some $f \in L^p$ and $g \in L^r$, so

$$Sh = S(f + g) = Sf + Sg = \widetilde{T}f + \widetilde{T}g = \widetilde{T}h$$

since because S is an extension we have $Sf = \widetilde{T}f$ for all $f \in L^p$ and $Sg = \widetilde{T}g$ for all $g \in L^q$. Thus $S = \widetilde{T}$, so the extension is unique. ☐

Exercise 6.83: Folland Exercise 6.42.

Prove the Marcinkiewicz theorem in the case $p_0 = p_1$. (Setting $p = p_0 = p_1$, we have $\lambda_{Tf}(\alpha) \leqslant (C_0\|f\|_p/\alpha)^{q_0}$ and $\lambda_{Tf}(\alpha) \leqslant (C_1\|f\|_p/\alpha)^{q_1}$. Use whichever estimate is better, depending on α, to majorize $q\int_0^\infty \alpha^{q-1}\lambda_{Tf}(\alpha)d\alpha$.)

Proof. Suppose that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are measure spaces; $p, q_0, q_1 \in [1, \infty]$ and

$p \leqslant q_0, q_1$, and $q_0 \neq q_1$; and

$$\frac{1}{q} = \frac{1-t}{q_0} + \frac{t}{q_1}, \qquad \text{where } 0 < t < 1.$$

Let $T\colon L^p(\mu) \to L^0(\nu)$ be[1] a sublinear map of weak types (p, q_0) and (p, q_1). We claim T is strong type (p, q). More precisely, suppose $[Tf]_{q_j} \leqslant C_j \|f\|_p$ for $j = 0, 1$. We claim $\|Tf\|_q \leqslant B_p \|f\|_p$ where B_p depends only on p, q_j, and C_j in addition to p.

Then for $\alpha > 0$ we have the estimates

$$\lambda_{Tf}(\alpha) \leqslant (C_0 \|f\|_p/\alpha)^{q_0} \quad \text{and} \quad \lambda_{Tf}(\alpha) \leqslant (C_1 \|f\|_p/\alpha)^{q_1},$$

so we obtain the estimate

$$\begin{aligned}
\|Tf\|_q^q = \int |Tf|^q &= q \int_0^\infty \alpha^{q-1} \mu\{|Tf| > \alpha\}\, \mathrm{d}\alpha \\
&= q \int_0^{\|f\|_p} \alpha^{q-1} \mu\{|Tf| > \alpha\}\, \mathrm{d}\alpha + q \int_{\|f\|_p}^\infty \alpha^{q-1} \mu\{|Tf| > \alpha\}\, \mathrm{d}\alpha \\
&\leqslant q \int_0^{\|f\|_p} \alpha^{q-1} \left(\frac{C_0 \|f\|_p}{\alpha} \right)^{q_0} \mathrm{d}\alpha + q \int_{\|f\|_p}^\infty \alpha^{q-1} \left(\frac{C_1 \|f\|_p}{\alpha} \right)^{q_1} \mathrm{d}\alpha \\
&\leqslant q C_0^{q_0} \|f\|_p^{q_0} \int_0^{\|f\|_p} \alpha^{q-q_0-1}\, \mathrm{d}\alpha + q C_1^{q_1} \|f\|_p^{q_1} \int_{\|f\|_p}^\infty \alpha^{q-q_1-1}\, \mathrm{d}\alpha \\
&\leqslant q C_0^{q_0} \|f\|_p^{q_0} \left[\frac{\alpha^{q-q_0}}{q-q_0} \right]_{\alpha=0}^{\alpha=\|f\|_p} + q C_1^{q_1} \|f\|_p^{q_1} \left[\frac{\alpha^{q-q_1}}{q-q_1} \right]_{\alpha=\|f\|_p}^{\alpha=\infty}. \\
&= \left(\frac{q C_0^{q_0} \|f\|_p^{q_0} \|f\|^{q-q_0}}{q-q_0} \right) - \left(\frac{q C_1^{q_1} \|f\|_p^{q_1} \|f\|_p^{q-q_1}}{q-q_1} \right) \\
&= \left(\frac{q C_0^{q_0}}{q-q_0} + \frac{q C_1^{q_1}}{q_1-q} \right) \|f\|_p^q.
\end{aligned}$$

Thus T is strong type (p, q), as claimed, and moreover $B_p \coloneqq \left(\frac{qC_0^{q_0}}{q-q_0} + \frac{qC_1^{q_1}}{q_1-q} \right)^{1/q}$ depends only on q_j and C_j for $j = 0, 1$. □

Exercise 6.84: Folland Exercise 6.43.

Let H be the Hardy-Littlewood maximal operator on $\mathbb{R}$. Compute $H\chi_{(0,1)}$ explicitly. Show that it is in L^p for all $p > 1$ and in weak L^1 but not in L^1, and that its L^p norm tends to ∞ like $(p-1)^{-1}$ as $p \to 1$, although $\|\chi_{(0,1)}\|_p = 1$ for all p.

Exercise 6.85: Folland Exercise 6.44.

Let I_α be the fractional integration operator of Folland Exercise 2.61. If $0 < \alpha < 1$, $1 < p < \alpha^{-1}$, and $r^{-1} = p^{-1} - \alpha$, then I_α is weak type $(1, (1-\alpha)^{-1})$ and strong type (p, r) with respect to Lebesgue measure on $(0, \infty)$.

[1] Here $L^0(\nu)$ is the space of measurable functions on Y.

Exercise 6.86: Folland Exercise 6.45, Altered.

The following concerns Folland Exercise 6.45, which reads as follows:

> If $0 < \alpha < n$, define an operator T_α on functions on $\mathbb{R}^n$ by
> $$T_\alpha f(x) := \int |x-y|^{-\alpha} f(y)\,\mathrm{d}y$$
> Then T_α is weak type $(1,(n-\alpha)^{-1})$ and strong type (p,r) with respect to Lebesgue measure on $\mathbb{R}^n$, where $1 < p < n\alpha^{-1}$ and $r^{-1} = p^{-1} - \alpha n^{-1}$. (The case $n = 3$, $\alpha = 1$ is of particular interest in physics: If f represents the density of a mass or charge distribution, $-(4\pi)^{-1}T_1 f$ represents the induced gravitational or electrostatic potential.)

The following aims to correct this exercise.

(a) Use a scaling argument to show that the exercise is incorrect as stated.

(b) Replace the exponent $-\alpha$ in the definition of with $-n+\alpha$ in the question. Prove that (this version of) T_α is weak type $(1, 1(n-\alpha)^{-1})$ and strong type (p,r) under the conditions on α, p, and r as stated in the exercise. Hint: First show that T_α is of *weak* type (p,r).

Solution.

(a) Suppose for a contradiction T_α is strong type (p,r), so that $\|T_\alpha\|_{L^p\to L^r} < \infty$. Now fix $\varepsilon > 0$. Since $\|T_\alpha\|_{L^p\to L^r} = \sup\{\|T_\alpha f\|_r \mid \|f\|_p = 1\} < \infty$, there exists $f \in L^p$ such that $\|f\|_p = 1$ and
$$\|T_\alpha f\|_r > (1-\varepsilon)\|T_\alpha\|_{L^p\to L^r} \tag{6.86.1}$$
For each $b \in \mathbb{R}_{>0}$ define $g_b \colon \mathbb{R}^n \to \mathbb{C}$ by
$$g_b(x) = f(bx).$$
Then $g_b(x) \in L^p$ and for a fixed $b \in \mathbb{R}_{>0}$ be fixed. We have
$$\|g_b\|_p^p = \int |g_b(x)|^p\,\mathrm{d}x = \frac{1}{b^{np}} \int |f(x)|^p\,\mathrm{d}x = \frac{1}{b^{np}},$$
so $\|g_b\|_p = 1/b^n$. And for each $x \in \mathbb{R}^n$, we have
$$\begin{aligned} T_\alpha g_b(x) &= \int |x-y|^{-\alpha} f(by)\,\mathrm{d}y \\ &= b^{-n}\int |x-y/b|^{-\alpha} f(y)\,\mathrm{d}y && (\text{substitute } by \mapsto y) \\ &= b^{-n}\int \left|\frac{bx-y}{b}\right|^{-\alpha} f(y)\,\mathrm{d}y = b^{\alpha-n}\int |bx-y|^{-\alpha} f(y)\,\mathrm{d}y, \end{aligned}$$

so

$$\begin{aligned}\|T_\alpha g_b\|_r^r &= b^{r(\alpha-n)} \int \left| \int |bx-y|^{-\alpha} f(y)\,\mathrm{d}y \right|^r \mathrm{d}x \\ &= b^{r(\alpha-n)} \int \left| b^{-n} \int |x-y|^{-\alpha} f(y)\,\mathrm{d}y \right|^r \mathrm{d}x \qquad \text{(substitute } bx \mapsto x) \\ &= b^{r(\alpha-2n)} \int |T_\alpha f(x)|^r \,\mathrm{d}x = b^{r(\alpha-2n)} \|T_\alpha f\|_r^r.\end{aligned}$$

Thus

$$b^{\alpha-n}\|T_\alpha f\|_r = \frac{b^{\alpha-2n}\|T_\alpha f\|_r}{b^{-n}} = \frac{\|T_\alpha g_b\|_r}{\|g_b\|_p} \leqslant \|T_\alpha\|_{L^p\to L^r}.$$

Therefore, since $0 < \alpha < n$ and in particular $\alpha \neq n$, we can choose $f \in L^p$ and $b > 0$ sufficiently large such that the left-hand side is strictly larger than the right-hand side (since otherwise T_α is the zero operator, contrary to the given definition of T_α), which contradicts the assumed boundedness of T on L^p. It follows that Folland Exercise 6.45 is incorrect as stated.

(b) Define $K \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{C}$ by

$$K(x,y) \coloneqq |x-y|^{-\alpha}.$$

Then K is $m \times m$-measurable, and for each $x \in \mathbb{R}^n$ and $\beta > 0$ we have

$$\begin{aligned}\lambda_{K(x,-)}(\beta) &= m(\{y \in \mathbb{R}^n \mid |x-y|^{-\alpha} > \beta\}) \\ &= m(\{y \in \mathbb{R}^n \mid |x-y| < \beta^{-1/\alpha}\}) \\ &\leqslant m(B_{\beta^{-1/\alpha}}(x))\end{aligned}$$

Since the measure of a ball of radius r in $\mathbb{R}^n$ is a scalar multiple of the radius to the power of n, there exists $C > 0$ such that for all $x \in \mathbb{R}^n$ and all $\beta > 0$,

$$m(B_{\beta^{-1/\alpha}}(x)) = C\,\beta^{-n/\alpha}$$

and thus

$$\beta^{n/\alpha}\lambda_{K(x,-)}(\beta) \leqslant \beta^{n/\alpha} m(B_{\beta^{-1/\alpha}}(x)) = \beta^{-n/\alpha}\beta^{n/\alpha}C = C.$$

Thus, by taking the $1/(n/a)$th power of both sides and taking the supremum over all $\beta \in \mathbb{R}_{>0}$, we obtain for all $x \in \mathbb{R}^n$ that

$$[K(x,-)]_q = \sup_{q>0}(\beta^q \lambda_{K(x,-)}(\beta))^{1/q} \leqslant C^{1/q}.$$

Arguing identically (but replacing $K(x,-)$ with $K(-,y)$ and x with y), there exists $C' > 0$ such that $[K(-,y)]_q \leqslant C'^{1/q}$ for all $y \in \mathbb{R}^n$. Now replacing C with the maximum of $C^{1/q}, C'^{1/q}$, the result then follows immediately from Folland Theorem 6.36. □

8 Elements of Fourier Analysis

It is easy to say that Fourier analysis, or harmonic analysis, originated in the work of Euler, Fourier, and others on trigonometric series; it is much harder to describe succinctly what the subject comprises today, for it is a meeting ground for ideas from many parts of analysis and has applications in such diverse areas as partial differential equations and algebraic number theory. Two of the central ingredients of harmonic analysis, however, are convolution operators and the Fourier transform, which we study in this chapter.

8.1 Preliminaries

We begin by making some notational conventions. Throughout this chapter we shall be working on $\mathbb{R}^n$, and n will always refer to the dimension. In any measure-theoretic considerations we always have Lebesgue measure in mind unless we specify otherwise. Thus, if E is a measurable set in $\mathbb{R}^n$, we shall denote $L^p(E, m)$ by $L^p(E)$. If U is open in $\mathbb{R}^n$ and $k \in \mathbb{R}$, we denote by $C^k(U)$ the space of all functions on U whose partial derivatives of order $\leqslant k$ all exist and are continuous, and we set $C^\infty(U) = \bigcap_1^\infty C^k(U)$. Furthermore, for any $E \subset \mathbb{R}^n$ we denote by $C_c^\infty(E)$ the space of all C^∞ functions on $\mathbb{R}^n$ whose support is compact and contained in E. If $E = \mathbb{R}^n$ or $U = \mathbb{R}^n$, we shall usually omit it in naming function spaces: thus, $L^p = L^p(\mathbb{R}^n), C^k = C^k(\mathbb{R}^n), C_c^\infty = C_c^\infty(\mathbb{R}^n)$. If $x, y \in \mathbb{R}^n$, we set

$$x \cdot y = \sum\nolimits_1^n x_j y_j, \quad |x| = \sqrt{x \cdot x}$$

8.1.1 Multi-index notation

It will be convenient to have a compact notation for partial derivatives. We shall write

$$\partial_j = \frac{\partial}{\partial x_j}$$

and for higher-order derivatives we use multi-index notation. A multi-index is an ordered n-tuple of nonnegative integers. If $\alpha = (\alpha_1, \ldots, \alpha_n)$ is a multi-index, we set

$$|\alpha| = \sum\nolimits_1^n \alpha_j, \quad \alpha! = \prod\nolimits_1^n \alpha_j!, \quad \partial^\alpha = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n}$$

and if $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$x^\alpha = \prod\nolimits_1^n x_j^{\alpha_j}$$

(The notation $|\alpha| = \sum \alpha_j$ is inconsistent with the notation $|x| = (\sum x_j^2)^{1/2}$, but the meaning will always be clear from the context.) Thus, for example, Taylor's formula for functions $f \in C^k$ reads

$$f(x) = \sum\nolimits_{|\alpha| \leqslant k} (\partial^\alpha f)(x_0) \frac{(x - x_0)^\alpha}{\alpha!} + R_k(x), \quad \lim_{x \to x_0} \frac{|R_k(x)|}{|x - x_0|^k} = 0$$

and the product rule for derivatives becomes

$$\partial^\alpha(fg) = \sum_{\beta+\gamma=\alpha} \frac{\alpha!}{\beta!\gamma!}(\partial^\beta f)(\partial^\gamma g)$$

(see Folland Exercise 8.1).

We shall often avail ourselves of the sloppy but handy device of using the same notation for a function and its value at a point. Thus, "x" may be used to denote the function whose value at any point x is x^α.

8.1.2 Existence of nonzero functions in C_c^∞

Two spaces of C^∞ functions on $\mathbb{R}^n$ will be of particular importance for us. The first is the space C_c^∞ of C^∞ functions with compact support. The existence of nonzero functions in C_c^∞ is not quite obvious; the standard construction is based on the fact that the function $\eta(t) = e^{-1/t}\chi_{(0,\infty)}(t)$ is C^∞ even at the origin (Folland Exercise 8.3). If we set

$$\psi(x) = \eta(1-|x|^2) = \begin{cases} \exp[(|x|^2-1)^{-1}] & \text{if } |x| < 1 \\ 0 & \text{if } |x| \geqslant 1 \end{cases} \tag{8.0.1}$$

it follows that $\psi \in C^\infty$, and $\text{supp}(\psi)$ is the closed unit ball. In the next section we shall use this single function to manufacture elements of C_c^∞ in great profusion; see Proposition 19 and Theorem 20.

8.1.3 Schwartz space

The other space of C^∞ functions we shall need is the Schwartz space $\mathcal{S}$ consisting of those C^∞ functions which, together with all their derivatives, vanish at infinity faster than any power of $|x|$. More precisely, for any nonnegative integer N and any multi-index α we define

$$\|f\|_{(N,\alpha)} = \sup_{x\in\mathbb{R}^n}(1+|x|)^N|\partial^\alpha f(x)|$$

then

$$\mathcal{S} = \{f \in C^\infty \mid \|f\|_{(N,\alpha)} < \infty \text{ for all } N, \alpha\}$$

Examples of functions in $\mathcal{S}$ are easy to find: for instance, $f_\alpha(x) = x^\alpha e^{-|x|^2}$ where α is any multi-index. Also, clearly $C_c^\infty \subset \mathcal{S}$.

It is an important observation that if $f \in \mathcal{S}$, then $\partial^\alpha f \in L^p$ for all α and all $p \in [1,\infty]$. Indeed, $|\partial^\alpha f(x)| \leqslant C_N(1+|x|)^{-N}$ for all N, and $(1+|x|)^{-N} \in L^p$ for $N > n/p$ by Corollary 101.

Proposition 8.1: 8.2.

$\mathcal{S}$ is a Fréchet space with the topology defined by the norms $\|\cdot\|_{(N,\alpha)}$.

Proof. The only nontrivial point is completeness. If $\{f_k\}$ is a Cauchy sequence in $\mathcal{S}$, then $\|f_j - f_k\|_{(N,\alpha)} \to 0$ for all N, α. In particular, for each α the sequence $\{\partial^\alpha f_k\}$ converges uniformly to a function g_α. Denoting by e_j the vector $(0, \ldots, 1, \ldots, 0)$ with the 1 in the jth position, we have

$$f_k(x + te_j) - f_k(x) = \int_0^t \partial_j f_k(x + se_j) ds$$

Letting $k \to \infty$, we obtain

$$g_0(x + te_j) - g_0(x) = \int_0^t g_{e_j}(x + se_j) ds$$

The fundamental theorem of calculus implies that $g_{e_j} = \partial_j g_0$, and an induction on $|\alpha|$ then yields $g_\alpha = \partial^\alpha g_0$ for all α. It is then easy to check that $\|f_k - g_0\|_{(N,\alpha)} \to 0$ for all α. □

Another useful characterization of $\mathcal{S}$ is the following.

Proposition 8.2: 8.3.

If $f \in C^\infty$, then $f \in \mathcal{S}$ if and only if $x^\beta \partial^\alpha f$ is bounded for all multi-indices α, β if and only if $\partial^\alpha(x^\beta f)$ is bounded for all multi-indices α, β.

Proof. Obviously $|x^\beta| \leqslant (1 + |x|)^N$ for $|\beta| \leqslant N$. On the other hand, $\sum_1^n |x_j|^N$ is strictly positive on the unit sphere $|x| = 1$, so it has a positive minimum δ there. It follows that $\sum_1^n |x_j|^N \geqslant \delta |x|^N$ for all x since both sides are homogeneous of degree N, and hence

$$(1 + |x|)^N \leqslant 2^N(1 + |x|^N) \leqslant 2^N\left[1 + \delta^{-1} \sum\nolimits_1^n |x_j^N|\right] \leqslant 2^N \delta^{-1} \sum\nolimits_{|\beta| \leqslant N} |x^\beta|$$

This establishes the first equivalence. The second one follows from the fact that each $\partial^\alpha(x^\beta f)$ is a linear combination of terms of the form $x^\gamma \partial^\delta f$ and vice versa, by the product rule (Folland Exercise 8.1). □

We next investigate the continuity of translations on various function spaces. The following notation for translations will be used throughout this chapter and the next one: If f is a function on $\mathbb{R}^n$ and $y \in \mathbb{R}^n$,

$$\tau_y f(x) = f(x - y)$$

We observe that $\|\tau_y f\|_p = \|f\|_p$ for $1 \leqslant p \leqslant \infty$ and that $\|\tau_y f\|_u = \|f\|_u$. A function f is called uniformly continuous if $\|\tau_y f - f\|_u \to 0$ as $y \to 0$. (The reader should pause to check that this is equivalent to the usual ε-δ definition of uniform continuity.)

Lemma 8.3: 8.4.

If $f \in C_c(\mathbb{R}^n)$, then f is uniformly continuous.

Proof. Given $\varepsilon > 0$, for each $x \in \operatorname{supp}(f)$ there exists $\delta_x > 0$ such that $|f(x-y)-f(x)| < \frac{1}{2}\varepsilon$ if $|y| < \delta_x$. Since $\operatorname{supp}(f)$ is compact, there exist $x_1, \ldots, x_N$ such that the balls of radius $\frac{1}{2}\delta_{x_j}$ about x_j cover $\operatorname{supp}(f)$. If $\delta = \frac{1}{2}\min\{\delta_{x_j}\}$, then, one easily sees that $\|\tau_y f - f\|_u < \varepsilon$ whenever $|y| < \delta$. □

Proposition 8.4: 8.5.

If $1 \leqslant p < \infty$, translation is continuous in the L^p norm; that is, if $f \in L^p$ and $z \in \mathbb{R}^n$, then $\lim_{y\to 0}\|\tau_{y+z}f - \tau_z f\|_p = 0$.

Proof. Since $\tau_{y+z} = \tau_y\tau_z$, by replacing f by $\tau_z f$ it suffices to assume that $z = 0$. First, if $g \in C_c$, for $|y| \leqslant 1$ the functions $\tau_y g$ are all supported in a common compact set K, so by Lemma 3,

$$\int |\tau_y g - g|^p \leqslant \|\tau_y g - g\|_u^p m(K) \to 0 \text{ as } y \to 0$$

Now suppose $f \in L^p$. If $\varepsilon > 0$, by **??** there exists $g \in C_c$ with $\|g - f\|_p < \varepsilon/3$, so

$$\|\tau_y f - f\|_p \leqslant \|\tau_y(f-g)\|_p + \|\tau_y g - g\|_p + \|g - f\|_p < \frac{2}{3}\varepsilon + \|\tau_y g - g\|_p$$

and $\|\tau_y g - g\|_p < \varepsilon/3$ if y is sufficiently small.

Proposition 4 is false for $p = \infty$, as one should expect since the L^∞ norm is closely related to the uniform norm; see Folland Exercise 8.4.

Some of our results will concern multiply periodic functions in $\mathbb{R}^n$, and for simplicity we shall take the fundamental period in each variable to be 1. That is, we define a function f on $\mathbb{R}^n$ to be periodic if $f(x+k) = f(x)$ for all $x \in \mathbb{R}^n$ and $k \in \mathbb{R}^n$. Every periodic function is thus completely determined by its values on the unit cube

$$Q = \left[-\frac{1}{2}, \frac{1}{2}\right)^n.$$

Periodic functions may be regarded as functions on the space $\mathbb{R}^n/\mathbb{Z}^n \cong (\mathbb{R}/\mathbb{Z})^n$ of cosets of $\mathbb{R}^n$, which we call the n-dimensional torus and denote by $\mathbb{T}^n$. (When $n = 1$ we write $\mathbb{T}$ rather than $\mathbb{T}^1$.) $\mathbb{T}^n$ is a compact Hausdorff space; it may be identified with the set of all $z = (z_1, \ldots, z_n) \in \mathbb{C}^n$ such that $|z_j| = 1$ for all j, via the map

$$(x_1, \ldots, x_n) \mapsto (e^{2\pi i x_1}, \ldots, e^{2\pi i x_n})$$

On the other hand, for measure-theoretic purposes we identify $\mathbb{T}^n$ with the unit cube Q, and when we speak of Lebesgue measure on $\mathbb{T}^n$ we mean the measure induced on $\mathbb{T}^n$ by Lebesgue measure on Q. In particular, $m(\mathbb{T}^n) = 1$. Functions on $\mathbb{T}^n$ may be considered as periodic functions on $\mathbb{T}^n$ or as functions on Q; the point of view will be clear from the context when it matters.

Exercise 8.5: Folland Exercise 8.1.

Prove the product rule for partial derivatives as stated in the text. Deduce that

$$\partial^\alpha(x^\beta f) = x^\beta \partial^\alpha f + \sum c_{\gamma\delta} x^\delta \partial^\gamma f, \quad x^\beta \partial^\alpha f = \partial^\alpha(x^\beta f) + \sum c'_{\gamma\delta} \partial^\gamma(x^\delta f)$$

for some constants $c_{\gamma\delta}$ and $c'_{\gamma\delta}$ with $c_{\gamma\delta} = c'_{\gamma\delta} = 0$ unless $|\gamma| < |\alpha|$ and $|\delta| < |\beta|$.

Exercise 8.6: Folland Exercise 8.2.

Observe that the binomial theorem can be written as follows:

$$(x_1 + x_2)^k = \sum_{|\alpha|=k} \frac{k!}{\alpha!} x^\alpha \quad (x = (x_1, x_2), \alpha = (\alpha_1, \alpha_2))$$

Prove the following generalizations:

(a) The multinomial theorem: If $x \in \mathbb{R}^n$,

$$(x_1 + \cdots + x_n)^k = \sum_{|\alpha|=k} \frac{k!}{\alpha!} x^\alpha$$

(b) The n-dimensional binomial theorem: If $x, y \in \mathbb{R}^n$,

$$(x + y)^\alpha = \sum_{\beta+\gamma=\alpha} \frac{\alpha!}{\beta!\gamma!} x^\beta y^\gamma.$$

Exercise 8.7: Folland Exercise 8.3.

Let $\eta(t) = e^{-1/t}$ for $t > 0, \eta(t) = 0$ for $t \leqslant 0$.

(a) For $k \in \mathbb{Z}_{\geqslant 1}$ and $t > 0, \eta^{(k)}(t) = P_k(1/t)e^{-1/t}$ where P_k is a polynomial of degree $2k$.

(b) $\eta^{(k)}(0)$ exists and equals zero for all $k \in \mathbb{Z}_{\geqslant 1}$.

Exercise 8.8: Folland Exercise 8.4.

If $f \in L^\infty$ and $\|\tau_y f - f\|_\infty \to 0$ as $y \to 0$, then f agrees a.e. with a uniformly continuous function. (Let $A_r f$ be as in Theorem 45. Then $A_r f$ is uniformly continuous for $r > 0$ and uniformly Cauchy as $r \to 0$.)

Solution. The statement of Exercise 8 follows immediately from the following points:

(i) $A_{1/n} f(x) \to f(x)$ a.e. as $n \to \infty$.

(ii) For all $n \in \mathbb{Z}_{\geqslant 1}$, $A_{1/n} f(x)$ is uniformly continuous as a function of $x \in \mathbb{R}^n$.

(iii) The sequence $\{A_{1/n} f\}_{n=1}^\infty$ is uniformly Cauchy.

(iv) If $\{f_n : \mathbb{R}^n \to \mathbb{C}\}_{n=1}^\infty$ is a uniformly Cauchy sequence of uniformly continuous functions, then $\lim_{n\to\infty} f_n$ is uniformly continuous.

Proof of (i). This is just Folland Theorem 3.18 since L^∞ functions are L^1_{loc}. □

Proof of (ii). Let $n \in \mathbb{Z}_{\geqslant 1}$. Fix $\varepsilon > 0$. It suffices to show $\|\tau_y A_r f - A_r f\|_u \to 0$ as $y \to 0$. For any x, we have

$$\begin{aligned} |\tau_y A_{1/n} f(x) - A_{1/n} f(x)| &= \frac{1}{m(B_r(0))}\left|\int_{B_r(x-y)} |f(z)|\,\mathrm{d}z - \int_{B_r(x)} |f(z)|\,\mathrm{d}z\right| \\ &= \frac{1}{m(B_r(0))}\left|\int_{B_r(x-y)} |f(z)|\,\mathrm{d}z - \int_{B_r(x-y)} |f(z-y)|\,\mathrm{d}z\right| \\ &\qquad \text{(substitute } z \mapsto z-y) \\ &\leqslant \frac{1}{m(B_r(0))}\int_{B_r(x-y)} |\tau_y f(z) - f(z)|\,\mathrm{d}z \\ &\leqslant \frac{1}{m(B_r(0))}\int_{B_r(x-y)} \|\tau_y f - f\|_\infty\,\mathrm{d}z \\ &\qquad \text{(since } |\tau_y f(z) - f(z)| \leqslant \|\tau_y f - f\|_\infty \text{ for a.e. } z \in \mathbb{R}^n) \\ &= \|\tau_y f - f\|_\infty. \end{aligned}$$

Taking the supremum of both sides over all $x \in \mathbb{R}^n$, we obtain

$$\|\tau_y A_{1/n} f - A_{1/n} f\|_u \leqslant \|\tau_y f - f\|_\infty.$$

Since $\|\tau_y f - f\|_\infty \to 0$ as $y \to 0$ by hypothesis, we conclude $A_{1/n} f$ is uniformly continuous. ☐

Proof of (iii). We claim $\|A_{1/n} f - A_{1/m} f\|_u \to 0$ as $m, n \to \infty$. Fix $\varepsilon > 0$. Since $A_{1/n}$ By Folland Lemma 3.16, $A_r f$ is a continuous function of $r \in \mathbb{R}_{>0}$; thus $A_{1/n} f - A_{1/m} f$ is continuous for all $n, m \in \mathbb{Z}_{\geqslant 1}$, so its supremum norm equals its infinity norm. Hence

$$\|A_{1/n} f - A_{1/m} f\|_u = \|A_{1/n} f - A_{1/m} f\|_\infty \leqslant \|A_{1/n} f - f\|_\infty + \|A_{1/m} f - f\|_\infty. \tag{8.8.1}$$

Where $n \in \mathbb{Z}_{\geqslant 1}$, we have

$$\begin{aligned} \|A_{1/n} f - f\|_\infty &= \left\| x \mapsto \frac{1}{m(B_{1/n}(x))}\int_{B_{1/n}(x)} |f(y)|\,\mathrm{d}y - f(x)\right\|_\infty \\ &\leqslant \left\| x \mapsto \frac{1}{m(B_{1/n}(x))}\int_{B_{1/n}(x)} |f(y) - f(x)|\,\mathrm{d}y\right\|_\infty \\ &\qquad \text{(by the triangle inequality)} \\ &\leqslant \left\| x \mapsto \frac{1}{m(B_{1/n}(x))}\int_{B_{1/n}(0)} |\tau_y f(x) - f(x)|\,\mathrm{d}y\right\|_\infty \\ &\leqslant \frac{1}{m(B_{1/n}(x))}\int_{B_{1/n}(0)} \|x \mapsto |\tau_y f(x) - f(x)|\|_\infty\,\mathrm{d}y \\ &\leqslant \frac{1}{m(B_{1/n}(x))}\int_{B_{1/n}(0)} \|\tau_y f - f\|_\infty\,\mathrm{d}y, \longrightarrow 0 \text{ as } n \to \infty \end{aligned}$$

where we used Minkowski's inequality for integrals (Folland Theorem 6.19) since $\tau_y f - f \in L^\infty$ for a.e. $y \in \mathbb{R}^n$ and $[y \mapsto \|\tau_y f - f\|_p] \in L^1$.

Thus both terms on the right-hand side of Equation (8.8.1) tend to 0 as $m, n \to \infty$, so

$\{A_{1/n}f\}_{n=1}^{\infty}$ is uniformly Cauchy. □

Proof of (iv). Fix $\varepsilon > 0$ and $g = \lim_{n\to\infty} f_n$. Then for all sufficiently large n, $|f_n(x) - g(x)| < \varepsilon/3$. Since each f_n is uniformly continuous, there exists $\delta > 0$ such that $|f_n(x) - f_n(y)| < \varepsilon/3$ whenever $|x - y| < \delta$. Thus, for any x, y such that $|x - y| < \delta$ and all sufficiently large n, we have

$$\begin{aligned}|g(x) - g(y)| &\leqslant |g(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - g(y)| \\ &< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon,\end{aligned}$$

so g is uniformly continuous. □

8.2 Convolutions

Definition 9 (Convoltuion). *Define the* ***convolution*** *by the assignment* $*\colon L^0 \times L^0 \longrightarrow L^0$, *written* $(f, g) \mapsto f * g$, *where*

$$f * g(x) := \int f(y)g(x - y)\,\mathrm{d}y.$$

for all x such that the integral exists.[2]

Proposition 8.10.

The convolution $*$ is well-defined. That is, for any $f, g \in L^0$, $[y \mapsto f(y)g(x - y)] \in L^1$ for a.e. $x \in \mathbb{R}^n$ and $f * g \in L^0$.

Proof. We shall need the fact that if f is a measurable function on $\mathbb{R}^n$, then the function $K(x, y) = f(x - y)$ is measurable on $\mathbb{R}^n \times \mathbb{R}^n$. We have $K = f \circ s$ where $s(x, y) = x - y$; since s is continuous, K is Borel measurable if f is Borel measurable. This can always be assumed without affecting the definition of $f * g$, by Proposition 22. However, the Lebesgue measurability of K also follows from the Lebesgue measurability of f; see Folland Exercise 8.5.

TODO. □

The elementary properties of convolutions are summarized in the following proposition.

Proposition 8.11: Properties of the convolution.

Let $f, g, h \in L^0$. Provided that the integrals in the following assertions exist, they are true:

(a) $f * g = g * f$.

[2]Various conditions can be imposed on f and g to guarantee that $f * g$ is defined at least almost everywhere. For example, if f is bounded and compactly supported, g can be any locally integrable function; see also Theorem 12, Theorem 14, and Proposition 13.

(b) If $f * g(x)$, $g * h(x)$, and $f * g * h(x) := \iint f(x-y-z)g(y)h(z)\,\mathrm{d}y\,\mathrm{d}z$ exist and are finite for a.e. $x \in \mathbb{R}^n$, then[a] $(f * g) * h = f * (g * h)$.
(c) For $z \in \mathbb{R}^n$, $\tau_z(f * g) = (\tau_z)(f * g) = (\tau_z f) * g = f * (\tau_z)g$.
(d) $\operatorname{supp}(f * g) \subset \overline{\operatorname{supp}(f) + \operatorname{supp}(g)}$.

[a] Associativity of the convolution can fail without these extra hypotheses. For example, for $f = \chi_{\mathbb{R}_{>0}} - \chi_{\mathbb{R}_{>0}}$, $g = \chi_{[0,1]} - \chi_{[-1,0)}$ and $h = \chi_{\mathbb{R}}$, we have $f * g(x) = g * f(x) = \max\{0, 2 - 2|x|\}$ and $g * h = h * g = 0$. However for every $x \in \mathbb{R}$, $(f * g) * h(x) = \int_{\mathbb{R}} f * g(y)dy = \int_{-1}^{1}(2 - 2|y|)dy = 2$, while $f * (g * h)(x) = f * 0(x) = 0$, so $(f * g) * h \neq f * (g * h)$.

Proof.

(a) Substituting $z \mapsto x - y$, we obtain
$$f * g(x) = \int f(x-y)g(y)\,\mathrm{d}y = \int f(z)g(x-z)\,\mathrm{d}z = g * f(x).$$
(b) At any $x \in \mathbb{R}^n$ where $f * g * h(x)$ is defined, by Fubini's theorem we can write $f * g * h(x)$ as
$$\begin{aligned}\int h(z)\,\mathrm{d}z \int f(x-z-y)g(y)\,\mathrm{d}y &= \int f * g(x-z)h(z)\,\mathrm{d}z = (f * g) * h(x)\\ &= \int f(x-y)g(y-z)h(z)\,\mathrm{d}y\,\mathrm{d}z = \int f(x-y)\,\mathrm{d}y \int g(y-z)h(z)\,\mathrm{d}z\\ &= \int f(x-y)(g * h)(y)\,\mathrm{d}y = f * (g * h)(x).\end{aligned}$$
In particular $(f * g) * h$ and $f * (g * h)$ are defined and equal to each other.

(c) We have
$$\tau_z(f * g)(x) = \int f(x-z-y)g(y)\,\mathrm{d}y = \int \tau_z f(x-y)g(y)\,\mathrm{d}y = (\tau_z f) * g(x),$$
and by (a),
$$\tau_z(f * g) = \tau_z(g * f) = (\tau_z g) * f = f * (\tau_z g).$$
(d) If $x \notin \overline{\operatorname{supp}(f) + \operatorname{supp}(g)}$, then for any $y \in \operatorname{supp}(g)$ we have $x - y \notin \operatorname{supp}(f)$; thus $f(x-y)g(y) = 0$ for all y, so $f * g(x) = 0$. ☐

The following two propositions contain the basic facts about convolutions of L^p functions.

8.12 Young's convolution inequality.

For any $p \in [1, \infty]$, if $f \in L^1$ and $g \in L^p$, then $f * g(x)$ exists for a.e. $x \in \mathbb{R}^n$, $f * g \in L^p$, and
$$\|f * g\|_p \leqslant \|f\|_1 \|g\|_p.$$

Proof. By Minkowski's inequality for integrals,

$$\|f * g\|_p = \left\| x \mapsto \int f(y)g(x-y)\,\mathrm{d}y \right\|_p \leqslant \int |f(y)| \|\tau_y g\|_p \,\mathrm{d}y = \|f\|_1 \|g\|_p. \qquad \square$$

Proposition 8.13.

For any $p \in [1, \infty]$, if $q = p'$, $f \in L^p$, and $g \in L^q$, then $f * g(x)$ exists for a.e. $x \in \mathbb{R}^n$, $f * g$ is bounded and uniformly continuous, and

$$\|f * g\|_u \leqslant \|f\|_p \|g\|_q.$$

If $p \in (1, \infty)$, then also $f * g \in C_0(\mathbb{R}^n)$.

Proof. By Hölder's inequality,

$$|f * g(x)| = \int f(y)g(x-y)\,\mathrm{d}y \leqslant \|f\|_p \|g\|_q,$$

so by taking the supremum over all $x \in \mathbb{R}^n$ we obtain $\|f * g\|_u \leqslant \|f\|_p \|g\|_q$. And $f * g$ is uniformly continuous by Propositions 4 and 11, since

$$\|\tau_y (f * g) - f * g\|_u = \|(\tau_y f - f) * g\|_\infty \leqslant \|\tau_y f - f\|_p \|g\|_q \to 0 \text{ as } y \to 0,$$

where if $p = \infty$ we swap f and g. Finally, if $p, q \in (1, \infty)$ then choose sequences $\{f_n\}$, $\{g_n\}$ of compactly supported functions such that $\|f_n - f_p\|, \|g_n - g_q\| \to 0$. Then by the above and Proposition 11(d), we obtain $f_n * g_n \in C_c$. But

$$\|f_n * g_n - f * g\|_u \leqslant \|f_n - f\|_p \|g_n\| + \|f\| \|g_n - g\|_q \to 0,$$

so $f * g \in C_0$ by Proposition 116. $\square$

Theorem 8.14.

Suppose $p, q, r \in [1, \infty]$ and $1/p + 1/q = 1/r + 1$.

(a) (Young's convolution inequality—general form), if $f \in L^p$ and $g \in L^q$, then $f * g \in L^r$ and

$$\|f * g\|_r \leqslant \|f\|_p \|g\|_r.$$

(b) Suppose also $p, q > 1$ and $r < \infty$. If $f \in L^p$ and $g \in \text{weak}(L^p)$, then $f * g \in L^r$ and there exists a constant $C_{pq} > 0$ independent of f and g such that

$$\|f * g\|_r \leqslant C_{pq} \|f\|_p \|g\|_q.$$

(c) Suppose $p = 1$ and $r = q > 0$. If $f \in L^1$ and $g \in \text{weak}(L^q)$, then $f * g \in \text{weak}(L^q)$ and there exists a constant C_q independent of f and g such that

$$[f * g]_q \leqslant C_q \|f\|_1 \|g\|_q.$$

Proof. To prove (a), let q be fixed. The special cases $p = 1, r = q$ and $p = q/(q-1), r = \infty$ are Theorem 12 and Proposition 13. The general case then follows from the Riesz-Thorin interpolation theorem. (See also Folland Exercise 8.6 for a direct proof.) (b) and (c) are special cases of Theorem 78. □

8.2.1 Smoothness of convolutions

One of the most important properties of convolution is that, roughly speaking, $f * g$ is at least as smooth as either f or g, because formally we have

$$\partial^\alpha (f * g)(x) = \partial^\alpha \int f(x-y)g(y)dy = \int \partial^\alpha f(x-y)g(y)dy = (\partial^\alpha f) * g(x)$$

and similarly $\partial^\alpha(f * g) = f * (\partial^\alpha g)$. To make this precise, one needs only to impose conditions on f and g so that differentiation under the integral sign is legitimate. One such result is the following; see also Exercises 23 and 24.

Proposition 8.15: 8.10.

If $f \in L^1$, $g \in C^k$, and $\partial^\alpha g$ is bounded for $|\alpha| \leqslant k$, then $f * g \in C^k$ and for all $|\alpha| \leqslant k$,

$$\partial^\alpha (f * g) = f * (\partial^\alpha g).$$

Proof. This is clear from Theorem 50. □

Proposition 8.16: 8.11.

If $f, g \in \mathcal{S}$, then $f * g \in \mathcal{S}$.

Proof. First, $f * g \in C^\infty$ by Proposition 15. Since

$$1 + |x| \leqslant 1 + |x-y| + |y| \leqslant (1+|x-y|)(1+|y|) \tag{8.16.1}$$

we have

$$\begin{aligned}(1+|x|)^N |\partial^\alpha (f * g)(x)| &\leqslant \int (1+|x-y|)^N |\partial^\alpha f(x-y)| (1+|y|)^N |g(y)| dy \\ &\leqslant \|f\|_{(N,\alpha)} \|g\|_{(N+n+1,\alpha)} \int (1+|y|)^{-n-1} dy\end{aligned}$$

which is finite by Corollary 101. □

Convolutions of functions on the torus $\mathbb{T}^n$ are defined just as for functions on $\mathbb{R}^n$. (If one regards functions on $\mathbb{T}^n$ as periodic functions on $\mathbb{R}^n$, of course, the integration is to be extended over the unit cube rather than $\mathbb{T}^n$.) All of the preceding results remain valid, with the same proofs.

8.2.2 Approximate identities

The following theorem underlies many of the important applications of convolutions on $\mathbb{R}^n$. We introduce a bit of notation that will be used frequently hereafter: If ϕ is any

function on $\mathbb{R}^n$ and $t > 0$, we set
$$\phi_t(x) := \frac{1}{t^n}\phi(x/t).$$
We observe that if $\phi \in L^1$, then $\int \phi_t$ is independent of t by Theorem 87, since
$$\int \phi_t = \int \phi(t^{-1}x)\frac{1}{t^n}dx = \int \phi(y)dy = \int \phi.$$
Moreover, the "mass" of ϕ_t becomes concentrated at the origin as $t \to 0$. (Draw a picture if this isn't clear.)

Theorem 8.17: 8.14.

Suppose $p \in [1, \infty)$, $\phi \in L^1$, and $\int \phi(x)dx = a$.

(a) If $p \in [1, \infty)$ and $f \in L^p$, then
$$f * \phi_t \to af \text{ in } L^p \text{ as } t \searrow 0.$$
(b) If f is bounded and uniformly continuous, then
$$f * \phi_t \to af \text{ uniformly as } t \searrow 0.$$
(c) If $f \in L^\infty$ and f is continuous on an open set U, then
$$f * \phi_t \to af \text{ uniformly on compact subsets of } U \text{ as } t \to 0.$$

Proof. Setting $y = tz$, we have
$$\begin{aligned} f * \phi_t(x) - af(x) &= \int [f(x-y) - f(x)]\phi_t(y)dy \\ &= \int [f(x-tz) - f(x)]\phi(z)dz = \int [\tau_{tz}f(x) - f(x)]\phi(z)dz \end{aligned}$$
Apply Minkowski's inequality for integrals:
$$\|f * \phi_t - af\|_p \leqslant \int \|\tau_{tz}f - f\|_p |\phi(z)|dz$$
Now, $\|\tau_{tz}f - f\|_p$ is bounded by $2\|f\|_p$ and tends to 0 as $t \to 0$ for each z, by Proposition 4. Assertion (a) therefore follows from the dominated convergence theorem.

The proof of (b) is exactly the same, with $\|\cdot\|_p$ replaced by $\|\cdot\|_u$. The estimate for $\|f * \phi_t - af\|_u$ is obvious, and $\|\tau_{tz}f - f\|_u \to 0$ as $t \to 0$ by the uniform continuity of f.

As for (c), given $\varepsilon > 0$ let us choose a compact $E \subset \mathbb{R}^n$ such that $\int_{E^c} |\phi| < \varepsilon$. Also, let K be a compact subset of U. If t is sufficiently small, then, we will have $x - tz \in U$ for all $x \in K$ and $z \in E$, so from the compactness of K it follows as in Lemma 3 that
$$\sup_{x\in K, z\in E} |f(x-tz) - f(x)| < \varepsilon$$

for small t. But then

$$\sup_{x\in K}|f*\phi_t(x)-af(x)| \leqslant \sup_{x\in K}\left[\int_E+\int_{E^c}\right]|f(x-tz)-f(x)||\phi(z)|dz$$
$$\leqslant \varepsilon\int|\phi|+2\|f\|_\infty\varepsilon$$

from which (c) follows. □

If we impose slightly stronger conditions on ϕ, we can also show that $f*\phi_t \to af$ almost everywhere for $f\in L^p$. The device in the following proof of breaking up an integral into pieces corresponding to the dyadic intervals $[2^k,2^{k+1}]$ and estimating each piece separately is a standard trick of the trade in Fourier analysis.

Theorem 8.18: 8.15.

Suppose $|\phi(x)|\leqslant C(1+|x|)^{-n-\varepsilon}$ for some $C,\varepsilon>0$ (which implies that $\phi\in L^1$ by Corollary 101), and $\int\phi(x)dx=a$. If $f\in L^p(1\leqslant p\leqslant\infty)$, then $f*\phi_t(x)\to af(x)$ as $t\to 0$ for every x in the Lebesgue set of f—in particular, for almost every x, and for every x at which f is continuous.

Proof. If x is in the Lebesgue set of f, for any $\delta>0$ there exists $\eta>0$ such that

$$\int_{|y|<r}|f(x-y)-f(x)|dy\leqslant\delta r^n \text{ for } r\leqslant\eta. \tag{8.18.1}$$

Let us set

$$I_1=\int_{|y|<\eta}|f(x-y)-f(x)||\phi_t(y)|dy \quad\text{and}\quad I_2=\int_{|y|\geqslant\eta}|f(x-y)-f(x)||\phi_t(y)|dy.$$

We claim that I_1 is bounded by $A\delta$ for some A independent of t, whereas $I_2\to 0$ as $t\to 0$. Since

$$|f*\phi_t(x)-af(x)|\leqslant I_1+I_2,$$

we will have

$$\limsup_{t\to 0}|f*\phi_t(x)-af(x)|\leqslant A\delta,$$

and since δ is arbitrary, this will complete the proof.

To estimate I_1, let K be the integer such that $2^K\leqslant\eta/t<2^{K+1}$ if $\eta/t\geqslant 1$, and $K=0$ if $\eta/t<1$. We view the ball $|y|<\eta$ as the union of the annuli $2^{-k}\eta\leqslant|y|<2^{1-k}\eta(1\leqslant k\leqslant K)$ and the ball $|y|<2^{-K}\eta$. On the kth annulus we use the estimate

$$|\phi_t(y)|\leqslant Ct^{-n}\left|\frac{y}{t}\right|^{-n-\varepsilon}\leqslant Ct^{-n}\left[\frac{2^{-k}\eta}{t}\right]^{-n-\varepsilon}$$

and on the ball $|y| < 2^{-K}\eta$ we use the estimate $|\phi_t(y)| \leqslant Ct^{-n}$. Thus

$$I_1 \leqslant \sum_1^K Ct^{-n}\left[\frac{2^{-k}\eta}{t}\right]^{-n-\varepsilon} \int_{2^{-k}\eta\leqslant|y|<2^{1-k}\eta} |f(x-y)-f(x)|dy + Ct^{-n}\int_{|y|<2^{-K}\eta} |f(x-y)-f(x)|dy.$$

Therefore, by Equation (8.18.1) and the fact that $2^K \leqslant \eta/t < 2^{K+1}$,

$$\begin{aligned} I_1 &\leqslant C\delta \sum_1^K (2^{1-k}\eta)^n t^{-n}\left[\frac{2^{-k}\eta}{t}\right]^{-n-\varepsilon} + C\delta t^{-n}(2^{-K}\eta)^n \\ &= 2^n C\delta\left[\frac{\eta}{t}\right]^{-\varepsilon} \sum_1^K 2^{k\varepsilon} + C\delta\left[\frac{2^{-K}\eta}{t}\right]^n \\ &= 2^n C\delta\left[\frac{\eta}{t}\right]^{-\varepsilon} \frac{2^{(K+1)\varepsilon}-2^\varepsilon}{2^\varepsilon-1} + C\delta\left[\frac{2^{-K}\eta}{t}\right]^n \\ &\leqslant 2^n C[2^\varepsilon(2^\varepsilon-1)^{-1}+1]\delta. \end{aligned}$$

As for I_2, if p' is the conjugate exponent to p and χ is the characteristic function of $\{y \mid |y| \geqslant \eta\}$, by Hölder's inequality we have

$$\begin{aligned} I_2 &\leqslant \int_{|y|\geqslant\eta} (|f(x-y)|+|f(x)|)|\phi_t(y)|dy \\ &\leqslant \|f\|_p\|\chi\phi_t\|_{p'} + |f(x)|\|\chi\phi_t\|_1 \end{aligned}$$

so it suffices to show that for $1 \leqslant q \leqslant \infty$, and in particular for $q=1$ and $q=p'$, $\|\chi\phi_t\|_q \to 0$ as $t \to 0$. If $q=\infty$, this is obvious:

$$\|\chi\phi_t\|_\infty \leqslant Ct^{-n}[1+(\eta/t)]^{-n-\varepsilon} = Ct^\varepsilon(t+\eta)^{-n-\varepsilon} \leqslant C\eta^{-n-\varepsilon}t^\varepsilon.$$

If $q<\infty$, by Corollary 100 we have

$$\begin{aligned} \|\chi\phi_t\|_q^q &= \int_{|y|\geqslant\eta} t^{-nq}|\phi(t^{-1}y)|^q dy = t^{n(1-q)}\int_{|z|\geqslant\eta/t} |\phi(z)|^q dz \\ &\leqslant C_1 t^{n(1-q)}\int_{\eta/t}^\infty r^{n-1-(n+\varepsilon)q}dr = C_2 t^{n(1-q)}\left[\frac{\eta}{t}\right]^{n-(n+\varepsilon)q} = C_3 t^{\varepsilon q}. \end{aligned}$$

In either case, $\|\chi\phi_t\|_q$ is dominated by t^ε, so we are done. □

In most of the applications of the preceding two theorems one has $a=1$, although the case $a=0$ is also useful. If $a=1$, $\{\phi_t\}_{t>0}$ is called an ***approximate identity***, as it furnishes an approximation to the identity operator on L^p by convolution operators. This construction is useful for approximating L^p functions by functions having specified regularity properties. For example, we have the following two important results:

Proposition 8.19: 8.17.

If $p \in [1,\infty)$, then C_c^∞ (and hence also $\mathcal{S}$) is dense in L^p and in C_0.

Proof. Given $f \in L^p$ and $\varepsilon > 0$, there exists $g \in C_c$ with $\|f - g\|_p < \varepsilon/2$, by **??**. Let ϕ be a function in C_c^∞ such that $\int \phi = 1$—for example, take $\phi = \left(\int \psi\right)^{-1} \psi$ where ψ is as in Equation (8.0.1). Then $g * \phi_t \in C_c^\infty$ by Proposition 11(d) and Proposition 15, and $\|g * \phi_t - g\|_p < \varepsilon/2$ for sufficiently small t by Theorem 17. The same argument applies if L^p is replaced by C_0, $\|\cdot\|_p$ by $\|\cdot\|_u$, and **??** by Proposition 116. □

Theorem 8.20: 8.18: The C^∞ Urysohn Lemma.

If $K \subset \mathbb{R}^n$ is compact and U is an open set containing K, there exists $f \in C_c^\infty$ such that $0 \leqslant f \leqslant 1$, $f = 1$ on K, and $\operatorname{supp}(f) \subset U$.

Proof. Let $\delta = \rho(K, U^c)$ (the distance from K to U^c, which is positive since K is compact), and let $V = \{x \mid \rho(x, K) < \delta/3\}$. Choose a nonnegative $\phi \in C_c^\infty$ such that $\int \phi = 1$ and $\phi(x) = 0$ for $|x| \geqslant \delta/3$ (for example, $\left(\int \psi\right)^{-1} \psi_{\delta/3}$ with ψ as in Equation (8.0.1)), and set $f = \chi_V * \phi$. Then $f \in C_c^\infty$ by Proposition 11(d) and Proposition 15, and it is easily checked that $0 \leqslant f \leqslant 1$, $f = 1$ on K, and $\operatorname{supp}(f) \subset \{x \mid \rho(x, K) \leqslant 2\delta/3\} \subset U$. □

Exercise 8.21: Folland Exercise 8.5.

If $s\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is defined by $s(x, y) = x - y$, then $s^{-1}(E)$ is Lebesgue measurable whenever E is Lebesgue measurable. (For $n = 1$, draw a picture of $s^{-1}(E) \subset \mathbb{R}^2$. It should be clear that after rotation through an angle $\pi/4$, $s^{-1}(E)$ becomes $F \times \mathbb{R}$ where $F = \left\{x \mid \sqrt{2}x \in E\right\}$, and Theorem 87 can be applied. The same idea works in higher dimensions.)

Exercise 8.22: Folland Exercise 8.6.

Prove Theorem 14(a) by using Folland Exercise 6.31 to show that

$$|f * g(x)|^r \leqslant \|f\|_p^{r-p} \|g\|_q^{r-q} \int |f(y)|^p |g(x - y)|^q dy$$

Exercise 8.23: Folland Exercise 8.7.

If f is locally integrable on $\mathbb{R}^n$ and $g \in C^k$ has compact support, then $f * g \in C^k$.

Exercise 8.24: Folland Exercise 8.8.

Suppose that $f \in L^p(\mathbb{R})$. If there exists $h \in L^p(\mathbb{R})$ such that

$$\lim_{y \to 0} \|y^{-1}(\tau_{-y} f - f) - h\|_p = 0$$

we call h the ***strong L^p derivative*** of f. If $f \in L^p(\mathbb{R}^n)$, strong L^p partial derivatives of f are defined similarly. Suppose that p and q are conjugate exponents, $f \in L^p$, $g \in L^q$, and the L^p derivative $\partial_j f$ exists. Then $\partial_j(f * g)$ exists (in the ordinary sense) and equals $(\partial_j f) * g$.

Exercise 8.25: Folland Exercise 8.9.

If $f \in L^p(\mathbb{R})$, the strong L^p derivative of f (call it h; see Folland Exercise 8.8) exists if and only if f is absolutely continuous on every bounded interval (perhaps after modification on a null set) and its pointwise derivative f' is in L^p, in which case $h = f'$ a.e. (For "only if," use Folland Exercise 8.8: If $g \in C_c$ with $\int g = 1$, then $f * g_t \to f$ and $(f * g_t)' \to h$ as $t \to 0$. For "if," write

$$\frac{f(x+y) - f(x)}{y} - f'(x) = \frac{1}{y}\int_0^y [f'(x+t) - f'(x)]dt$$

and use Minkowski's inequality for integrals.)

Exercise 8.26: Folland Exercise 8.10.

Let ϕ satisfy the hypotheses of Theorem 18. If $f \in L^p(1 \leqslant p \leqslant \infty)$, define the ***$\phi$-maximal function*** of f to be $M_\phi f(x) = \sup_{t>0}|f * \phi_t(x)|$. (Observe that the Hardy-Littlewood maximal function Hf is $M_\phi|f|$ where ϕ is the characteristic function of the unit ball divided by the volume of the ball.) Show that there is a constant C, independent of f, such that $M_\phi f \leqslant C \cdot Hf$. (Break up the integral $\int f(x-y)\phi_t(y)dy$ as the sum of the integrals over $|y| \leqslant t$ and over $2^k t < |y| \leqslant 2^{k+1}t$ $(k = 0, 1, 2, \ldots)$, and estimate ϕ_t on each region.) It follows from Theorem 44 that M_ϕ is weak type $(1, 1)$, and the proof of Theorem 45 can then be adapted to give an alternate demonstration that $f * \phi_t \to \left(\int \phi\right) f$ a.e.

Exercise 8.27: Folland Exercise 8.11.

Young's inequality shows that L^1 is a Banach algebra, the product being convolution.

(a) If $\mathfrak{J}$ is an ideal in the algebra L^1, so is its closure in L^1.

(b) If $f \in L^1$, the smallest closed ideal in L^1 containing f is the smallest closed subspace of L^1 containing all translates of f. (If $g \in C_c$, $f * g(x)$ can be approximated by sums $\sum f(x - y_j)g(y_j)\Delta y_j$. On the other hand, if $\{\phi_t\}$ is an approximate identity, $f * \tau_y(\phi_t) \to \tau_y f$ as $t \to 0$.)

8.3 The Fourier Transform

One of the fundamental principles of harmonic analysis is the exploitation of symmetry. To be more specific, if one is doing analysis on a space on which a group acts, it is a good idea to study functions (or other analytic objects) that transform in simple ways under the group action, and then try to decompose arbitrary functions as sums or integrals of these basic functions.

The spaces we are studying are $\mathbb{R}^n$ and $\mathbb{T}^n$, which are abelian groups under addition and act on themselves by translation. The building blocks of harmonic analysis on these spaces are the functions that transform under translation by multiplication by a factor of absolute value one, that is, functions f such that for each x there is a number $\phi(x)$ with $|\phi(x)| = 1$ such that $f(y+x) = \phi(x)f(y)$. If f and ϕ have this property, then $f(x) = \phi(x)f(0)$, so f is completely determined by ϕ once $f(0)$ is given; moreover,

$$\phi(x)\phi(y)f(0) = \phi(x)f(y) = f(x+y) = \phi(x+y)f(0)$$

so that (unless $f = 0$) $\phi(x+y) = \phi(x)\phi(y)$. In short, to find all fs that transform as described above, it suffices to find all ϕs of absolute value one that satisfy the functional equation $\phi(x+y) = \phi(x)\phi(y)$. Upon imposing the natural requirement that ϕ should be measurable, we have a complete solution to this problem.

Theorem 8.28: 8.19.

If ϕ is a measurable function on $\mathbb{T}^n$ (resp. $\mathbb{R}^n$) such that $\phi(x+y) = \phi(x)\phi(y)$ and $|\phi| = 1$, there exists $\xi \in \mathbb{T}^n$ (resp. $\xi \in \mathbb{R}^n$) such that $\phi(x) = e^{2\pi i\xi\cdot x}$.

Proof. We first prove this assertion on $\mathbb{R}$. Let $a \in \mathbb{R}$ be such that $\int_0^a \phi(t)dt \neq 0$; such an a surely exists, for otherwise the Lebesgue differentiation theorem would imply that $\phi = 0$ a.e. Setting $A = \left(\int_0^a \phi(t)dt\right)^{-1}$, then, we have

$$\phi(x) = A\int_0^a \phi(x)\phi(t)dt = A\int_0^a \phi(x+t)dt = A\int_x^{x+a} \phi(t)dt.$$

Thus ϕ, being the indefinite integral of a locally integrable function, is continuous; and then, being the integral of a continuous function, it is C^1. Moreover,

$$\phi'(x) = A[\phi(x+a) - \phi(x)] = B\phi(x), \text{ where } B = A[\phi(a) - 1].$$

It follows that $(d/dx)(e^{-Bx}\phi(x)) = 0$, so that $e^{-Bx}\phi(x)$ is constant. Since $\phi(0) = 1$, we have $\phi(x) = e^{Bx}$, and since $|\phi| = 1, B$ is purely imaginary, so $B = 2\pi i\xi$ for some $\xi \in \mathbb{R}$. This completes the proof for $\mathbb{R}$; as for $\mathbb{R}$, the ϕ we have been considering will be periodic (with period 1) if and only if $e^{2\pi i\xi} = 1$ if and only if $\xi \in \mathbb{R}$.

The n-dimensional case follows easily, for if $e_1, \ldots, e_n$ is the standard basis for $\mathbb{R}^n$, the functions $\psi_j(t) = \phi(te_j)$ satisfy $\psi_j(t+s) = \psi_j(t)\psi_j(s)$ on $\mathbb{R}$, so that $\psi_j(t) = e^{2\pi i\xi_j t}$, and hence

$$\phi(x) = \phi\left(\sum_1^n x_j e_j\right) = \prod_1^n \psi_j(x_j) = e^{2\pi i\xi\cdot x}. \qquad \square$$

8.3.1 Fourier transform on $\mathbb{T}^n$

The idea now is to decompose more or less arbitrary functions on $\mathbb{R}^n$ or $\mathbb{T}^n$ in terms of the exponentials $e^{2\pi i\xi\cdot x}$. In the case of $\mathbb{T}^n$ this works out very simply for L^2 functions:

Theorem 8.29: 8.20.

Let $E_\kappa(x) = e^{2\pi i\kappa\cdot x}$. Then $\{E_\kappa \mid \kappa \in \mathbb{Z}^n\}$ is an orthonormal basis of $L^2(\mathbb{T}^n)$.

Proof. Verification of orthonormality is an easy exercise in calculus; by Fubini's theorem it boils down to the fact that $\int_0^1 e^{2\pi ikt}dt$ equals 1 if $k = 0$ and equals 0 otherwise. Next, since $E_\kappa E_\lambda = E_{\kappa+\lambda}$, the set of finite linear combinations of the E_κs is an algebra. It clearly separates points on $\mathbb{T}^n$; also, $E_0 = 1$ and $\overline{E}_\kappa = E_{-\kappa}$. Since $\mathbb{T}^n$ is compact, the Stone-Weierstrass theorem implies that this algebra is dense in $C(\mathbb{T}^n)$ in the uniform norm and hence in the L^2 norm, and $C(\mathbb{T}^n)$ is itself dense in $L^2(\mathbb{T}^n)$ by **??**. It follows that $\{E_\kappa\}$ is a basis. □

We can restate Theorem 29 as follows. If $f \in L^2(\mathbb{T}^n)$, we define its ***Fourier transform*** $\widehat{f}$, a function on $\mathbb{T}^n$, by

$$\widehat{f}(\kappa) = \langle f, E_\kappa\rangle = \int_{\mathbb{T}^n} f(x)e^{-2\pi i\kappa\cdot x}dx,$$

and we call the series

$$\sum_{\kappa\in\mathbb{Z}^n} \widehat{f}(\kappa)E_\kappa$$

the ***Fourier series*** of f.[3] Theorem 29 then says that the Fourier transform maps $L^2(\mathbb{T}^n)$ onto $\ell^2(\mathbb{T}^n)$, that $\|\widehat{f}\|_2 = \|f\|_2$ (Parseval's identity), and that the Fourier series of f converges to f in the L^2 norm. We shall consider the question of pointwise convergence in the next two sections.

Actually, the definition of $\widehat{f}(\kappa)$ makes sense if f is merely in $L^1(\mathbb{T}^n)$, and $|\widehat{f}(\kappa)| \leqslant \|f\|_1$, so the Fourier transform extends to a norm-decreasing map from $L^1(\mathbb{T}^n)$ to $\ell^\infty(\mathbb{T}^n)$. (The Fourier series of an L^1 function may be quite badly behaved, but there are still methods for recovering f from $\widehat{f}$ when $f \in L^1$, as we shall see in the next section.) By interpolating between L^1 and L^2, we obtain the following result.

Theorem 8.30: 8.21: The Hausdorff-Young Inequality.

Suppose that $1 \leqslant p \leqslant 2$ and q is the conjugate exponent to p. If $f \in L^p(\mathbb{T}^n)$, then $\widehat{f} \in \ell^q(\mathbb{Z}^n)$ and $\|\widehat{f}\|_q \leqslant \|f\|_p$.

Proof. Since $\|\widehat{f}\|_\infty \leqslant \|f\|_1$ and $\|\widehat{f}\|_2 = \|f\|_2$ for $f \in L^1$ or $f \in L^2$, the assertion follows from the Riesz-Thorin interpolation theorem. □

[3]The term "Fourier transform" is also used to mean the map $f \mapsto \widehat{f}$.

8.3.2 Fourier transform on $\mathbb{R}^n$

The situation on $\mathbb{R}^n$ is more delicate. The formal analogue of Theorem 29 should be

$$f(x) = \int_{\mathbb{R}^n} \widehat{f}(\xi) e^{2\pi i \xi \cdot x} d\xi, \text{ where } \widehat{f}(\xi) = \int_{\mathbb{R}^n} f(x) e^{-2\pi i \xi \cdot x} dx$$

These relations turn out to be valid when suitably interpreted, but some care is needed. In the first place, the integral defining $\widehat{f}(\xi)$ is likely to diverge if $f \in L^2$. However, it certainly converges if $f \in L^1$. We therefore begin by defining the Fourier transform of $f \in L^1(\mathbb{R}^n)$ by

$$\mathcal{F} f(\xi) = \widehat{f}(\xi) = \int_{\mathbb{R}^n} f(x) e^{-2\pi i \xi \cdot x} dx$$

(We use the notation $\mathcal{F}$ for the Fourier transform only in certain situations where it is needed for clarity.) Clearly $\|\widehat{f}\|_u \leqslant \|f\|_1$, and $\widehat{f}$ is continuous by Theorem 50; thus

$$\mathcal{F} \colon L^1(\mathbb{R}^n) \to BC(\mathbb{R}^n)$$

We summarize the elementary properties of $\mathcal{F}$ in a theorem.

Theorem 8.31: 8.22.

Suppose $f, g \in L^1(\mathbb{R}^n)$.

(a) $(\tau_y f)^\wedge(\xi) = e^{-2\pi i \xi \cdot y} \widehat{f}(\xi)$ and $\tau_\eta(\widehat{f}) = \widehat{h}$ where $h(x) = e^{2\pi i \eta \cdot x} f(x)$.

(b) If T is an invertible linear transformation of $\mathbb{R}^n$ and $S = (T^*)^{-1}$ is its inverse transpose, then $(f \circ T)^\wedge = |\det T|^{-1} \widehat{f} \circ S$. In particular, if T is a rotation, then $(f \circ T)^\wedge = \widehat{f} \circ T$; and if $Tx = t^{-1}x (t > 0)$, then $(f \circ T)^\wedge(\xi) = t^n \widehat{f}(t\xi)$, so that $(f_t)^\wedge(\xi) = \widehat{f}(t\xi)$ in the notation of (8.13).

(c) $(f * g)^\wedge = \widehat{f}\widehat{g}$.

(d) If $x^\alpha f \in L^1$ for $|\alpha| \leqslant k$, then $\widehat{f} \in C^k$ and $\partial^\alpha \widehat{f} = ((-2\pi i x)^\alpha f)^\wedge$.

(e) If $f \in C^k$, $\partial^\alpha f \in L^1$ for $|\alpha| \leqslant k$, and $\partial^\alpha f \in C_0$ for $|\alpha| \leqslant k-1$, then $(\partial^\alpha f)^\wedge(\xi) = (2\pi i \xi)^\alpha \widehat{f}(\xi)$.

(f) (The Riemann–Lebesgue Lemma) $\mathcal{F}(L^1(\mathbb{R}^n)) \subset C_0(\mathbb{R}^n)$.

Proof.

(a) We have

$$(\tau_y f)(\xi) = \int f(x-y) e^{-2\pi i \xi \cdot x} dx = \int f(x) e^{-2\pi i \xi \cdot (x+y)} dx = e^{-2\pi i \xi \cdot y} \widehat{f}(\xi)$$

and similarly for the other formula.

(b) By Theorem 87,

$$\begin{aligned}(f \circ T)\widehat{(\xi)} &= \int f(Tx) e^{-2\pi i \xi \cdot x} dx = |\det T|^{-1} \int f(x) e^{-2\pi i \xi \cdot T^{-1} x} dx \\ &= |\det T|^{-1} \int f(x) e^{-2\pi i S\xi \cdot x} dx = |\det T|^{-1} \widehat{f}(S\xi)\end{aligned}$$

(c) By Fubini's theorem,

$$\begin{aligned}
(f * g)\widehat{(\xi)} &= \iint f(x-y)g(y)e^{-2\pi i\xi\cdot x}dydx \\
&= \iint f(x-y)e^{-2\pi i\xi\cdot(x-y)}g(y)e^{-2\pi i\xi\cdot y}dxdy \\
&= \widehat{f}(\xi)\int g(y)e^{-2\pi i\xi\cdot y}dy \\
&= \widehat{f}(\xi)\widehat{g}(\xi).
\end{aligned}$$

(d) By Theorem 50 and induction on $|\alpha|$,

$$\partial^\alpha \widehat{f}(\xi) = \partial_\xi^\alpha \int f(x)e^{-2\pi i\xi\cdot x}dx = \int f(x)(-2\pi ix)^\alpha e^{-2\pi i\xi\cdot x}dx$$

(e) First assume $n = |\alpha| = 1$. Since $f \in C_0$, we can integrate by parts:

$$\begin{aligned}
\int f'(x)e^{-2\pi i\xi\cdot x}dx &= f(x)e^{-2\pi i\xi\cdot x}|_{-\infty}^{\infty} - \int f(x)(-2\pi i\xi)e^{-2\pi i\xi\cdot x}dx \\
&= 2\pi i\xi\widehat{f}(\xi).
\end{aligned}$$

The argument for $n > 1, |\alpha| = 1$ is the same—to compute $(\partial_j f)^\vee$, integrate by parts in the jth variable—and the general case follows by induction on $|\alpha|$.

(f) By (e), if $f \in C^1 \cap C_c$, then $|\xi|\widehat{f}(\xi)$ is bounded and hence $\widehat{f} \in C_0$. But the set of all such fs is dense in L^1 by Proposition 19, and $\widehat{f}_n \to \widehat{f}$ uniformly whenever $f_n \to f$ in L^1. Since C_0 is closed in the uniform norm, the result follows. Continuity of $\widehat{f}$ follows from the DCT, so we only need to show $\widehat{f}$ vanishes at infinity. For $f := \chi_{[a_1,b_1]\times\cdots\times[a_n,b_n]}$, we have

$$\begin{aligned}
\widehat{f}(\xi) &= \int_{\mathbb{R}^n} e^{-2\pi i\langle x,\xi\rangle}\xi_{[\cdots]}(x)dx \\
&= \prod_{k=1}^n \int_{a_k}^{b_k} e^{-2\pi ix_k\cdot\xi_k}dx_k = \prod_{k=1}^n \frac{1}{-2\pi i\xi_k}(e^{2\pi ib_k\xi_k} - e^{-2\pi ia_k\xi_k}),
\end{aligned}$$

which tends to 0 as $|\xi| \to \infty$. Next, if $f \in L^1(\mathbb{R}^n)$, pick simple functions $\{\phi_j\}_{j=1}^\infty$ such that $\|f - \phi_j\|_1 \to 0$ as $j \to \infty$.

$$|\widehat{f}(\xi)| \leqslant |\widehat{f}(\xi) - \widehat{\phi}_j(\xi)| + |\widehat{\phi}_j(\xi)| \leqslant \|f - \phi_j\|_1 + |\widehat{\phi}_j(\xi)|,$$

and both terms vanish as $j \to \infty$. □.

Parts (d) and (e) of Theorem 31 point to a fundamental property of the Fourier transform: Smoothness properties of f are reflected in the rate of decay of $\widehat{f}$ at infinity, and vice versa. Parts (a), (c), (e), and (f) of this theorem are valid also on $\mathbb{T}^n$, as is (b) provided that T leaves the lattice $\mathbb{T}^n$ invariant (Folland Exercise 8.12).

Corollary 8.32: 8.23.

$\mathcal{F}$ maps the Schwartz class $\mathcal{S}$ continuously into itself.

Proof. If $f \in \mathcal{S}$, then $x^\alpha \partial^\beta f \in L^1 \cap C_0$ for all α, β, so by Theorem 31d,e, $\widehat{f}$ is C^∞ and
$$(x^\alpha \partial^\beta f)^\wedge = (-1)^{|\alpha|}(2\pi i)^{|\beta|-|\alpha|}\partial^\alpha(\xi^\beta \widehat{f}).$$
Thus $\partial^\alpha(\xi^\beta \widehat{f})$ is bounded for all α, β, whence $\widehat{f} \in \mathcal{S}$ by Proposition 2. Moreover, since $\int (1+|x|)^{-n-1}dx < \infty$,
$$\|(x^\alpha \partial^\beta f)\|_u \leqslant \|x^\alpha \partial^\beta f\|_1 \leqslant C\|(1+|x|)^{n+1}x^\alpha \partial^\beta f\|_u.$$
It then follows that $\|\widehat{f}\|_{(N,\beta)} \leqslant C_{N,\beta}\sum_{|\gamma|\leqslant|\beta|}\|f\|_{(N+n+1,\gamma)}$ by the proof of Proposition 2, so the Fourier transform is continuous on $\mathcal{S}$. □

At this point we need to compute an important specific Fourier transform.

Proposition 8.33: 8.24.

If $f(x) = e^{-\pi a|x|^2}$ where $a > 0$, then $\widehat{f}(\xi) = a^{-n/2}e^{-\pi|\xi|^2/a}$.

Proof. First consider the case $n = 1$. Since the derivative of $e^{-\pi a x^2}$ is $-2\pi a e^{-\pi a x^2}$, by Theorem 31(d,e) we have
$$(\widehat{f})'(\xi) = (-2\pi i x e^{-\pi a x^2})^\wedge(\xi) = \frac{i}{a}(f')^\wedge \xi = \frac{i}{a}(2\pi i\xi)\widehat{f}(\xi) = -\frac{2\pi}{a}\xi\widehat{f}(\xi)$$
It follows from the product rule that $(d/d\xi)(e^{\pi\xi^2/a}\widehat{f}(\xi)) = 0$, so that $e^{\pi\xi^2/a}\widehat{f}(\xi)$ is constant. To evaluate the constant, set $\xi = 0$ and use Proposition 102:
$$\widehat{f}(0) = \int e^{-\pi a x^2}dx = a^{-1/2}$$
The n-dimensional case follows by Fubini's theorem, since $|x|^2 = \sum_1^n x_j^2$:
$$\begin{aligned}\widehat{f}(\xi) &= \prod_1^n \int \exp(-\pi a x_j^2 - 2\pi i\xi_j x_j)dx_j \\ &= \prod_1^n [a^{-1/2}\exp(-\pi\xi_j^2/a)] = a^{-n/2}\exp(-\pi|\xi|^2/a).\end{aligned}$$
□

We are now ready to invert the Fourier transform. If $f \in L^1$, we define
$$f^\vee(x) = \widehat{f}(-x) = \int f(\xi)e^{2\pi i\xi\cdot x}d\xi$$
and we claim that if $f \in L^1$ and $\widehat{f} \in L^1$ then $(\widehat{f})^\vee = f$. A simple appeal to Fubini's theorem fails because the integrand in
$$(\widehat{f})^\vee(x) = \iint f(y)e^{-2\pi i\xi\cdot y}e^{2\pi i\xi\cdot x}dyd\xi$$
is not in $L^1(\mathbb{R}^n \times \mathbb{R}^n)$. The trick is to introduce a convergence factor and then pass to the limit, using Fubini's theorem via the following lemma.

Lemma 8.34: 8.25.

If $f, g \in L^1$, then $\int \widehat{f} g = \int f \widehat{g}$.

Proof. Both integrals are equal to $\iint f(x) g(\xi) e^{-2\pi i \xi \cdot x} dx d\xi$. □

Theorem 8.35: 8.26: The Fourier Inversion Theorem.

If $f \in L^1$ and $\widehat{f} \in L^1$, then f agrees almost everywhere with a continuous function f_0, and $(\widehat{f})^\vee = (f^\vee)^\wedge = f_0$.

Proof. Given $t > 0$ and $x \in \mathbb{R}^n$, set

$$\phi(\xi) = \exp(2\pi i \xi \cdot x - \pi t^2 |\xi|^2)$$

By Theorem 31(a) and Proposition 33,

$$\widehat{\phi}(y) = t^{-n} \exp(-\pi |x - y|^2 / t^2) = g_t(x - y)$$

where $g(x) = e^{-\pi |x|^2}$ and the subscript t has the meaning in (8.13). By Lemma 34, then,

$$\int e^{-\pi t^2 |\xi|^2} e^{2\pi i \xi \cdot x} \widehat{f}(\xi) d\xi = \int \widehat{f} \phi = \int f \widehat{\phi} = f * g_t(x)$$

Since $\int e^{-\pi |x|^2} dx = 1$, by Theorem 17 we have $f * g_t \to f$ in the L^1 norm as $t \to 0$. On the other hand, since $\widehat{f} \in L^1$ the dominated convergence theorem yields

$$\lim_{t \to 0} \int e^{-\pi t^2 |\xi|^2} e^{2\pi i \xi \cdot x} \widehat{f}(\xi) d\xi = \int e^{2\pi i \xi \cdot x} \widehat{f}(\xi) d\xi = (\widehat{f})^\vee(x)$$

It follows that $f = (\widehat{f})^\vee$ a.e., and similarly $(f^\vee)^\wedge = f$ a.e. Since $(\widehat{f})^\vee$ and $(f^\vee)^\wedge$ are continuous, being Fourier transforms of L^1 functions, the proof is complete. □

Corollary 8.36: 8.27.

If $f \in L^1$ and $\widehat{f} = 0$, then $f = 0$ a.e.

Theorem 8.37: 8.28: Corollary.

$\mathcal{F}$ is an isomorphism of $\mathcal{S}$ onto itself.

Proof. By Corollary 32, $\mathcal{F}$ maps $\mathcal{F}$ continuously into itself, and hence so does $f \mapsto f^\vee$, since $f^\vee(x) = \widehat{f}(-x)$. By the Fourier inversion theorem, these maps are inverse to each other. □

At last we are in a position to derive the analogue of Theorem 29 on $\mathbb{R}^n$.

Theorem 8.38: 8.29: The Plancherel Theorem.

If $f \in L^1 \cap L^2$, then $\widehat{f} \in L^2$; and $\mathcal{F}|_{L^1 \cap L^2}$ extends uniquely to a unitary isomorphism on L^2.

Proof. Let $X = \{f \in L^1 \mid \widehat{f} \in L^1\}$. Since $\widehat{f} \in L^1$ implies $f \in L^\infty$, we have $X \subset L^2$ by Proposition 16, and X is dense in L^2 because $\mathcal{S} \subset X$ and $\mathcal{S}$ is dense in L^2 by Proposition 19. Given $f, g \in X$, let $h = \overline{\widehat{g}}$. By the inversion theorem,

$$\widehat{h}(\xi) = \int e^{-2\pi i\xi\cdot x}\overline{\widehat{g}(x)}dx = \int \overline{e^{2\pi i\xi\cdot x}\widehat{g}(x)}dx = \overline{g(\xi)}$$

Hence, by Lemma 34,

$$\int f\overline{g} = \int f\widehat{h} = \int \widehat{f}h = \widehat{f}\overline{\widehat{g}}.$$

Thus $\mathcal{F}|_X$ preserves the L^2 inner product; in particular, by taking $g = f$, we obtain $\|\widehat{f}\|_2 = \|f\|_2$. Since $\mathcal{F}(X) = X$ by the Fourier inversion theorem, $\mathcal{F}|_X$ extends by continuity to a unitary isomorphism on L^2.

It remains only to show that this extension agrees with $\mathcal{F}$ on all of $L^1 \cap L^2$. But if $f \in L^1 \cap L^2$ and $g(x) = e^{-\pi|x|^2}$ as in the proof of the inversion theorem, we have $f * g_t \in L^1$ by Young's inequality and $(f * g_t)\widehat{e} \in L^1$ because $\widehat{(f * g_t)}((\xi) = e^{-\pi t^2|\xi|^2}\widehat{f}(\xi)$ and $\widehat{f}$ is bounded. Hence $f * g_t \in X$; moreover, by Theorem 17, $f * g_t \to f$ in both the L^1 and L^2 norms. Therefore $(f * g_t) \to \widehat{f}$ both uniformly and in the L^2 norm, and we are done. ☐

We have thus extended the domain of the Fourier transform from L^1 to $L^1 + L^2$. Just as on $\mathbb{T}^n$, the Riesz-Thorin theorem yields the following result for the intermediate L^p spaces.

Theorem 8.39: 8.30: The Hausdorff-Young Inequality.

Suppose that $1 \leqslant p \leqslant 2$ and q is the conjugate exponent to p. If $f \in L^p(\mathbb{R}^n)$, then $\widehat{f} \in L^q(\mathbb{R}^n)$ and $\|\widehat{f}\|_q \leqslant \|f\|_p$.

If $f \in L^1$ and $\widehat{f} \in L^1$, the inversion formula

$$f(x) = \int \widehat{f}(\xi)e^{2\pi i\xi\cdot x}d\xi$$

exhibits f as a superposition of the basic functions $e^{2\pi i\xi\cdot x}$; it is often called the ***Fourier integral representation*** of f. This formula remains valid in spirit for all $f \in L^2$, although the integral (as well as the integral defining $\widehat{f}$) may not converge pointwise. The interpretation of the inversion formula will be studied further in the next section.

We conclude this section with a beautiful theorem that involves an interplay of Fourier series and Fourier integrals. To motivate it, consider the following problem: Given a function $f \in L^1(\mathbb{R}^n)$, how can one manufacture a periodic function (that is, a function on $\mathbb{R}^n$) from it? Two possible answers suggest themselves. One way is to "average" f over all periods, producing the series $\sum_{k\in\mathbb{R}^n} f(x-k)$. This series, if it converges, will surely define a periodic function. The other way is to restrict f to the lattice $\mathbb{R}^n$ and use it to form a Fourier series $\sum_{\kappa\in\mathbb{R}^n} \widehat{f}(\kappa)e^{2\pi i\kappa\cdot x}$. The content of the following theorem is that these methods both work and both give the same answer.

Theorem 8.40: 8.31.

If $f \in L^1(\mathbb{R}^n)$, the series $\sum_{k\in\mathbb{R}^n} \tau_k f$ converges pointwise a.e. and in $L^1(\mathbb{R}^n)$ to a function Pf such that $\|Pf\|_1 \leqslant \|f\|_1$. Moreover, for $\kappa \in \mathbb{R}^n$, $(Pf)^\wedge(\kappa)$ (Fourier transform on $\mathbb{R}^n$) equals $\widehat{f}(\kappa)$ (Fourier transform on $\mathbb{R}^n$).

Proof. Let $Q = \left[-\frac{1}{2}, \frac{1}{2}\right)^n$. Then $\mathbb{R}^n$ is the disjoint union of the cubes $Q + k = \{x + k \mid x \in Q\}, k \in \mathbb{R}^n$, so

$$\int_Q \sum_{k\in\mathbb{Z}^n} |f(x-k)|dx = \sum_{k\in\mathbb{Z}^n} \int_{Q+k} |f(x)|dx = \int_{\mathbb{Z}^n} |f(x)|dx$$

Now apply Theorem 48. First, it shows that the series $\sum \tau_k f$ converges a.e. and in $L^1(\mathbb{T}^n)$ to a function $Pf \in L^1(\mathbb{T}^n)$ such that $\|Pf\|_1 \leqslant \|f\|_1$, since $\mathbb{T}^n$ is measure-theoretically identical to Q. Second, it yields

$$\begin{aligned}\mathcal{F}(Pf)(\kappa) &= \int_Q \sum_{k\in\mathbb{Z}^n} f(x-k)e^{-2\pi i\kappa\cdot x}dx = \sum_{k\in\mathbb{Z}^n} \int_{Q+k} f(x)e^{-2\pi i\kappa\cdot(x+k)}dx \\ &= \sum_{k\in\mathbb{Z}^N} \int_{Q+k} f(x)e^{-2\pi i\kappa\cdot x}dx = \int_{\mathbb{Z}^n} f(x)e^{-2\pi i\kappa\cdot x}dx = \widehat{f}(\kappa).\end{aligned}$$

□

If we impose conditions on f to guarantee that the series in question converge absolutely, we obtain a more refined result.

Theorem 8.41: 8.32: The Poisson Summation Formula.

Suppose $f \in C(\mathbb{R}^n)$ satisfies $|f(x)| \leqslant C(1+|x|)^{-n-\varepsilon}$ and $|\widehat{f}(\xi)| \leqslant C(1+|\xi|)^{-n-\varepsilon}$ for some $C, \varepsilon > 0$. Then

$$\sum_{k\in\mathbb{Z}^n} f(x+k) = \sum_{\kappa\in\mathbb{Z}^n} \widehat{f}(\kappa)e^{2\pi i\kappa\cdot x}$$

where both series converge absolutely and uniformly on $\mathbb{T}^n$. In particular, taking $x = 0$,

$$\sum_{k\in\mathbb{Z}^n} f(k) = \sum_{\kappa\in\mathbb{Z}^n} \widehat{f}(\kappa)$$

Proof. The absolute and uniform convergence of the series follows from the fact that $\sum_{k\in\mathbb{Z}^n}(1+|k|)^{-n-\varepsilon} < \infty$, which can be seen by comparing the latter series to the convergent integral $\int(1+|x|)^{-n-\varepsilon}dx$. Thus the function Pf $= \sum_k \tau_k f$ is in $C(\mathbb{Z}^n)$ and hence in $L^2(\mathbb{Z}^n)$, so Theorem 56 implies that the series $\sum \widehat{f}(\kappa)e^{2\pi i\kappa\cdot x}$ converges in $L^2(\mathbb{Z}^n)$ to Pf. Since it also converges uniformly, its sum equals Pf pointwise. (The replacement of k by $-k$ in the formula for Pf is immaterial since the sum is over all $k \in \mathbb{Z}^n$.) □

Exercise 8.42: Folland Exercise 8.12.

Work out the analogue of Theorem 31 for the Fourier transform on $\mathbb{T}^n$.

Exercise 8.43: Folland Exercise 8.13.

Let $f(x) = \frac{1}{2} - x$ on the interval $[0, 1)$, and extend f to be periodic on $\mathbb{R}$.

(a) $\widehat{f}(0) = 0$, and $\widehat{f}(\kappa) = (2\pi i\kappa)^{-1}$ if $\kappa \neq 0$.

(b) $\sum_1^\infty k^{-2} = \pi^2/6$. (Use the Parseval identity.)

Solution.

(a) First note $f \in L^2(\mathbb{T})$, since

$$\|f\|_2^2 = \int_{\mathbb{T}} |f(x)|^2 \, \mathrm{d}x = \left[\frac{x}{4} - \frac{x^2}{2} + \frac{x^3}{3}\right]_{x=0}^{x=1} = \frac{1}{12}. \tag{8.43.1}$$

We have

$$\widehat{f}(0) = \int_{\mathbb{T}} f(x)e^{-2\pi i 0\cdot x} \, \mathrm{d}x = \int_0^1 f(x) = \frac{1}{2} - \left[\frac{x^2}{2}\right]_{x=0}^{x=1} = 0,$$

and if $k \in \mathbb{Z} \setminus \{0\}$,

$$\begin{aligned}\widehat{f}(k) &= \int_{\mathbb{T}} \left(\frac{1}{2} - x\right) e^{-2\pi ikx} \, \mathrm{d}x = \int_0^1 \frac{1}{2} e^{-2\pi ikx} \, \mathrm{d}x - \int_0^1 x e^{-2\pi ikx} \, \mathrm{d}x \\ &= \frac{-1}{4\pi ik} - \left[\frac{x}{-2\pi ike^{-2\pi ikx}}\right]_{x=0}^{x=1} + \frac{-1}{2\pi ik} \int_0^1 e^{-2\pi ikx} \, \mathrm{d}x \\ &= \frac{-1}{4\pi ik} - \frac{-1}{2\pi ik} + \frac{1}{4\pi ik} = \frac{1}{2\pi ik}.\end{aligned}$$

(b) By part (a) $|\widehat{f}(k)|^2 = 1/(4\pi^2 k^2)$, so by Plancherel's theorem

$$\sum_{k=1}^\infty \frac{1}{k^2} = 4\pi^2 \sum_{k=0}^\infty |\widehat{f}(k)|^2 = 2\pi^2 \sum_{k\in\mathbb{Z}} |\widehat{f}(k)|^2 = 2\pi^2 \|f\|_2^2 \overset{(8.43.1)}{=} \frac{\pi^2}{6}.$$ □

Exercise 8.44: Folland Exercise 8.14.

(Wirtinger's Inequality) If $f \in C^1([a, b])$ and $f(a) = f(b) = 0$, then

$$\int_a^b |f(x)|^2 dx \leqslant \left(\frac{b-a}{\pi}\right)^2 \int_a^b |f'(x)|^2 dx$$

(By a change of variable it suffices to assume $a = 0, b = \frac{1}{2}$. Extend f to $\left[-\frac{1}{2}, \frac{1}{2}\right]$ by setting $f(-x) = -f(x)$, and then extend f to be periodic on $\mathbb{R}$. Check that f, thus extended, is in $C^1(\mathbb{R})$ and apply the Parseval identity.)

Exercise 8.45: Folland Exercise 8.15.

Let $\operatorname{sinc} x = (\sin \pi x)/\pi x (\operatorname{sinc} 0 = 1)$.

(a) If $a > 0, \widehat{\chi}_{[-a,a]}(x) = \chi^{\vee}_{[a,a]}(x) = 2a \operatorname{sinc} 2ax$.

(b) Let $\mathcal{H}_a = \{f \in L^2 \mid \widehat{f}(\xi) = 0 \text{ (a.e.) for } |\xi| > a\}$. Then $\mathcal{H}$ is a Hilbert space and $\{\sqrt{2a}\operatorname{sinc}(2ax - k) \mid k \in \mathbb{Z}\}$ is an orthonormal basis for $\mathcal{H}$.

(c) (The Sampling Theorem) If $f \in \mathcal{H}_a$, then $f \in C_0$ (after modification on a null set), and $f(x) = \sum_{-\infty}^{\infty} f(k/2a)\operatorname{sinc}(2ax - k)$, where the series converges both uniformly and in L^2. (In the terminology of signal analysis, a signal of bandwidth $2a$ is completely determined by sampling its values at a sequence of points $\{k/2a\}$ whose spacing is the reciprocal of the bandwidth.)

Solution.

(a) We have

$$\widehat{\chi}_{[-a,a]}(\xi) = \int_{-a}^{a} e^{-2\pi i\xi x}\,dx = \frac{-1}{2\pi i\xi}(e^{-2\pi i\xi a} - e^{2\pi i\xi a}) = \frac{\sin(2\pi a\xi)}{\pi\xi} = 2a\operatorname{sinc}(2a\xi)$$

and, by changing variables $x \mapsto -x$ in the integrand of $\chi^{\vee}_{[-a,a]}(\xi)$, we find

$$\chi^{\vee}_{[-a,a]}(\xi) = \int_{-a}^{a} e^{2\pi i\xi x} = -\int_{a}^{-a} e^{-2\pi i\xi x}\,dx = \int_{-a}^{a} e^{-2\pi i\xi x}\,dx = \chi^{\wedge}_{[-a,a]}(\xi).$$

(b) $\mathscr{H}_a$ is a linear subspace: If $f, g \in \mathscr{H}_a$ and $\lambda \in \mathbb{C}$, then for all $|\xi| > a$ we have $\widehat{f}(\xi) = \widehat{g}(\xi) = 0$, so

$$(f + \lambda g)^{\wedge}(\xi) = \widehat{f}(\xi) + \lambda\widehat{g}(\xi) = 0 + \lambda 0 = 0.$$

Thus $f + \lambda g \in \mathscr{H}_a$, so $\mathscr{H}_a$ is a linear subspace of L^2.

$\mathscr{H}_a$ is closed: Suppose $\{f_n\}_{n=1}^{\infty} \subset \mathscr{H}_a$ and $\|f_n - f\|_2 \to 0$. Since the Fourier transform is unitary on L^2 (hence an isometry), $\|\widehat{f_n} - \widehat{f}\|_2 \to 0$, that is, $f_n \to f$ in L^2. Thus there exists a subsequence $\widehat{f}_{n_k} \to \widehat{f}$ pointwise a.e., so for a.e. $x \in \mathbb{R}$, we have for $|\xi| > a$

$$\widehat{f}(\xi) = \lim_{k\to\infty} \widehat{f}_{n_k}(x) = \lim_{k\to\infty} 0 = 0.$$

Thus $f \in \mathscr{H}_a$, so $\mathscr{H}_a$ is a closed linear subspace of the Hilbert space L^2, and thus $\mathscr{H}_a$ is a Hilbert space.

Now for $k \in \mathbb{Z}$ and $x \in \mathbb{R}$, define $\zeta_k(x) := \sqrt{2a}\operatorname{sinc}(2ax - k)$. We claim $\{\zeta_k\}_{k\in\mathbb{Z}}$ is an orthonormal basis of $\mathscr{H}_a$. We first show $\{\zeta_k\}_{k\in\mathbb{Z}} \subset \mathscr{H}_a$. For any $k \in \mathbb{Z}$,

$$\zeta_k(x) = \sqrt{2a}\operatorname{sinc}(2ax - k) = \frac{1}{\sqrt{2a}}(2a\operatorname{sinc}(2a(x - k/2a))) \overset{(a)}{=} \frac{1}{\sqrt{2a}}\chi^{\vee}_{[-a,a]}(x - k/2a). \tag{8.45.1}$$

Taking the Fourier transform, we obtain

$$\zeta_k^\wedge(\xi) = \frac{1}{\sqrt{2a}}(\tau_{k/2a}\chi_{[-a,a]}^\vee)^\wedge(\xi) = \frac{e^{-2\pi i\xi(k/2a)}}{\sqrt{2a}}(\chi_{[-a,a]}^\vee)^\wedge(\xi) = \frac{e^{-2\pi i(k/2a)\xi}}{\sqrt{2a}}\chi_{[-a,a]}(\xi), \tag{8.45.2}$$

where for the last equality we used $\chi_{[-a,a]} \in L^2$ and that the Fourier transform is a unitary isomorphism on L^2. In particular, Equation (8.45.2) shows both that $\zeta_k \in L^2$ (since its Fourier transform is) and that $\widehat{\zeta}_k(\xi) = 0$ whenever $|\xi| > a$, so $\zeta_k \in \mathscr{H}_a$.

$\{\zeta_k\}_{k\in\mathbb{Z}}$ is an orthonormal set in $\mathscr{H}_a$: Since the Fourier transform is a unitary operator $L^2 \to L^2$, we have for all $k \in \mathbb{Z}$ that

$$\langle\zeta_k|\zeta_k\rangle = \langle\zeta_k^\wedge|\zeta_k^\wedge\rangle = \frac{1}{2a}\int_{-a}^{a} e^{2\pi i(k-k)\xi}\,\mathrm{d}\xi = \frac{1}{2a}\int_{-a}^{a} 1\,\mathrm{d}\xi = 1,$$

and if $\ell \in \mathbb{Z} \setminus \{k\}$,

$$\begin{aligned}\langle\zeta_k|\zeta_\ell\rangle = \langle\zeta_k^\wedge|\zeta_\ell^\wedge\rangle &\overset{(8.45.2)}{=} \frac{1}{2a}\int e^{-2\pi i(k/2a)\xi}\chi_{[-a,a]}(\xi)\overline{e^{-2\pi i\xi(\ell/2a)}\chi_{[-a,a]}(\xi)}\,\mathrm{d}\xi \\ &= \frac{1}{2a}\int_{-a}^{a} e^{2\pi i\xi\left(\frac{k-\ell}{2a}\right)}\,\mathrm{d}\xi = \frac{1}{2a}\left(\frac{2a}{2\pi i(k-\ell)}(e^{\pi i(k-\ell)} - e^{\pi i(\ell-k)})\right) = \frac{\sin(\pi(k-\ell))}{2\pi i(k-\ell)} = 0.\end{aligned}$$

Thus $\{\zeta_k\}_{k\in\mathbb{Z}}$ is an orthonormal set in $\mathscr{H}_a$.

$\{\zeta_k\}_{k\in\mathbb{Z}}$ is a basis of $\mathscr{H}_a$: Suppose $f \in \mathscr{H}_a$ satisfies $\langle f|\zeta_k\rangle = 0$ (and hence also $\langle f^\wedge|\zeta_k^\wedge\rangle = 0$) for all $k \in \mathbb{Z}$. Then for each $k \in \mathbb{Z}$,

$$\begin{aligned}0 &= \int f^\wedge(\xi)\overline{\zeta_k^\wedge(\xi)}\,\mathrm{d}\xi \overset{(8.45.2)}{=} \frac{1}{\sqrt{2a}}\int_{-a}^{a} f^\wedge(\xi)e^{2\pi i(k/2a)\xi}\,\mathrm{d}\xi \\ &= \frac{1}{\sqrt{2a}}\int_{-1/2}^{1/2} f^\wedge(\eta/2a)e^{2\pi ik\eta}\,\mathrm{d}\eta = \sqrt{2a}\int_{\mathbb{T}} f^\wedge(-\eta/2a)\overline{E_k}(\eta)\,\mathrm{d}\eta = \sqrt{2a}\langle\widehat{f}\circ s|E_k\rangle,\end{aligned}$$

where $s\colon \eta \mapsto -\eta/2a$, and $E_k(\eta) = e^{2\pi ik\eta}$. In particular $\langle\widehat{f}\chi_{[-a,a]}|E_k\rangle = 0$ for all $k \in \mathbb{Z}$. But by Folland Theorem 8.20 $\{E_k\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{T})$, so $\widehat{f}\chi_{[-a,a]} = 0$ a.e. Therefore, since $\widehat{f} \in L^2(\mathbb{T})$, by the Fourier inversion theorem (namely since the Fourier transform is an isomorphism $L^2 \to L^2$), $\widehat{f}\circ s = 0$ a.e. on $[-1/2, 1/2]$. Thus $\widehat{f}\chi_{[-a,a]} = 0$ a.e., and hence $\widehat{f} = 0$ for a.e. $\xi \in \mathbb{R}$ (since already $\widehat{f}(\xi) = 0$ for all $\xi > a$). It follows that $\{\zeta_k\}_{k\in\mathbb{Z}}$ is a basis of $\mathscr{H}_a$.

(c) Fix $f \in \mathscr{H}_a$. By part (b) $\{\zeta_k\}_{k\in\mathbb{Z}}$ is an orthonormal basis of $\mathscr{H}_a$, so

$$f = \sum\nolimits_{k\in\mathbb{Z}}\langle f|\zeta_k\rangle\zeta_k = \sum\nolimits_{k\in\mathbb{Z}}\langle\widehat{f}|\widehat{\zeta}_k\rangle\zeta_k,$$

where the series converge in L^2. Thus it is is enough to show $\langle f|\zeta_k\rangle = \frac{1}{\sqrt{2a}}f(k/2a)$ for $k \in \mathbb{Z}$ and that $\sum_{k\in\mathbb{Z}} f(k/2a)\zeta_k$ converges to f uniformly. We have

$$\begin{aligned}\langle f^\wedge|\zeta_k^\wedge\rangle &= \frac{1}{\sqrt{2a}}\int f^\wedge(x)e^{2\pi i(k/2a)x}\chi_{[-a,a]}(x)\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2a}}\int_{-a}^{a} f^\wedge(x)e^{2\pi i(k/2a)x}\,\mathrm{d}x = \frac{1}{\sqrt{2a}}\widehat{\widehat{f}}(-k/2a) = \frac{1}{\sqrt{2a}}f(k/2a),\end{aligned}$$

so it only remains to show the series converges uniformly, and for this it is enough to show the sequence $\{\sum_{k=-N}^{N} f(k/2a)\zeta_k\}_{N\in\mathbb{Z}}$ is uniformly Cauchy.

Fix $\varepsilon > 0$. By Parseval's identity $\sum_{k\in\mathbb{Z}}|\langle f|\zeta_k\rangle| = \|f\|_2^2 < \infty$, so for all sufficiently large $N \in \mathbb{Z}_{\geqslant 0}$

$$\sum\nolimits_{k\in\mathbb{Z}}|\langle f|\zeta_k\rangle|^2 < \varepsilon. \tag{8.45.3}$$

Now fix $x \in \mathbb{R}$ and $M, N \in \mathbb{Z}$ with $M \leqslant N$. For all sufficiently large $M, N \in \mathbb{Z}$, we have

$$\begin{aligned}\left|\sum\nolimits_{k=M}^{N} f(k/2a)\operatorname{sinc}(2ax-k)\right| &= \left|\sum\nolimits_{k=M}^{N}\langle f|\zeta_k\rangle\zeta_k(x)\right| = \sum\nolimits_{k=M}^{N}|\langle f|\zeta_k\rangle||\zeta_k(x)|\\ &\leqslant \left(\sum\nolimits_{k=M}^{N}|\langle f|\zeta_k\rangle|^2\right)^{1/2}\left(\sum\nolimits_{k=M}^{N}|\zeta_k(x)|^2\right)^{1/2} \overset{(8.45.3)}{<} \varepsilon^{1/2}\left(\sum\nolimits_{k=M}^{N}|\zeta_k(x)|^2\right)^{1/2}\end{aligned}$$

where we used the Cauchy–Schwarz inequality. Since $\chi_{[-a,a]}$ is a factor of ζ_k, we may assume $x \in [-a, a]$, and hence that $0 \leqslant |x| \leqslant a$. But we only know this (the previous sentence) for ζ_k, not $\widehat{\zeta_k}$! This requires a correction before the rest of the argument to work. It thus only remains to show the remaining sum term on the right-hand side is uniformly bounded for all $x \in [-a, a]$ as $M, N \to \infty$. For all sufficiently large $M, N \in \mathbb{Z}_{\geqslant 0}$ sufficiently large and $k \in \{M+1, \ldots, N\}$, we have

$$|2ax-k|^2 = |k-2ax|^2 \geqslant ||k|^2 - 2a|x|| \geqslant \frac{|k|^2}{2} = \frac{k^2}{2},$$

and hence

$$\frac{1}{|2ax-k|^2} \leqslant \frac{2}{k^2},$$

so that

$$\sum\nolimits_{k=M}^{N}|\zeta_k(x)|^2 = \frac{2a}{\pi^2}\sum\nolimits_{k=M}^{N}\frac{|\sin(\pi(2ax-k))|^2}{|2ax-k|^2} \leqslant \frac{2a}{\pi^2}\sum\nolimits_{k=M}^{N}\frac{1}{|2ax-k|^2} \leqslant \frac{4a}{\pi^2}\sum\nolimits_{k=M}^{N}\frac{1}{k^2} < \varepsilon,$$

where the final step is by Folland Exercise 8.13(b). The argument that $\|x \mapsto \sum_{k=M}^{N} f(k/2a)\operatorname{sinc}(2ax-k)\|_u < \varepsilon$ for all sufficiently large $M, N \in \mathbb{Z}$ with $M \leqslant N$ is similar. Thus the series $\sum_{k=-\infty}^{\infty} f(k/2a)\operatorname{sinc}(2ax-k)$ is uniformly Cauchy, and hence converges uniformly.

Lastly, we show $f(x) = \sum_{k\in\mathbb{Z}} f(k/2a)\operatorname{sinc}(2ax-k)$ a.e. and that $f \in C_0$. We already know the partial sums converge to f in L^2, so some subsequence of the partial sums converge to f pointwise a.e., so, after modification of f on a null set f equals the given series. Thus f is the uniform limit of the partial sums—which are themselves continuous since sinc is—so f is continuous. To see f vanishes at infinity, note that if we take the Fourier transformation of Equation (8.45.2) once more, we obtain

$$\begin{aligned}\widehat{\widehat{\zeta_k}}(x) &= \frac{1}{\sqrt{2a}}\int e^{-2\pi i(k/2a)\xi}\chi_{[-a,a]}(\xi)e^{-2\pi ix\xi}\,\mathrm{d}\xi = \frac{1}{\sqrt{2a}}\int \chi_{[-a,a]}(\xi)e^{2\pi i(-x-k/2a)\xi}\,\mathrm{d}\xi\\ &= \frac{1}{\sqrt{2a}}\chi_{[-a,a]}^{\vee}(-x-k/2a) \overset{(8.45.1)}{=} \zeta_k(-x).\end{aligned}$$

But $\{\zeta_k\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $\mathscr{H}_a$, so we have a convergent series in L^2 given by

$$f(-x) = \sum_{k\in\mathbb{Z}} \langle f|\zeta_k\rangle \zeta_k(-x) = \sum_{k\in\mathbb{Z}} \langle f|\zeta_k\rangle \hat{\hat{\zeta}}_k(x) = \mathcal{F}^2\Big(x \mapsto \sum_{k\in\mathbb{Z}} \langle f|\zeta_k\rangle \zeta_k(x)\Big)(x) = \hat{\hat{f}}(x),$$

where the penultimate equality is by the DCT, so in particular $\hat{f} \in L^1$ by the Fourier inversion theorem. Thus $\hat{\hat{f}}(-x) = f(x)$ is the Fourier transform of an L^1 function, so $f \in C_0$ by the Riemann–Lebesgue lemma. ☐

Exercise 8.46: Folland Exercise 8.16.

Let $f_k = \chi_{[-1,1]} * \chi_{[-k,k]}$.

(a) Compute $f_k(x)$ explicitly and show that $\|f\|_u = 2$.

(b) $f_k^\vee(x) = (\pi x)^{-2} \sin 2\pi k x \sin 2\pi x$, and $\|f_k^\vee\|_1 \to \infty$ as $k \to \infty$. (Use Folland Exercise 8.15(a), and substitute $y = 2\pi k x$ in the integral defining $\|f_k^\vee\|_1$.)

(c) $\mathcal{F}(L^1)$ is a proper subset of C_0. (Consider $g_k = f_k^\vee$ and use the open mapping theorem.)

Solution.

(a) Let $[a, b], [c, d] \subset \mathbb{R}$. Then

$$\chi_{[c,d]}(x - y) = \delta_{c \leqslant x-y \leqslant d} = \delta_{x-d \leqslant y \leqslant x-c} = \chi_{[x-d,x-c]}(y),$$

so

$$\begin{aligned}\chi_{[a,b]} * \chi_{[c,d]}(x) &= \int \chi_{[a,b]}(y)\chi_{[c,d]}(x-y)\,\mathrm{d}y = \chi_{[a,b]}(y)\chi_{[x-d,x-c]}(y)\,\mathrm{d}y \\ &= \int \chi_{[a,b]\cap[x-d,x-c]}(y)\,\mathrm{d}y = m([a,b] \cap [x-d, x-c]).\end{aligned}$$

Thus

$$\|f\|_u = \sup_{x\in\mathbb{R}} |m([-1,1] \cap [x-k, x+k])| \leqslant m([-1,1]) = 2.$$

(b) By Folland Exercise 8.15(a),

$$\begin{aligned}f_k^\vee(x) &= (\chi_{[-1,1]} * \chi_{[-k,k]})^\vee(x) = \chi_{[-1,1]}^\vee(x)\chi_{[-k,k]}^\vee(x) \\ &= 2\operatorname{sinc}(2x)2k\operatorname{sinc}(2kx) = (\pi x)^{-2}\sin(2\pi x)\sin(2\pi k x)\end{aligned}$$

and, making the substitution $y \mapsto 2k\pi x$, we obtain

$$\begin{aligned}\int |f_k^\vee(x)|\,\mathrm{d}x &= \pi^{-2} \int \left|\frac{1}{x^2}\sin(2\pi x)\sin(2\pi k x)\right| \mathrm{d}x \\ &= 4|k|^2 \int \left|\frac{1}{y^2}\sin(y)\sin(y/k)\right| \mathrm{d}y = 4|k| \lim_{N\to\infty} \int_{-\infty}^{\infty} \left|\frac{\sin y}{y}\right| \left|\frac{\sin(y/k)}{y/k}\right| \mathrm{d}y.\end{aligned}$$

For all $N \in \mathbb{Z}_{\geqslant 0}$, $\left|\frac{\sin y}{y}\right|\left|\frac{\sin(y/k)}{y/k}\right|\chi_{[N,N]} \leqslant \chi_{[N,N]} \in L^1$, so by the DCT we have

$$\lim_{k\to\infty} \int_{-N}^{N} \left|\frac{\sin y}{y}\right| \left|\frac{\sin(y/k)}{y/k}\right| \mathrm{d}y = \int_{-N}^{N} \lim_{k\to\infty} \left|\frac{\sin y}{y}\right| \left|\frac{\sin(y/k)}{y/k}\right| \mathrm{d}y = \int_{-N}^{N} \left|\frac{\sin y}{y}\right| \mathrm{d}y.$$

Hence

$$\|f_k^\vee\|_1 \geqslant \int_{-N}^{N} \left|\frac{\sin y}{y}\right| \mathrm{d}y \tag{8.46.1}$$

for all $N \in \mathbb{Z}_{\geqslant 0}$. But the right-hand side diverges to ∞ as $N \to \infty$, which we now show (or, alternatively, by Folland Exercise 2.59). Note that $|\sin x| \geqslant 1/2$ for all $x \in \mathbb{R}$ such that $|x| \in [\pi/6, 5\pi/6], [7\pi/6, 11\pi/6], [13\pi/6, 17\pi/6], \ldots$. On these respective intervals, we have $|1/x| \geqslant 6/5\pi, 6/11\pi, 6/17\pi, \ldots$, and thus $\left|\frac{\sin x}{x}\right| \geqslant 3/5\pi, 3/11\pi, 3/17\pi, \ldots$. Therefore, for all $k \in \mathbb{Z}_{\geqslant 0}$, by taking the limit of Equation (8.46.1) as $N \to \infty$, we obtain

$$\int_{-\infty}^{\infty} \left|\frac{\sin y}{y}\right| \mathrm{d}y \geqslant 3\left(\frac{1}{5\pi} + \frac{1}{11\pi} + \frac{1}{17\pi} + \cdots\right) = \frac{3}{\pi}\sum_{N=1}^{\infty} \frac{1}{6N-1} = \infty.$$

(c) Any $\widehat{f} \in \mathfrak{F}(L^1)$ is continuous since $\mathfrak{F}$ maps L^1 into C_0. Now suppose for a contradiction $\mathfrak{F}(L^1) = C_0$. By the Hausdorff–Young inequality, $\mathfrak{F}$ is bounded as a map $L^1 \to C_0$ (since $\widehat{f} \in C_0 \subset C_b$, hence $\|\widehat{f}\|_u = \|\widehat{f}\|_\infty \leqslant \|f\|_1$ for all $f \in L^1$). Thus $\mathfrak{F}$ is a bounded surjection, so by the open mapping theorem $\mathfrak{F}$ is invertible on L^1 and $\mathfrak{F}^{-1}\colon C_0 \to L^1$ is bounded. Then there exists $C > 0$ such that for all $k \in \mathbb{Z}_{\geqslant 0}$,

$$\|\widehat{f}_k\|_1 \leqslant C\|f_k\|_u \overset{(a)}{=} 2C,$$

contradicting part (b) since $\|\widehat{f}_k\|_1 \to \infty$ as $k \to \infty$. □

Exercise 8.47: Folland Exercise 8.17.

Given $a > 0$, let $f(x) = e^{-2\pi x}x^{a-1}$ for $x > 0$ and $f(x) = 0$ for $x \leqslant 0$.

(a) $f \in L^1$, and $f \in L^2$ if $a > \frac{1}{2}$.

(b) $\widehat{f}(\xi) = \Gamma(a)[(2\pi)(1+i\xi)]^{-a}$. (Here we are using the branch of z^a in the right half plane that is positive when z is positive. Cauchy's theorem may be used to justify the complex substitution $y = (1+i\xi)x$ in the integral defining $\widehat{f}$.)

(c) If $a, b > \frac{1}{2}$ then

$$\int_{-\infty}^{\infty} (1-ix)^{-a}(1+ix)^{-b}dx = \frac{2^{2-a-b}\pi\Gamma(a+b-1)}{\Gamma(a)\Gamma(b)}.$$

Exercise 8.48: Folland Exercise 8.18.

Suppose $f \in L^2(\mathbb{R})$.

(a) The L^2 derivative f' (see Exercises 24 and 25) exists if and only if $\xi\widehat{f} \in L^2$, in which case $\widehat{f'}(\xi) = 2\pi i\xi\widehat{f}(\xi)$.

(b) If the L^2 derivative f' exists, then

$$\left[\int |f(x)|^2 dx\right] \leqslant 4\int |xf(x)|^2 dx \int |f'(x)|^2 dx$$

(If the integrals on the right are finite, one can integrate by parts to obtain

$\int |f|^2 = -2\,\mathrm{Re}\int x\overline{f}f'.$)

(c) (Heisenberg's Inequality) For any $b, \beta \in \mathbb{R}$,

$$\int (x-b)^2|f(x)|^2 dx \int (\xi-\beta)^2|\widehat{f}(\xi)|^2 d\xi \geqslant \frac{\|f\|_2^4}{16\pi^2}$$

(The inequality is trivial if either integral on the right is infinite; if not, reduce to the case $b = \beta = 0$ by considering $g(x) = e^{-2\pi i\beta x}f(x+b)$.) This inequality, a form of the quantum uncertainty principle, says that f and $\widehat{f}$ cannot both be sharply localized about single points b and β.

Exercise 8.49: Folland Exercise 8.19.

(A variation on the theme of Folland Exercise 8.18) If $f \in L^2(\mathbb{R}^n)$ and the set $S = \{x \mid f(x) \neq 0\}$ has finite measure, then for any measurable

$$E \subset \mathbb{R}^n, \int_E |\widehat{f}|^2 \leqslant \|f\|_2^2 m(S)m(E).$$

Solution. Given that the measure of the set S is finite ($m(S) < \infty$), it follows that $L^p(S) \subset L^q(S)$ for $1 \leqslant q \leqslant p$. Thus, since $f \in L^2(S)$, we have $f \in L^1(S)$. And for any fixed $\xi \in \mathbb{R}^n$, we have

$$\int_S |e^{2\pi i x\cdot\xi}|^2 \,\mathrm{d}x = \int_S 1 \,\mathrm{d}x = m(S) < \infty,$$

so the map $x \mapsto e^{2\pi i x\cdot\xi}$ is also in $L^2(S)$. Now by Hölder's inequality

$$|\widehat{f}(\xi)| = \left|\int f(x)e^{-2\pi i\xi\cdot x}\,\mathrm{d}x\right| = \left|\int \chi_S(x)f(x)e^{-2\pi i\xi\cdot x}\,\mathrm{d}x\right| \leqslant \|f\|_2\|\chi_S\|_2 = \|f\|_2 m(S)^{1/2}, \tag{8.49.1}$$

where the second equality is because $f|_{\mathbb{R}^n\setminus S} = 0$ (by definition of S). Thus

$$\|\widehat{f}\chi_E\|_2^2 = \int_E |\widehat{f}(\xi)|^2 d\xi \overset{(8.49.1)}{\leqslant} \|f\|_2^2 m(S)\int_E 1\,\mathrm{d}\xi = \|f\|_2^2 m(S)m(E). \qquad \square$$

Exercise 8.50: Folland Exercise 8.20.

If $f \in L^1(\mathbb{R}^{n+m})$, define $Pf(x) = \int f(x,y)dy$. (Here $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$.) Then $Pf \in L^1(\mathbb{R}^n)$, $\|Pf\|_1 \leqslant \|f\|_1$, and $(Pf)^\wedge(\xi) = \widehat{f}(\xi, 0)$.

Exercise 8.51: Folland Exercise 8.21.

State and prove a result that encompasses both Theorem 40 and Folland Exercise 8.20, in the setting of Fourier transforms on closed subgroups and quotient groups of $\mathbb{R}^n$.

Exercise 8.52: Folland Exercise 8.22.

Since $\mathcal{F}$ commutes with rotations, the Fourier transform of a radial function is radial; that is, if $F \in L^1(\mathbb{R}^n)$ and $F(x) = f(|x|)$, then $\widehat{F}(\xi) = g(|\xi|)$, where f and g are related as follows.

(a) Let $J(\xi) = \int_S e^{ix\xi} d\sigma(x)$ where σ is surface measure on the unit sphere S in $\mathbb{R}^n$ (Theorem 99). Then J is radial—say, $J(\xi) = j(|\xi|)$—and $g(\rho) = \int_0^\infty j(2\pi r\rho) f(r) r^{n-1} dr$.

(b) J satisfies $\sum_1^n \partial_k^2 J + J = 0$.

(c) j satisfies $\rho j''(\rho) + (n-1)j'(\rho) + \rho j(\rho) = 0$. (This equation is a variant of Bessel's equation. The function j is completely determined by the fact that it is a solution of this equation, is smooth at $\rho = 0$, and satisfies $j(0) = \sigma(S) = 2\pi^{n/2}/\Gamma(n/2)$. In fact, $j(\rho) = (2\pi)^{n/2}\rho^{(2-n)/2} J_{(n-2)/2}(\rho)$ where J_α is the Bessel function of the first kind of order α.)

(d) If $n = 3, j(\rho) = 4\pi\rho^{-1}\sin\rho$. (Set $f(\rho) = \rho j(\rho)$ and use (c) to show that $f'' + f = 0$. Alternatively, use spherical coordinates to compute the integral defining $J(0, 0, \rho)$ directly.)

Exercise 8.53: Folland Exercise 8.23.

In this exercise we develop the theory of Hermite functions.

(a) Define operators T, T^* on $\mathcal{S}(\mathbb{R})$ by $Tf(x) = 2^{-1/2}[xf(x) - f'(x)]$ and $T^*f(x) = 2^{-1/2}[xf(x) + f'(x)]$. Then $\int (Tf)\overline{g} = \int f\left(\overline{T^* g}\right)$ and $T^*T^k - T^kT^* = kT^{k-1}$.

(b) Let $h_0(x) = \pi^{-1/4}e^{-x^2/2}$, and for $k \geqslant 1$ let $h_k = (k!)^{-1/2}T^k h_0$. (h_k is the kth normalized Hermite function.) We have $Th_k = \sqrt{k+1}h_{k+1}$ and $T^*h_k = \sqrt{k}h_{k-1}$, and hence $TT^*h_k = kh_k$.

(c) Let $S = 2TT^* + I$. Then $Sf(x) = x^2 f(x) - f''(x)$ and $Sh_k = (2k+1)h_k$. (S is called the Hermite operator.)

(d) $\{h_k\}_0^\infty$ is an orthonormal set in $L^2(\mathbb{R})$. (Check directly that $\|h_0\|_2 = 1$, then observe that for $k > 0, \int h_k \overline{h}_m = k^{-1}\int (TT^*h_k)\overline{h}_m$ and use (a) and (b).)

(e) We have

$$T^k f(x) = (-1)^k 2^{-k/2} e^{x^2/2}\left(\frac{d}{dx}\right)^k [e^{-x^2/2} f(x)]$$

(use induction on k), and in particular,

$$h_k(x) = \frac{(-1)^k}{[\pi^{1/2}2^k k!]^{1/2}} e^{x^2/2}\left(\frac{d}{dx}\right)^k e^{-x^2}$$

(f) Let $H_k(x) = e^{x^2/2}h_k(x)$. Then H_k is a polynomial of degree k, called the kth normalized Hermite polynomial. The linear span of $H_0, \ldots, H_m$ is the set of all polynomials of degree $\leqslant m$. (The kth Hermite polynomial as usually defined is $[\pi^{1/2}2^k k!]^{1/2} H_k$.)

(g) $\{h_k\}_0^\infty$ is an orthonormal basis for $L^2(\mathbb{R})$. (Suppose $f \perp h_k$ for all k, and let $g(x) = f(x)e^{-x^2/2}$. Show that $\widehat{g} = 0$ by expanding $e^{-2\pi i\xi\cdot x}$ in its Maclaurin series and using (f).)

(h) Define $A\colon L^2 \to L^2$ by $Af(x) = (2\pi)^{1/4}f(x\sqrt{2\pi})$, and define $\widetilde{f} = A^{-1}\mathfrak{F}Af$ for $f \in L^2$. Then A is unitary and $\widetilde{f}(\xi) = (2\pi)^{-1/2}\int f(x)e^{-i\xi x}dx$. Moreover, $\widetilde{Tf} = -iT(\widetilde{f})$ for $f \in \mathfrak{F}$, and $\widetilde{h}_0 = h_0$; hence $\widetilde{h}_k = (-i)^k h_k$. Therefore, if $\phi_k = Ah_k$, $\{\phi_k\}_0^\infty$ is an orthonormal basis for L^2 consisting of eigenfunctions for $\mathfrak{F}$; namely, $\widehat{\phi}_k = (-i)^k\phi_k$.

Q5.

Suppose that $f \in L^1(\mathbb{R})$ and both f and $\widehat{f}$ have compact support. Prove that $f = 0$.

Solution. Since we can translate and compose with scalar multiplication, we may assume without loss of generality $\operatorname{supp} f \subset [0, 1/2]$. Since $f \in L^1$, By the Hausdorff–Young theorem $\widehat{f} \in L^\infty$ and $\|\widehat{f}\|_\infty \leqslant \|f\|_1$. Hence $\widehat{f}$ is a.e. bounded, and in particular

$$\|\widehat{f}\|_1 = \int |\widehat{f}| \leqslant \int \|f\|_1 \chi_{\operatorname{supp}(\widehat{f})} < \infty.$$

Thus $\widehat{f} \in L^1$, so by the Fourier inversion theorem f is a.e. continuous and $\widehat{f}^\wedge = (f^\vee)^\wedge = f$.

Since $\operatorname{supp}\widehat{f}$ is bounded, there exists $N \in \mathbb{Z}_{\geqslant 0}$ such that $\widehat{f}(\kappa) = 0$ whenever $|\kappa| \geqslant N$. In particular, the Fourier series of f is $\sum_{m=-N}^N \widehat{f}(m)e^{2\pi imx}$. By a corollary to the Fourier inversion theorem (namely Folland Corollary 8.27), to see $f = \sum_{m=-N}^N \widehat{f}(m)e^{2\pi imx}$ a.e. it suffices to show for $\kappa \in \mathbb{Z}$ that

$$\mathfrak{F}\Big(x \mapsto \sum_{m=-N}^N \widehat{f}(m)e^{2\pi imx}\Big)(\kappa) = \widehat{f}(\kappa).$$

And indeed,

$$\begin{aligned}\mathfrak{F}\Big(x \mapsto \sum_{m=-N}^N \widehat{f}(m)e^{2\pi imx}\Big)(\kappa) &= \int_0^1 \Big(\sum_{m=-N}^N \widehat{f}(m)e^{2\pi imx}\Big)e^{-2\pi i\kappa x}\,dx \\ &= \sum_{m=-N}^N \widehat{f}(m)\int_0^1 e^{2\pi i(m-\kappa)x}\,dx = \sum_{m=-N}^N \widehat{f}(m)\delta_{m,\kappa} = \widehat{f}(\kappa),\end{aligned}$$

so $f = \sum_{m=-N}^N \widehat{f}(m)e^{2\pi imx}$ a.e. But f vanishes on the interval $(1/2, 1)$, so the sum $\sum_{m=-N}^N \widehat{f}(m)e^{2\pi imx} = 0$ must also; but any trigonometric polynomial that vanishes on an interval must be identically zero (e.g., by the identity principle, since trigonometric polynomials are holomorphic), so $f = 0$. □

Q6.

Show that the conditions $\frac{1}{p} + \frac{1}{q} = 1$ and $1 \leqslant p \leqslant 2$ in the Hausdorff–Young inequality (Folland Theorem 8.30) are both necessary for such an inequality to hold. [a]

[a] *Hint: For the second condition, consider the functions* $f_s(x) = s^{-\frac{n}{2}} e^{-\pi|x|^2/s}$ *for* $s = 1 + it$ *and* $t > 0$.

Solution. Suppose $p, q \in [1, \infty]$ satisfy

$$\|\widehat{f}\|_q \leqslant \|f\|_p \text{ for all } f \in L^p(\mathbb{R}^n). \tag{8.53.1}$$

- *Necessity that the exponents are conjugate*: Suppose $p, q \in [1, \infty]$, and consider an arbitrary $f \in L^p(\mathbb{R}^n)$. For $t > 0$, define $f_t(x) = t^{-n} f(t^{-1}x)$.

$$\|f_t\|_p = \left(\int t^{-np}|f(t^{-1}x)|^p \,\mathrm{d}x\right)^{1/p} = t^{-n}\left(\int t^n |f(x)|^p \,\mathrm{d}x\right)^{1/p} = t^{-n(1-1/p)}\|f\|_p, \tag{8.53.2}$$

and this equation still holds if $p = \infty$ with the convention $1/p = 0$. Now in particular we know $f_t \in L^p$. Now write

$$\widehat{f}_t(\xi) = t^{-n}\int f(t^{-1}x)e^{-2\pi i\xi\cdot x}\,\mathrm{d}t = \int f(y)e^{-2\pi i\xi\cdot(y/t)}\,\mathrm{d}y = \widehat{f}(t\xi).$$

Then

$$\|\widehat{f}_t\|_q = \left(\int |\widehat{f}(t\xi)|^q \,\mathrm{d}\xi\right)^{1/q} = t^{-n/q}\left(\int |\widehat{f}(\xi)|^q \,\mathrm{d}\xi\right)^{1/q} = t^{-n/q}\|\widehat{f}\|_q \tag{8.53.3}$$

so

$$\|\widehat{f}\|_q \overset{(8.53.3)}{=} t^{n/q}\|\widehat{f}_t\|_q \overset{(8.53.1)}{\leqslant} t^{n/q}\|f_t\|_p \overset{(8.53.2)}{=} t^{n/q}t^{-n(1-1/p)}\|f\|_p = t^{n\left(\frac{1}{p}+\frac{1}{q}-1\right)}\|f\|_p,$$

where we use the condition that $1/q = 0$ for $q = \infty$. But $t > 0$ was arbitrary, so this must hold for all such t; thus $1/p + 1/q - 1 = 0$, so p and q are conjugate exponents. Thus the conjugate exponent condition in the Hausdorff–Young inequality is necessary for $p, q \in [1, \infty]$.

- *Necessity that* $p \in [1, 2]$: Suppose for a contradiction $p \in (2, \infty]$ and again consider an arbitrary $f \in L^1(\mathbb{R}^n)$. First note $p \neq \infty$, since otherwise by Folland Exercise 8.15 the $L^1(\mathbb{R})$ function $\chi_{[-\frac{1}{2},\frac{1}{2}]}$ satisfies

$$\infty = \int_{-\infty}^{\infty}\left|\frac{\sin(\xi)}{\xi}\right|\mathrm{d}\xi = \|\widehat{\chi}_{[-\frac{1}{2},\frac{1}{2}]}\|_1 \overset{(8.53.1)}{\leqslant} \|\chi_{[-\frac{1}{2},\frac{1}{2}]}\|_\infty = 1,$$

a contradiction (and the case of general $n \in \mathbb{Z}_{\geqslant 1}$ is similar by considering $\chi_{[-1/2,1/2]^n}$), so we may assume $p \in (2, \infty)$.

Let $f_s(x) = s^{-n/2}e^{-\pi|x|^2/s}$ and let $h(x) = e^{-\pi s|x|^2}$, so that $f_s = \widehat{h}$ by Folland Proposition 8.24. By our assumption (8.53.1) and the previous point, $1/p + 1/q = 1$. Then $q \in (1, 2)$, and in particular $q < p$. We have

$$\|h\|_p = \left(\int |e^{-\pi s|x|^2}|^p \,\mathrm{d}x\right)^{1/p} = \left(\int e^{-\pi p|x|^2}\,\mathrm{d}x\right)^{1/p} \overset{\text{Folland Prop. 2.53}}{=} p^{-n/2p} \tag{8.53.4}$$

and

$$\|\widehat{h}\|_q = \|f_s\|_q = \left(\int |s^{-n/2}e^{-\pi|x|^2/s}|^q\right)^{1/q} = |s|^{-n/2}\left(\int e^{-\pi q(1+t^2)^{-1}|x|^2}\,\mathrm{d}x\right)^{1/q}$$

$$\overset{\text{Folland Prop. 2.53}}{=} |s|^{-n/2}\left(\frac{\pi}{\pi q(1+t^2)^{-1}}\right)^{n/2q} = (1+t^2)^{-n/4}q^{-n/2q}(1+t^2)^{n/2q}$$

$$= q^{-n/2q}(1+t^2)^{\frac{n}{4}\left(\frac{2}{q}-1\right)} = q^{-n/2q}(1+t^2)^{\frac{n}{4}\left(\frac{1}{q}-\frac{1}{p}\right)}, \tag{8.53.5}$$

where for the last equality we used the requirement from the previous point that $1/p + 1/q = 1$. In particular $h \in L^p(\mathbb{R}^n)$, so by our assumption (8.53.1)

$$p^{-n/2p} \overset{(8.53.4)}{=} \|h\|_p \geqslant \|\widehat{h}\|_q \overset{(8.53.5)}{=} q^{-n/2q}(1+t^2)^{\frac{n}{4}\left(\frac{1}{q}-\frac{1}{p}\right)}.$$

Raising both sides to the power of $-2/n$, we obtain

$$p^{1/p} \leqslant q^{1/q}(1+t^2)^{-\frac{1}{2}\left(\frac{1}{q}-\frac{1}{p}\right)}. \tag{8.53.6}$$

Since $p < q$ by assumption, $-1/2(1/q - 1/p) < 0$, so by choosing $t > 0$ appropriately we can make $(1+t)^{-\frac{1}{2}\left(\frac{1}{q}-\frac{1}{p}\right)}$ arbitrarily small. But $p^{1/p}$ is strictly positive, so this contradicts Equation (8.53.6). Thus $p \notin (2, \infty]$, so $p \in [1, 2]$. □

8.4 Summation of Fourier Integrals and Series

The Fourier inversion theorem shows how to express a function f on $\mathbb{R}^n$ in terms of $\widehat{f}$ provided that f and $\widehat{f}$ are in L^1. The same result holds for periodic functions. Namely, if $f \in L^1(\mathbb{R}^n)$ and $\widehat{f} \in \ell^1(\mathbb{R}^n)$, then the Fourier series $\sum_\kappa \widehat{f}(\kappa)e^{2\pi i\kappa\cdot x}$ converges absolutely and uniformly to a function g. Since $\ell^1 \subset \ell^2$, it follows from Theorem 29 that $f \in L^2$ and that the series converges to f in the L^2 norm. Hence $f = g$ a.e., and $f = g$ everywhere if f is assumed continuous at the outset.

Two questions therefore arise. What conditions on f will guarantee that $\widehat{f}$ is integrable? And how can f be recovered from $\widehat{f}$ if $\widehat{f}$ is not integrable?

As for the first question, since $\widehat{f}$ is bounded for $f \in L^1$, the issue is the decay of $\widehat{f}$ at infinity, and this is related to the smoothness properties of f. For example, by Theorem 31e, if $f \in C^{n+1}(\mathbb{R}^n)$ and $\partial^\alpha f \in L^1 \cap C_0$ for $|\alpha| \leqslant n+1$, then $|\widehat{f}(\xi)| \leqslant C(1+|\xi|)^{-n-1}$ and hence $\widehat{f} \in L^1(\mathbb{R}^n)$ by Corollary 101. The same result holds for periodic functions, for the same reason: If $f \in C^{n+1}(\mathbb{R}^n)$, then $|\widehat{f}(\kappa)| \leqslant C(1+|\kappa|)^{-n-1}$ and hence $\widehat{f} \in \ell^1(\mathbb{R}^n)$.

To obtain sharper results when $n > 1$ requires a generalized notion of partial derivatives, so we shall postpone this task until §9.3. (See Theorem 43.) However, for $n = 1$ we can easily obtain a better theorem that covers the useful case of functions that are continuous and piecewise C^1. We state it for periodic functions and leave the nonperiodic case to the reader (Folland Exercise 8.24).

Theorem 8.54: 8.33.

Suppose that f is periodic and absolutely continuous on $\mathbb{R}$, and that $f' \in L^p(\mathbb{R})$ for some $p > 1$. Then $\widehat{f} \in \ell^1(\mathbb{R})$.

Proof. Since $p > 1$, we have $C_p = \sum_1^\infty \kappa^{-p} < \infty$; and since $L^p(\mathbb{T}) \subset L^2(\mathbb{T})$ for $p > 2$, we may assume that $p \leqslant 2$. Integration by parts (Theorem 79) shows that $(f')\widehat{(}(\kappa) = 2\pi i\kappa\widehat{f}(\kappa)$. Hence, by the inequalities of Hölder and Hausdorff-Young, if q is the conjugate exponent to p,

$$\sum_{\kappa\neq 0} |\widehat{f}(\kappa)| \leqslant \left[\sum_{\kappa\neq 0}(2\pi|\kappa|)^{-p}\right]^{1/p}\left[\sum_{\kappa\neq 0}(2\pi|\kappa\widehat{f}(\kappa)|)^q\right]^{1/q}$$
$$= \frac{(2C_p)^{1/p}}{2\pi}\|(f')\|_q \leqslant \frac{(2C_p)^{1/p}}{2\pi}\|f'\|_p$$

Adding $|\widehat{f}(0)|$ to both sides, we see that $\|\widehat{f}\|_1 < \infty$.

We now turn to the problem of recovering f from $\widehat{f}$ under minimal hypotheses on f, and we consider first the case of $\mathbb{R}^n$. The proof of the Fourier inversion theorem contains the essential idea: Replace the divergent integral $\int \widehat{f}(\xi)e^{2\pi i\xi\cdot x}d\xi$ by $\int \widehat{f}(\xi)\Phi(t\xi)e^{2\pi i\xi\cdot x}d\xi$ where Φ is a continuous function that vanishes rapidly enough at infinity to make the integral converge. If we choose Φ to satisfy $\Phi(0) = 1$, then $\Phi(t\xi) \to 1$ as $t \to 0$, and with any luck the corresponding integral will converge to f in some sense. One Φ that works is the function $\Phi(\xi) = e^{-\pi|\xi|^2}$ used in the proof of the inversion theorem, but we shall see below that there are others of independent interest. We therefore formulate a fairly general theorem, for which we need the following lemma that complements Theorem 31(c).

Lemma 8.55: 8.34.

If $f, g \in L^2(\mathbb{R}^n)$, then $(\widehat{f}\widehat{g})^\vee = f * g$.

Proof. $\widehat{f}\widehat{g} \in L^1$ by Plancherel's theorem and Hölder's inequality, so $(\widehat{f}\widehat{g})^\vee$ makes sense. Given $x \in \mathbb{R}^n$, let $h(y) = \overline{g(x-y)}$. It is easily verified that $\widehat{h}(\xi) = \overline{\widehat{g}(\xi)e^{-2\pi i\xi\cdot x}}$, so since $\mathcal{F}$ is unitary on L^2,

$$f * g(x) = \int f\overline{h} = \int \widehat{f}\overline{\widehat{h}} = \int \widehat{f}(\xi)\widehat{g}(\xi)e^{2\pi i\xi\cdot x}d\xi = (\widehat{f}\widehat{g})^\vee(x)$$

Theorem 8.56: 8.35.

Suppose that $\Phi \in L^1 \cap C_0$, $\Phi(0) = 1$, and $\phi = \Phi^\vee \in L^1$. Given $f \in L^1 + L^2$, for $t > 0$ set

$$f^t(x) = \int \widehat{f}(\xi)\Phi(t\xi)e^{2\pi i\xi\cdot x}d\xi$$

(a) If $f \in L^p(1 \leqslant p < \infty)$, then $f^t \in L^p$ and $\|f^t - f\|_p \to 0$ as $t \to 0$.

(b) If f is bounded and uniformly continuous, then so is f^t, and $f^t \to f$ uniformly as $t \to 0$.

(c) Suppose also that $|\phi(x)| \leqslant C(1+|x|)^{-n-\epsilon}$ for some $C, \epsilon > 0$. Then $f^t(x) \to f(x)$ for every x in the Lebesgue set of f.

Proof. We have $f = f_1 + f_2$ where $f_1 \in L^1$ and $f_2 \in L^2$. Since $\widehat{f}_1 \in L^\infty$, $\widehat{f}_2 \in L_2$, and $\Phi \in (L^1 \cap C_0) \subset (L^1 \cap L^2)$, the integral defining f^t converges absolutely for every x. Moreover, if $\phi_t(x) = t^{-n}\phi(t^{-1}x)$, we have $\Phi(t\xi) = (\phi_t)^\wedge(\xi)$ by the inversion theorem and Theorem 31b, and $\int \phi(x)dx = \Phi(0) = 1$. Since $\phi, \Phi \in L^1$ we have $f_1 * \phi \in L^1$ and $\widehat{f}_1\Phi \in L^1$, so by Theorem 8.22c and the inversion formula,

$$\int \widehat{f}_1(\xi)\Phi(t\xi)e^{2\pi i\xi\cdot x}d\xi = f_1 * \phi_t(x)$$

Also, $\phi \in L^2$ by the Plancherel theorem, so by Lemma 55,

$$\int \widehat{f}_2(\xi)\Phi(t\xi)e^{2\pi i\xi\cdot x}d\xi = f_2 * \phi_t(\xi)$$

In short, $f^t = f * \phi_t$, so the assertions follow from Theorems 8.14 and 8.15.

By combining this theorem with the Poisson summation formula, we obtain a corresponding result for periodic functions.

Theorem 8.57: 8.36.

Suppose that $\Phi \in C(\mathbb{R}^n)$ satisfies $|\Phi(\xi)| \leqslant C(1+|\xi|)^{-n-\varepsilon}$, $|\Phi^\vee(x)| \leqslant C(1+|x|)^{-n-\varepsilon}$, and $\Phi(0) = 1$. Given $f \in L^1(\mathbb{R}^n)$, for $t > 0$ set

$$f^t(x) = \sum\nolimits_{\kappa\in\mathbb{Z}^n} \widehat{f}(\kappa)\Phi(t\kappa)e^{2\pi i\kappa\cdot x}$$

(which converges absolutely since $\sum_\kappa |\Phi(t\kappa)| < \infty$).

(a) If $f \in L^p(\mathbb{T}^n)(1 \leqslant p < \infty)$, then $\|f^t - f\|_p \to 0$ as $t \to 0$, and if $f \in C(\mathbb{T}^n)$, then $f^t \to f$ uniformly as $t \to 0$.

(b) $f^t(x) \to f(x)$ for every x in the Lebesgue set of f.

Proof. Let $\phi = \Phi^\vee$ and $\phi_t(x) = t^{-n}\phi(t^{-1}x)$. Then $(\phi_t)^\wedge(\xi) = \Phi(t\xi)$, and ϕ_t satisfies the hypotheses of the Poisson summation formula, so

$$\sum\nolimits_{k\in\mathbb{Z}^n} \phi_t(x-k) = \sum\nolimits_{k\in\mathbb{Z}^n} \Phi(t\kappa)e^{2\pi i\kappa\cdot x}$$

Let us denote the common value of these sums by $\psi_t(x)$. Then

$$(f * \psi_t)^\vee(\kappa) = \widehat{f}(\kappa)\widehat{\psi}_t(\kappa) = \widehat{f}(\kappa)\Phi(t\kappa) = (f^t)^\wedge(\kappa).$$

so $f^t = f * \psi_t$. Hence, by Young's inequality and Theorem 40 we have

$$\|f^t\|_p \leqslant \|f\|_p\|\psi_t\|_1 \leqslant \|f\|_p\|\phi_t\|_1 = \|f\|_p\|\phi\|_1$$

so the operators $f \mapsto f^t$ are uniformly bounded on $L^p, 1 \leqslant p \leqslant \infty$.

Now, since Φ is continuous and $\Phi(0) = 1$, we clearly have $f^t \to f$ uniformly (and hence in $L^p(\mathbb{T}^n)$) if f is a trigonometric polynomial—that is, if $\widehat{f}(\kappa) = 0$ for all but finitely many κ. But the trigonometric polynomials are dense in $C(\mathbb{T}^n)$ in the uniform norm by the Stone-Weierstrass theorem, and hence also dense in $L^p(\mathbb{T}^n)$ in the L^p norm for $p < \infty$. Assertion (a) therefore follows from Proposition 89.

To prove (b), suppose that x is in the Lebesgue set of f; by translating f we may assume that $x = 0$, which simplifies the notation. With $Q = \left[-\frac{1}{2}, \frac{1}{2}\right)^n$, we have

$$\begin{aligned} f^t(0) = f * \psi_t(0) &= \int_Q f(x)\psi_t(-x)dx \\ &= \int_Q f(x)\phi_t(-x)dx + \sum\nolimits_{k\neq 0} \int_Q f(x)\phi_t(-x+k)dx \end{aligned}$$

Since

$$|\phi_t(x)| \leqslant Ct^{-n}(1 + t^{-1}|x|)^{-n-\varepsilon} \leqslant Ct^{\varepsilon}|x|^{-n-\varepsilon}$$

for $x \in Q$ and $k \neq 0$ we have $|\phi_t(-x + k)| \leqslant C2^{n+\varepsilon}t^{\varepsilon}|k|^{-n-\varepsilon}$, and hence

$$\sum\nolimits_{k\neq 0} \int_Q |f(x)\phi_t(-x+k)|dx \leqslant \left[C2^{n+\varepsilon}\|f\|_1 \sum\nolimits_{k\neq 0} |k|^{-n-\varepsilon}\right]t^{\varepsilon}$$

which vanishes as $t \to 0$. On the other hand, if we define $g = f\chi_Q \in L^1(\mathbb{R}^n)$, then 0 is in the Lebesgue set of g (because 0 is in the interior of Q, and the condition that 0 be in the Lebesgue set of g depends only on the behavior of g near 0), so by Theorem 18,

$$\lim_{t\to 0} \int_Q f(x)\phi_t(-x)dx = \lim_{t\to 0} g * \phi_t(0) = g(0) = f(0)$$

Let us examine some specific examples of functions Φ that can be used in Theorems 8.35 and 8.36. The first is the one already used in the proof of the inversion theorem,

$$\Phi(\xi) = e^{-\pi|\xi|^2}, \quad \phi(x) = \Phi^\vee(x) = e^{-\pi|x|^2}$$

This ϕ is called the Gauss kernel or Weierstrass kernel. It is important for a number of reasons, including its connection with the heat equation that we shall explain in Folland Section 8.7. When $n = 1$, its periodized version

$$\psi_t(x) = \frac{1}{t}\sum\nolimits_{k\in\mathbb{Z}} e^{-\pi|x-k|^2/t^2} = \sum\nolimits_{\kappa\in\mathbb{Z}} e^{-\pi t^2\kappa^2} e^{2\pi i\kappa\cdot x}$$

in terms of which the f^t in Theorem 57 is given by $f^t = f * \psi_t$, is essentially one of the Jacobi theta functions, which are connected with elliptic functions and have applications in number theory.

The second example is $\Phi(\xi) = e^{-2\pi|\xi|}$, whose inverse Fourier transform ϕ is called the Poisson kernel on $\mathbb{R}^n$. When $n = 1$, we have

$$\begin{aligned} \phi(x) &= \int_{-\infty}^0 e^{2\pi(1+ix)\xi}d\xi + \int_0^\infty e^{2\pi(-1+ix)\xi}d\xi \\ &= \frac{1}{2\pi}\left[\frac{1}{1+ix} + \frac{1}{1-ix}\right] = \frac{1}{\pi(1+x^2)} \end{aligned}$$

The formula for ϕ in higher dimensions is worked out in Folland Exercise 8.26; it turns out that $\phi(x)$ is a constant multiple of $(1+|x|^2)^{-(n+1)/2}$. Like the Gauss kernel, the Poisson kernel has an interpretation in terms of partial differential equations that we shall explain in Folland Section 8.7.

If we take $n=1$ and $\Phi(\xi)=e^{-2\pi|\xi|}$ in Theorem 57, make the substitution $r=e^{-2\pi t}$, and write A_rf in place of f^t, we obtain

$$\begin{aligned} A_rf(x) &= \sum\nolimits_{\kappa\in\mathbb{Z}} r^{|\kappa|}\widehat{f}(\kappa)e^{-2\pi i\kappa x} \\ &= \widehat{f}(0)+\sum\nolimits_{k=1}^{\infty} r^k[\widehat{f}(k)e^{2\pi ikx}+\widehat{f}(-k)e^{-2\pi ikx}] \end{aligned}$$

This formula is a special case of one of the classical methods for summing a (possibly) divergent series. Namely, if $\sum_0^\infty a_k$ is a series of complex numbers, for $0<r<1$ its $\boldsymbol{r}$th Abel mean is the series $\sum_0^\infty r^k a_k$. If the latter series converges for $r<1$ to the sum $S(r)$ and the limit $S=\lim_{r\nearrow 1}S(r)$ exists, the series $\sum_0^\infty a_k$ is said to be Abel summable to S. If $\sum_0^\infty a_k$ converges to the sum S, then it is also Abel summable to S (Folland Exercise 8.27), but the Abel sum may exist even when the series diverges.

In (8.38), $A_rf(x)$ is the rth Abel mean of the Fourier series of f, in which the kth and $(-k)$th terms are grouped together to make a series indexed by the nonnegative integers. It has the following complex-variable interpretation: If we set $z=re^{2\pi ix}$, then

$$A_rf(x)=\sum\nolimits_0^\infty \widehat{f}(k)z^k+\sum\nolimits_1^\infty \widehat{f}(-k)\overline{z}^k$$

The two series on the right define, respectively, a holomorphic and an antiholomorphic function on the unit disc $|z|<1$. In particular, $A_rf(x)$ is a harmonic function on the unit disc, and the fact that $A_rf\to f$ as $r\to 1$ means that f is the boundary value of this function on the unit circle. See also Folland Exercise 8.28.

Our final example is the function $\Phi(\xi)=\max(1-|\xi|,0)$ with $n=1$. Its inverse Fourier transform is

$$\begin{aligned} \phi(x) &= \int_{-1}^{0}(1+\xi)e^{2\pi i\xi\cdot x}d\xi+\int_0^1(1-\xi)e^{2\pi i\xi\cdot x}d\xi \\ &= \frac{e^{2\pi ix}+e^{-2\pi ix}-2}{(2\pi ix)^2}=\left(\frac{\sin\pi x}{\pi x}\right)^2 \end{aligned}$$

If we use this Φ in Theorem 57, take $t=(m+1)^{-1}(m=0,1,2,\ldots)$, and write $\sigma_mf(x)$ for $f^{1/(m+1)}(x)$, we obtain

$$\begin{aligned} \sigma_mf(x) &= \sum\nolimits_{\kappa=-m}^{m}\frac{m+1-|\kappa|}{m+1}\widehat{f}(\kappa)e^{2\pi i\kappa x} \\ &= \widehat{f}(0)+\sum\nolimits_{k=1}^{m}\frac{m+1-k}{m+1}[\widehat{f}(k)e^{2\pi ikx}+\widehat{f}(-k)e^{-2\pi ikx}] \end{aligned}$$

This is an instance of another classical method for summing divergent series. Namely, if $\sum_0^\infty a_k$ is a series of complex numbers, its $\boldsymbol{m}$th Cesr̀o mean is the average of its first $m+1$ partial sums, $(m+1)^{-1}\sum_0^m S_n$, where $S_n=\sum_0^n a_k$. If the sequence of Cesr̀o means converges as $m\to\infty$ to a limit S, the series is said to be Cesr̀o summable to S. It is easily verified that if $\sum_0^\infty a_k$ converges to S, then it is Cesr̀o summable to S (but perhaps not conversely), and that $\sigma_m f(x)$ is the mth Cesr̀o mean of the Fourier series of f with the kth and $(-k)$th terms grouped together. See Folland Exercise 8.29, and also Folland Exercise 8.33 in the next section.

Exercise 8.58: Folland Exercise 8.24.

State and prove an analogue of Theorem 54 for functions on $\mathbb{R}$. (In addition to the hypotheses that f be locally absolutely continuous and that $f' \in L^p$ for some $p > 1$, you will need some further conditions f and/or f' at infinity to make the argument work. Make them as mild as possible.)

Exercise 8.59: Folland Exercise 8.25.

For $0 < \alpha \leqslant 1$, let $\Lambda_\alpha(\mathbb{T})$ be the space of Hölder continuous functions on $\mathbb{T}$ of exponent α as in Folland Exercise 5.11. Suppose $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$.

(a) If f satisfies the hypotheses of Theorem 54, then $f \in \Lambda_{1/q}(\mathbb{T})$, but f need not lie in $\Lambda_\alpha(\mathbb{T})$ for any $\alpha > 1/q$. (Hint: $f(b) - f(a) = \int_a^b f'(t)dt$.)

(b) If $\alpha < 1, \Lambda_\alpha(\mathbb{T})$ contains functions that are not of bounded variation and hence are not absolutely continuous. (But see Folland Exercise 3.37.)

Exercise 8.60: Folland Exercise 8.26.

The aim of this exercise is to show that the inverse Fourier transform of $e^{-2\pi|\xi|}$ on $\mathbb{R}^n$ is

$$\phi(x) = \frac{\Gamma\left(\frac{1}{2}(n+1)\right)}{\pi^{(n+1)/2}}(1+|x|^2)^{-(n+1)/2}$$

(a) If $\beta \geqslant 0, e^{-\beta} = \pi^{-1}\int_{-\infty}^{\infty}(1+t^2)^{-1}e^{-i\beta t}dt$. (Use (8.37).)

(b) If $\beta \geqslant 0, e^{-\beta} = \int_0^\infty (\pi s)^{-1/2}e^{-s}e^{-\beta^2/4s}ds$. (Use (a), Proposition 33, and the formula $(1+t^2)^{-1} = \int_0^\infty e^{-(1+t^2)s}ds$.)

(c) Let $\beta = 2\pi|\xi|$ where $\xi \in \mathbb{R}^n$; then the formula in (b) expresses $e^{-2\pi|\xi|}$ as a superposition of dilated Gauss kernels. Use Proposition 33 again to derive the asserted formula for ϕ.

Exercise 8.61: Folland Exercise 8.27.

Suppose that the numerical series $\sum_0^\infty a_k$ is convergent.

(a) Let $S_m^n = \sum_m^n a_k$. Then $\sum_m^n r^k a_k = \sum_m^{n-1} S_m^j(r^j - r^{j+1}) + S_m^n r^n$ for $0 \leqslant r \leqslant 1$ ("summation by parts").

(b) $|\sum_m^n r^k a_k| \leqslant \sup_{j \geqslant m}|S_m^j|$.

(c) The series $\sum_0^\infty r^k a_k$ is uniformly convergent for $0 \leqslant r \leqslant 1$, and hence its sum $S(r)$ is continuous there. In particular, $\sum_0^\infty a_k = \lim_r > 1S(r)$.

Exercise 8.62: Folland Exercise 8.28.

Suppose that $f \in L^1(\mathbb{T})$, and let $A_r f$ be given by (8.38).

(a) $A_r f = f * P_r$ where $P_r(x) = \sum_{-\infty}^{\infty} r^{|\kappa|} e^{2\pi i \kappa x}$ is the Poisson kernel for $\mathbb{T}$.

(b) $P_r(x) = (1 - r^2)/(1 + r^2 - 2r\cos 2\pi x)$.

Exercise 8.63: Folland Exercise 8.29.

Given $\{a_k\}_0^\infty \subset \mathbb{C}$, let $S_n = \sum_0^n a_k$ and $\sigma_m = (m+1)^{-1} \sum_0^m S_n$.

(a) $\sigma_m = (m+1)^{-1} \sum_0^m (m+1-k) a_k$.

(b) If $\lim_{n\to\infty} S_n = \sum_0^\infty a_k$ exists, then so does $\lim_{m\to\infty} \sigma_m$, and the two limits are equal.

(c) The series $\sum_0^\infty (-1)^k$ diverges but is Abel and Cesr̀o summable to $\frac{1}{2}$.

Exercise 8.64: Folland Exercise 8.30.

If $f \in L^1(\mathbb{R}^n)$, f is continuous at 0, and $\widehat{f} \geqslant 0$, then $\widehat{f} \in L^1$. (Use Theorem 8.35c and Fatou's lemma.)

Exercise 8.65: Folland Exercise 8.31.

Suppose $a > 0$. Use (8.37) to show that

$$\sum_{-\infty}^{\infty} \frac{1}{k^2 + a^2} = \frac{\pi}{a} \frac{1 + e^{-2\pi a}}{1 - e^{-2\pi a}}$$

Then subtract a^{-2} from both sides and let $a \to 0$ to show that $\sum_1^\infty k^{-2} = \pi^2/6$.

Exercise 8.66: Folland Exercise 8.32.

A C^∞ function f on $\mathbb{R}$ is real-analytic if for every $x \in \mathbb{R}$, f is the sum of its Taylor series based at x in some neighborhood of x. If f is periodic and we regard f as a function on $S = \{z \in \mathbb{R} \mid |z| = 1\}$, this condition is equivalent to the condition that f be the restriction to S of a holomorphic function on some neighborhood of S. Show that $f \in C^\infty(\mathbb{R})$ is real-analytic if and only if $|\widehat{f}(\kappa)| \leqslant Ce^{-\varepsilon|\kappa|}$ for some $C, \varepsilon > 0$. (See the discussion of the Abel means $A_r f$ in the text, and note that $\overline{z} = z^{-1}$ when $|z| = 1$.)

8.5 Pointwise Convergence of Fourier Series

The techniques and results of the previous two sections, involving such things as L^p norms and summability methods, are relatively modern; they were preceded historically by the study of pointwise convergence of one-dimensional Fourier series. Although the latter is one of the oldest parts of Fourier analysis, it is also one of the most difficult—unfortunately for the mathematicians who developed it, but fortunately for us who are the beneficiaries of the ideas and techniques they invented in doing so. A thorough study of this issue is beyond the scope of this book, but we would be remiss not to present a few of the classic results.

To set the stage, suppose $f \in L^1(\mathbb{T})$. We denote by $S_m f$ the mth symmetric partial sum of the Fourier series of f:

$$S_m f(x) = \sum_{-m}^{m} \widehat{f}(k) e^{2\pi ikx}$$

From the definition of $\widehat{f}(k)$, we have

$$S_m f(x) = \sum_{-m}^{m} \int_0^1 f(y) e^{2\pi ik(x-y)} dy = f * D_m(x)$$

where D_m is the mth Dirichlet kernel:

$$D_m(x) = \sum_{-m}^{m} e^{2\pi ikx}$$

The terms in this sum form a geometric progression, so

$$D_m(x) = e^{-2\pi imx} \sum_{0}^{2m} e^{2\pi ikx} = e^{-2\pi imx} \frac{e^{2\pi(2m+1)x} - 1}{e^{2\pi ix} - 1}$$

Multiplying top and bottom by $e^{-\pi ix}$ yields the standard closed formula for D_m:

$$D_m(x) = \frac{e^{(2m+1)\pi ix} - e^{-(2m+1)\pi ix}}{e^{\pi ix} - e^{-\pi ix}} = \frac{\sin(2m+1)\pi x}{\sin \pi x} \tag{8.66.1}$$

The difficulty with the partial sums $S_m f$, as opposed to (for example) the Abel or Cesr̀o means, can be summed up in a nutshell as follows. $S_m f$ can be regarded as a special case of the construction in Theorem 57; in fact, with the notation used there, $S_m f = f^{1/m}$ if we take $\Phi = \chi_{[-1,1]}$. But $\chi_{[-1,1]}$ does not satisfy the hypotheses of Theorem 57, because its inverse Fourier transform $(\pi x)^{-1} \sin 2\pi x$ (Folland Exercise 8.15(a)) is not in $L^1(\mathbb{R})$. On the level of periodic functions, this is reflected in the fact that although $D_m \in L^1(\mathbb{R})$ for all m, $\|D_m\|_1 \to \infty$ as $m \to \infty$ (Folland Exercise 8.34).

Among the consequences of this is that the Fourier series of a continuous function f need not converge pointwise, much less uniformly, to f; see Folland Exercise 8.35. (This does not contradict the fact that trigonometric polynomials are dense in $C(\mathbb{T})$! It just means that if one wants to approximate a function $f \in C(\mathbb{T})$ uniformly by trigonometric polynomials, one should not count on the partial sums $S_m f$ to do the job; the Cesr̀o means defined by (8.39) work much better in general.) To obtain positive results for pointwise convergence, one must look in other directions.

The first really general theorem about pointwise convergence of Fourier series was obtained in 1829 by Dirichlet, who showed that $S_m f(x) \to \frac{1}{2}[f(x+) + f(x-)]$ for every x provided that f is piecewise continuous and piecewise monotone. Later refinements of the argument showed that what is really needed is for f to be of bounded variation. We now prove this theorem, for which we need two lemmas. The first one is a slight generalization of one of the more arcane theorems of elementary calculus, the "second mean value theorem for integrals."

Lemma 8.67: 8.41.

Let ϕ and ψ be real-valued functions on $[a, b]$. Suppose that ϕ is monotone and right continuous on $[a, b]$ and ψ is continuous on $[a, b]$. Then there exists $\eta \in [a, b]$ such that

$$\int_a^b \phi(x)\psi(x)dx = \phi(a)\int_a^\eta \psi(x)dx + \phi(b)\int_\eta^b \psi(x)dx$$

Proof. Adding a constant c to ϕ changes both sides of the equation by the amount $c\int_a^b \psi(x)dx$, so we may assume that $\phi(a) = 0$. We may also assume that ϕ is increasing; otherwise replace ϕ by $-\phi$. Let $\Psi(x) = \int_x^b \psi(t)dt$ (so that $\Psi' = -\psi$) and apply Theorem 79:

$$\int_a^b \phi(x)\psi(x)dx = -\phi(x)\Psi(x)|_a^b + \int_{(a,b]} \Psi(x)d\phi(x)$$

The endpoint evaluations vanish since $\phi(a) = \Psi(b) = 0$. Since ϕ is increasing and $\int_{(a,b]} d\phi = \phi(b) - \phi(a) = \phi(b)$, if m and M are the minimum and maximum values of Ψ on $[a, b]$ we have $m\phi(b) \leqslant \int_{(a,b]} \Psi d\phi \leqslant M\phi(b)$. By the intermediate value theorem, then, there exists $\eta \in [a, b]$ such that $\int_{(a,b]} \Psi d\phi = \Psi(\eta)\phi(b)$, which is the desired result.

Lemma 8.68: 8.42.

There is a constant $C < \infty$ such that for every $m \geqslant 0$ and every $[a, b] \subset [-\frac{1}{2}, \frac{1}{2}]$,

$$\left|\int_a^b D_m(x)dx\right| \leqslant C$$

Moreover, $\int_{-1/2}^0 D_m(x)dx = \int_0^{1/2} D_m(x)dx = \frac{1}{2}$ for all m.

Proof. By Equation (8.66.1), $\int_a^b D_m(x)dx = \int_a^b \frac{\sin(2m+1)\pi x}{\pi x}dx + \int_a^b \sin(2m + 1)\pi x[\frac{1}{\sin \pi x} - \frac{1}{\pi x}]dx$.

Since $(\sin \pi x)^{-1} - (\pi x)^{-1}$ is bounded on $[-\frac{1}{2}, \frac{1}{2}]$ and $|\sin(2m + 1)\pi x| \leqslant 1$, the second integral on the right is bounded in absolute value by a constant. With the substitution $y = (2m + 1)\pi x$, the first one becomes

$$\int_{(2m+1)\pi a}^{(2m+1)\pi b} \frac{\sin y}{\pi y}dy = \frac{\text{Si}[(2m + 1)\pi b] - \text{Si}[(2m + 1)\pi a]}{\pi}$$

where $\text{Si}(x) = \int_0^x y^{-1} \sin y dy$. But $\text{Si}(x)$ is continuous and approaches the finite limits $\pm\frac{1}{2}\pi$ as $x \to \pm\infty$ (see Folland Exercise 2.59(b)), so $\text{Si}(x)$ is bounded. This proves the first assertion. As for the second one,

$$\int_{-1/2}^{1/2} D_m(x)dx = \sum_{-m}^m \int_{-1/2}^{1/2} e^{2\pi ikx}dx = 1$$

(only the term with $k = 0$ is nonzero), so since D_m is even,

$$\int_{-1/2}^{0} D_m(x)dx = \int_{0}^{\frac{1}{2}} D_m(x)dx = \frac{1}{2}$$

Theorem 8.69: 8.43.

If $f \in BV(\mathbb{T})$—that is, if f is periodic on $\mathbb{T}$ and of bounded variation on $\left[-\frac{1}{2}, \frac{1}{2}\right]$—then

$$\lim_{m\to\infty} S_m f(x) = \frac{1}{2}[f(x+) + f(x-)] \text{ for every } x$$

In particular, $\lim_{m\to\infty} S_m f(x) = f(x)$ at every x at which f is continuous.

Proof. We begin by making some reductions. In examining the convergence of $S_m f(x)$, we may assume that $x = 0$ (by replacing f with the translated function $\tau_{-x} f$), that f is real-valued (by considering the real and imaginary parts separately), and that f is right continuous (since replacing $f(t)$ by $f(t+)$ affects neither $S_m f$ nor $\frac{1}{2}[f(0+) + f(0-)]$). In this case, by Theorem 3.27 b, on the interval $\left[-\frac{1}{2}, \frac{1}{2}\right)$ we can write f as the difference of two right continuous increasing functions g and h. If these functions are extended to $\mathbb{R}$ by periodicity, they are again of bounded variation, and it is enough to show that $S_m g(0) \to \frac{1}{2}[g(0+) + g(0-)]$ and likewise for h.

In short, it suffices to consider the case where $x = 0$ and f is increasing and right continuous on $\left[-\frac{1}{2}, \frac{1}{2}\right)$. Since D_m is even, we have $S_m f(0) = f * D_m(0) = \int_{-1/2}^{1/2} f(x) D_m(x)dx$, so by Lemma 68,

$$S_m f(0) - \tfrac{1}{2}[f(0+) + f(0-)]$$
$$= \int_0^{1/2} [f(x) - f(0+)] D_m(x)dx + \int_{-1/2}^{0} [f(x) - f(0-)] D_m(x)dx$$

We shall show that the first integral on the right tends to zero as $m \to \infty$; a similar argument shows that the second integral also tends to zero, thereby completing the proof.

Given $\varepsilon > 0$, choose $\delta > 0$ small enough so that $f(\delta) - f(0+) < \varepsilon/C$ where C is as in Lemma 68. Then by Lemma 67, for some $\eta \in [0, \delta]$,

$$\left|\int_0^{\delta} [f(x) - f(0+)] D_m(x)dx\right| = [f(\delta) - f(0+)]\left|\int_{\eta}^{\delta} D_m(x)dx\right|,$$

which is less than ε. On the other hand, by (8.40),

$$\int_{\delta}^{1/2} [f(x) - f(0+)] D_m(x)dx = \widehat{g}_+(-m) - \widehat{g}_-(m)$$

where $g_\pm$is the periodic function given on the interval $\left[-\frac{1}{2}, \frac{1}{2}\right)$ by

$$g_\pm(x) = \frac{[f(x) - f(0+)]e^{\pm\pi i x}}{2i \sin \pi x} \chi_{[\delta,1/2)}(x)$$

But $g_\pm \in L^1(\mathbb{T})$, so $\widehat{g}_\pm(\mp m) \to 0$ as $m \to \infty$ by the Riemann–Lebesgue lemma (the periodic analogue of Theorem 31f). Therefore,

$$\limsup_{m\to\infty}\left|\int_0^{1/2}[f(x)-f(0+)]D_m(x)dx\right|<\varepsilon$$

for every $\varepsilon>0$, and we are done.

One of the less attractive features of Fourier series is that bad behavior of a function at one point affects the behavior of its Fourier series at all points. For example, if f has even one jump discontinuity, then $\widehat{f}$ cannot be in $\ell^1(\mathbb{Z})$ and so the series $\sum \widehat{f}(k)e^{2\pi ikx}$ cannot converge absolutely at any point. However, to a limited extent the convergence of the series at a point x depends only on the behavior of f near x, as explained in the following localization theorem.

Theorem 8.70: 8.44.

If f and g are in $L^1(\mathbb{T})$ and $f=g$ on an open interval I, then $S_mf-S_mg\to 0$ uniformly on compact subsets of I.

Proof. It is enough to assume that $g=0$ (consider $f-g$), and by translating f we may assume that I is centered at 0, say $I=(-c,c)$ where $c\leqslant\frac{1}{2}$. Fix $\delta<c$; we shall show that if $f=0$ on I then $S_mf\to 0$ uniformly on $[-\delta,\delta]$.

The first step is to show that $S_mf\to 0$ pointwise on $[-\delta,\delta]$, and the argument is similar to the preceding proof. Namely, by (8.40) we have

$$S_mf(x)=\int_{-1/2}^{1/2}f(x-y)D_m(y)dy=\widehat{g}_{x,+}(-m)-\widehat{g}_{x,-}(m)$$

where

$$g_{x,\pm}(y)=\frac{f(x-y)e^{\pm\pi iy}}{2i\sin\pi y}$$

Since $f(x-y)=0$ on a neighborhood of the zeros of $\sin\pi y$, the functions $g_{x,\pm}$ are in $L^1(\mathbb{T})$, so $\widehat{g}_{x,\pm}(\mp m)\to 0$ by the Riemann–Lebesgue lemma.

The next step is to show that if $x_1,x_2\in[-\delta,\delta]$, then $S_mf(x_1)-S_mf(x_2)$ vanishes as $x_1-x_2\to 0$, uniformly in m. By (8.40) again,

$$S_mf(x_1)-S_mf(x_2)=\int_{-1/2}^{1/2}\frac{\sin(2m+1)\pi y}{\sin\pi y}[f(x_1-y)-f(x_2-y)]dy.$$

But $f(x_1-y)-f(x_2-y)=0$ for $|y|<c-\delta$, and for $c-\delta\leqslant|y|\leqslant\frac{1}{2}$ we have

$$\left|\frac{\sin(2m+1)\pi y}{\sin\pi y}\right|\leqslant\frac{1}{\sin\pi(c-\delta)}=A$$

where A is independent of m. Hence

$$|S_mf(x_1)-S_mf(x_2)|\leqslant A\int_{-1/2}^{1/2}|f(x_1-y)-f(x_2-y)|dy=A\|\tau_{x_1}f-\tau_{x_2}f\|_1$$

which vanishes as $x_1 - x_2 \to 0$ by (the periodic analogue of) Proposition 4.

Now, given $\varepsilon > 0$, we can choose η small enough so that if $x_1, x_2 \in [-\delta, \delta]$ and $|x_1 - x_2| < \eta$, then $|S_m f(x_1) - S_m f(x_2)| < \varepsilon/2$. Choose $x_1, \ldots, x_k \in [-\delta, \delta]$ so that the intervals $|x - x_j| < \eta$ cover $[-\delta, \delta]$. Since $S_m f(x_j) \to 0$ for each j, we can choose M large enough so that $|S_m f(x_j)| < \varepsilon/2$ for $m > M$ and $1 \leqslant j \leqslant k$. If $|x| \leqslant \delta$, then, we have $|x - x_j| < \eta$ for some j, so

$$|S_m f(x)| \leqslant |S_m f(x) - S_m f(x_j)| + |S_m f(x_j)| < \varepsilon$$

for $m > M$, and we are done.

Corollary 8.71: 8.45.

Suppose that $f \in L^1(\mathbb{T})$ and I is an open interval of length $\leqslant 1$.

(a) If f agrees on I with a function g such that $\widehat{g} \in \ell^1(\mathbb{Z})$, then $S_m f \to f$ uniformly on compact subsets of I.

(b) If f is absolutely continuous on I and $f' \in L^p(I)$ for some $p > 1$, then $S_m f \to f$ uniformly on compact subsets of I.

Proof. If $f = g$ on I, then $S_m f - f = S_m f - g = (S_m f - S_m g) + (S_m g - g)$ on I, and if $\widehat{g} \in \ell^1(\mathbb{Z})$, then $S_m g \to g$ uniformly on $\mathbb{Z}$; (a) follows. As for (b), given $[a_0, b_0] \subset I$, pick $a < a_0$ and $b > b_0$ so that $[a, b] \subset I$, and let g be the continuous periodic function that equals f on $[a, b]$ and is linear on $[b, a+1]$ (which is unique since $g(b) = f(b)$ and $g(a+1) = g(a) = f(a)$). Under the hypotheses of (b), g is absolutely continuous on $\mathbb{Z}$ and $g' \in L^p(\mathbb{Z})$, so $\widehat{g} \in \ell^1(\mathbb{Z})$ by Theorem 54. Thus $S_m f \to f$ uniformly on $[a_0, b_0]$ by (a).

Finally, we discuss the behavior of $S_m f$ near a jump discontinuity of f. Let us first consider a simple example: Let

$$\phi(x) = \frac{1}{2} - x - [x] \quad ([x] = \text{ greatest integer } \leqslant x).$$

Then ϕ is periodic and is C^∞ except for jump discontinuities at the integers, where $\phi(j+) - \phi(j-) = 1$. It is easy to check that $\widehat{\phi}(0) = 0$ and $\widehat{\phi}(k) = (2\pi i k)^{-1}$ for $k \neq 0$ (Folland Exercise 8.13(a)), so that

$$S_m \phi(x) = \sum\nolimits_{0<|k|\leqslant m} \frac{e^{2\pi i k x}}{2\pi i k} = \sum\nolimits_1^m \frac{\sin 2\pi k x}{\pi k}$$

From Corollary 71 it follows that $S_m \phi \to \phi$ uniformly on any compact set not containing an integer, and it is obvious that $S_m \phi(x) = 0$ when x is an integer. But near the integers a peculiar thing happens: $S_m \phi$ contains a sequence of spikes that overshoot and undershoot ϕ, as shown in Figure 8.1, and as $m \to \infty$ the spikes tend to zero in width but not in height. In fact, when m is large the value of $S_m \phi$ at its first maximum to the right of 0 is about 0.5895, about 18% greater than $\phi(0+) = \frac{1}{2}$. This is known as the Gibbs phenomenon; the precise statement and proof are given in Folland Exercise 8.37.

Now suppose that f is any periodic function on $\mathbb{R}$ having a jump discontinuity at

$x = a$ (that is, $f(a+)$ and $f(a-)$ exist and are unequal). Then the function

$$g(x) = f(x) - [f(a+) - f(a-)]\phi(x-a)$$

is continuous at every point where f is, and also at $x = a$ provided that we (re)define $g(a)$ to be $\frac{1}{2}[f(a+) + f(a-)]$, as the jumps in f and ϕ cancel out. If g satisfies one of the hypotheses of Corollary 71 on an interval I containing a, the Fourier series of g will converge uniformly near a, and hence the Fourier series of f will exhibit the same Gibbs phenomenon as that of ϕ.

Finally, suppose that f is periodic and continuous except at finitely many points $a_1, \ldots, a_k \in \mathbb{T}$, where f has jump discontinuities. We can then subtract off all the jumps to form a continuous function g:

$$g(x) = f(x) - \sum [f(a_j+) - f(a_j-)]\phi(x - a_j)$$

If f satisfies some mild smoothness conditions—for example, if f is absolutely continuous on any interval not containing any a_j and $f' \in L^p$ for some $p > 1$—then $\widehat{g}$ will be in $\ell^1(\mathbb{Z})$. Conclusion: $S_m f \to f$ uniformly on any interval not containing any a_j, $S_m(a_j) \to \frac{1}{2}[f(a_j+) + f(a_j-)]$, and $S_m f$ exhibits the Gibbs phenomenon near every a_j.

Exercise 8.72: Folland Exercise 8.33.

Let $\sigma_m f$ be the Cesr̀o means of the Fourier series of f given by (8.39).

(a) $\sigma_m f = f * F_m$ where $F_m = (m+1)^{-1} \sum_0^m D_k$ and D_k is the kth Dirichlet kernel. (See Folland Exercise 8.29(a).) F_m is called the mth Fejér kernel.

(b) $F_m(x) = \sin^2(m+1)\pi x / (m+1)\sin^2 \pi x$. (Use (8.40) and the fact that $\sin(2k+1)\pi x = \operatorname{Im} e^{(2k+1)\pi i x}$.)

Exercise 8.73: Folland Exercise 8.34.

If D_m is the mth Dirichlet kernel, $\|D_m\|_1 \to \infty$ as $m \to \infty$. (Make the substitution $y = (2m+1)\pi x$ and use Folland Exercise 2.59(a)..)

Exercise 8.74: Folland Exercise 8.35.

The purpose of this exercise is to show that the Fourier series of "most" continuous functions on $\mathbb{T}$ do not converge pointwise.

(a) Define $\phi_m(f) = S_m f(0)$. Then $\phi \in C(\mathbb{T})^*$ and $\|\phi\| = \|D_m\|_1$.

(b) The set of all $f \in C(\mathbb{T})$ such that the sequence $\{S_m f(0)\}$ converges is meager in $C(\mathbb{T})$. (Use Folland Exercise 8.34 and the uniform boundedness principle.)

(c) There exist $f \in C(\mathbb{T})$ (in fact, a residual set of such fs) such that $\{S_m f(x)\}$ diverges for every x in a dense subset of $\mathbb{T}$. (The result of (b) holds if the point 0 is replaced by any other point in $\mathbb{T}$. Apply Folland Exercise 5.40..)

Exercise 8.75: Folland Exercise 8.36.

The Fourier transform is not surjective from $L^1(\mathbb{T})$ to $C_0(\mathbb{T})$. (Use Folland Exercise 8.34, and confer with Folland Exercise 8.16(c).)

Exercise 8.76: Folland Exercise 8.37.

(a) Let ϕ be given by (8.46) and let $\Delta_m = S_m\phi - \phi$. Then $(d/dx)\Delta_m(x) = D_m(x)$ for $x \notin \mathbb{Z}$.

(b) The first maximum of Δ_m to the right of 0 occurs at $x = (2m+1)^{-1}$, and

$$\lim_{m\to\infty} \Delta_m\left(\frac{1}{2m+1}\right) = \frac{1}{\pi}\int_0^\pi \frac{\sin t}{t}dt - \frac{1}{2} \cong 0.0895$$

(Use (8.40) and the fact that $\Delta_m(x) = \int_0^x \Delta_m'(t)dt - \frac{1}{2}$.)

(c) More generally, the jth critical point of Δ_m to the right of 0 occurs at $x = j/(2m+1)(j = 1, \dots, 2m)$, and

$$\lim_{m\to\infty} \Delta_m\left(\frac{j}{2m+1}\right) = \frac{1}{\pi}\int_0^{j\pi} \frac{\sin t}{t}dt - \frac{1}{2}$$

These numbers are positive for j odd and negative for j even. (See Folland Exercise 2.59(b))

8.6 Fourier Analysis of Measures

We recall that $M(\mathbb{R}^n)$ is the space of complex Borel measures on $\mathbb{R}^n$ (which are automatically Radon measures by **??**), and we embed $L^1(\mathbb{R}^n)$ into $M(\mathbb{R}^n)$ by identifying $f \in L^1$ with the measure $d\mu = f dm$. We shall need to define products of complex measures on Cartesian product spaces, which can easily be done in terms of products of positive measures by using Radon-Nikodym derivatives. Namely, if $\mu, \nu \in M(\mathbb{R}^n)$, we define $\mu \times \nu \in M(\mathbb{R}^n \times \mathbb{R}^n)$ by

$$d(\mu\times\nu)(x,y) = \frac{d\mu}{d|\mu|}(x)\frac{d\nu}{d|\nu|}(y)d(|\mu|\times|\nu|)(x,y)$$

If $\mu, \nu \in M(\mathbb{R}^n)$, we define their convolution $\mu * \nu \in M(\mathbb{R}^n)$ by $\mu * \nu(E) = \mu\times\nu(\alpha^{-1}(E))$ where $\alpha\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is addition, $\alpha(x,y) = x+y$. In other words,

$$\mu\times\nu(E) = \iint \chi_E(x+y)d\mu(x)d\nu(y)$$

Proposition 8.77: 8.48.

(a) Convolution of measures is commutative and associative.

(b) For any bounded Borel measurable function h,

$$\int hd(\mu * \nu) = \iint h(x+y)d\mu(x)d\nu(y).$$

(c) $\|\mu * \nu\| \leqslant \|\mu\|\|\nu\|$.

(d) If $d\mu = fdm$ and $d\nu = gdm$, then $d(\mu * \nu) = (f * g)dm$; that is, on L^1 the new and old definitions of convolution coincide.

Proof. Commutativity is obvious from Fubini's theorem, as is associativity, for $\lambda * \mu * \nu$ is unambiguously defined by the formula

$$\lambda * \mu * \nu(E) = \iiint \chi_E(x+y+z)d\lambda(x)d\mu(y)d\nu(z)$$

Assertion (b) follows from (8.47) by the usual linearity and approximation arguments. In particular, taking $h = d|\mu * \nu|/d(\mu * \nu)$, since $|h| = 1$ we obtain

$$\|\mu * \nu\| = \int hd(\mu * \nu) \leqslant \iint |h|d|\mu|d|\nu| = \|\mu\|\|\nu\|$$

which proves (c). Finally, if $d\mu = fdm$ and $d\nu = gdm$, for any bounded measurable h we have

$$\begin{aligned}\int hd(\mu * \nu) &= \iint h(x+y)f(x)g(y)dxdy \\ &= \iint h(x)f(x-y)g(y)dxdy = \int h(x)(f * g)(x)dx\end{aligned}$$

whence $d(\mu * \nu) = (f * g)dm$.

We can also define convolutions of measures with functions in $L^p(\mathbb{R}^n, m)$, which we implicitly assume to be Borel measurable. (By Proposition 22, this is no restriction.)

Proposition 8.78: 8.49.

If $f \in L^p(\mathbb{R}^n)(1 \leqslant p \leqslant \infty)$ and $\mu \in M(\mathbb{R}^n)$, then the integral $f * \mu(x) = \int f(x-y)d\mu(y)$ exists for a.e. x, $f * \mu \in L^p$, and $\|f * \mu\|_p \leqslant \|f\|_p\|\mu\|$. (Here " L^p" and "a.e." refer to Lebesgue measure.)

Proof. If f and μ are nonnegative, then $f * \mu(x)$ exists (possibly being equal to ∞) for every x, and by Minkowski's inequality for integrals,

$$\|f * \mu\|_p \leqslant \int \|f(\cdot - y)\|_p d\mu(y) = \|f\|_p\|\mu\|$$

In particular, $f * \mu(x) < \infty$ for a.e. x. In the general case this argument applies to $|f|$ and $|\mu|$, and the result follows easily.

In the case $p = 1$, the definition of $f * \mu$ in Proposition 78 coincides with the definition given earlier in which f is identified with $f\,dm$, for

$$\int_E f * \mu(x)dx = \iint \chi_E(x)f(x-y)d\mu(y)dx = \iint \chi_E(x+y)f(x)dxd\mu(y)$$

for any Borel set E. Thus $L^1(\mathbb{R}^n)$ is not merely a subalgebra of $M(\mathbb{R}^n)$ with respect to convolution but an ideal.

We extend the Fourier transform from $L^1(\mathbb{R}^n)$ to $M(\mathbb{R}^n)$ in the obvious way: If $\mu \in M(\mathbb{R}^n)$, $\widehat{\mu}$ is the function defined by

$$\widehat{\mu}(\xi) = \int e^{-2\pi i\xi\cdot x}d\mu(x)$$

(The Fourier transform on measures is sometimes called the Fourier-Stieltjes transform.) Since $e^{-2\pi i\xi\cdot x}$ is uniformly continuous in x, it is clear that $\widehat{\mu}$ is a bounded continuous function and that $\|\widehat{\mu}\|_u \leqslant \|\mu\|$. Moreover, by taking $h(x) = e^{-2\pi i\xi\cdot x}$ in Proposition 77b, one sees immediately that $(\mu * \nu)^\wedge = \widehat{\mu}\widehat{\nu}$.

We conclude by giving a useful criterion for vague convergence of measures in terms of Fourier transforms.

Proposition 8.79: 8.50.

Suppose that $\mu_1, \mu_2, \ldots$, and μ are in $M(\mathbb{R}^n)$. If $\|\mu_k\| \leqslant C < \infty$ for all k and $\widehat{\mu}_k \to \widehat{\mu}$ pointwise, then $\mu_k \to \mu$ vaguely.

Proof. If $f \in \mathcal{S}$, then $f^\vee \in \mathcal{S}$ (Corollary 32), so by the Fourier inversion theorem,

$$\int f d\mu_k = \iint f^\vee(y)e^{-2\pi iy\cdot x}dyd\mu_k(x) = \int f^\vee(y)\widehat{\mu}_k(y)dy$$

Since $f^\vee \in L^1$ and $\|\widehat{\mu}_k\|_u \leqslant C$, the dominated convergence theorem implies that $\int f d\mu_k \to \int f d\mu$. But $\mathcal{S}$ is dense in $C_0(\mathbb{R}^n)$ (Proposition 19), so by Proposition 89, $\int f d\mu_k \to \int f d\mu$ for all $f \in C_0(\mathbb{R}^n)$, that is, $\mu_k \to \mu$ vaguely.

This result has a partial converse: If $\mu_k \to \mu$ vaguely and $\|\mu_k\| \to \|\mu\|$, then $\widehat{\mu}_k \to \widehat{\mu}$ pointwise. This follows from Folland Exercise 7.3.

Exercise 8.80: Folland Exercise 8.38.

Work out the analogues of the results in this section for measures on the torus $\mathbb{T}^n$.

Exercise 8.81: Folland Exercise 8.39.

If μ is a positive Borel measure on $\mathbb{T}$ with $\mu(\mathbb{T}) = 1$, then $|\widehat{\mu}(k)| < 1$ for all $k \neq 0$ unless μ is a linear combination, with positive coefficients, of the point masses at $0, \frac{1}{m}, \ldots, \frac{m-1}{m}$ for some $m \in \mathbb{T}$, in which case $\widehat{\mu}(jm) = 1$ for all $j \in \mathbb{T}$.

Exercise 8.82: Folland Exercise 8.40.

$L^1(\mathbb{R}^n)$ is vaguely dense in $M(\mathbb{R}^n)$. (If $\mu \in M(\mathbb{R}^n)$, consider $\phi_t * \mu$ where $\{\phi_t\}_{t>0}$ is an approximate identity.)

Exercise 8.83: Folland Exercise 8.41.

Let Δ be the set of finite linear combinations of the point masses $\delta_x, x \in \mathbb{R}^n$. Then Δ is vaguely dense in $M(\mathbb{R}^n)$. (If f is in the dense subset $C_c(\mathbb{R}^n)$ of $L^1(\mathbb{R}^n)$ and $g \in C_0(\mathbb{R}^n)$, approximate $\int fg$ by Riemann sums. Then use Folland Exercise 8.40.)

Exercise 8.84: Folland Exercise 8.42.

A function ϕ on $\mathbb{R}^n$ that satisfies $\sum_{j,k=1}^m z_j \overline{z}_k \phi(x_j - x_k) \geqslant 0$ for all $z_1, \ldots, z_m \in \mathbb{R}$ and all $x_1, \ldots, x_m \in \mathbb{R}^n$, for any $m \in \mathbb{R}$, is called positive definite. If $\mu \in M(\mathbb{R}^n)$ is positive, then $\widehat{\mu}$ is positive definite.

8.7 Applications to Partial Differential Equations

In this section we present a few of the many applications of Fourier analysis to the theory of partial differential equations; others will be found in Chapter 9. We shall use the term differential operator to mean a linear partial differential operator with smooth coefficients, that is, an operator L of the form

$$Lf(x) = \sum_{|\alpha| \leqslant m} a_\alpha(x) \partial^\alpha f(x), \quad a_\alpha \in C^\infty$$

If the a_αs are constants, we call L a constant-coefficient operator. In this case, if for all sufficiently well-behaved functions f (for example, $f \in \mathcal{S}$) we have

$$\widehat{(Lf)(\xi)} = \sum_{|\alpha| \leqslant m} a_\alpha (2\pi i \xi)^\alpha \widehat{f}(\xi)$$

It is therefore convenient to write L in a slightly different form: We set $b_\alpha = (2\pi i)^{|\alpha|} a_\alpha$ and introduce the operators

$$D^\alpha = (2\pi i)^{-|\alpha|} \partial^\alpha$$

so that

$$L = \sum_{|\alpha| \leqslant m} b_\alpha D^\alpha, \quad (Lf) = \sum_{|\alpha| \leqslant m} b_\alpha \xi^\alpha \widehat{f}$$

Thus, if P is any polynomial in n complex variables, say $P(\xi) = \sum_{|\alpha| \leqslant m} b_\alpha \xi^\alpha$, we can form the constant-coefficient operator $P(D) = \sum_{|\alpha| \leqslant m} b_\alpha D^\alpha$, and we then have $\widehat{[P(D)f]} = P\widehat{f}$. The polynomial P is called the symbol of the operator $P(D)$.

eaClearly, one potential application of the Fourier transform is in finding solutions of the differential equation $P(D)u = f$. Indeed, application of the Fourier transform to both sides yields $\widehat{u} = P^{-1}\widehat{f}$, whence $u = (P^{-1}\widehat{f})^\vee$. Moreover, if P^{-1} is the Fourier transform of a function ϕ, we can express u directly in terms of f as $u = f * \phi$. For these calculations to make sense, however, the functions f and $P^{-1}\widehat{f}$ (or P^{-1}) must be ones to which the Fourier transform can be applied, which is a serious limitation within the theory we have developed so far. The full power of this method becomes available only when the the domain of the Fourier transform is substantially extended. We shall do this in Folland Section 9.2; for the time being, we invite the reader to work out a fairly simple example in Folland Exercise 8.43. (It must also be pointed out that even when this method works, $u = (P^{-1}\widehat{f})^\vee$ is far from being the only solution of $P(D)u = f$; there are others that grow too fast at infinity to be within the scope even of the extended Fourier transform.)

Let us turn to some more concrete problems. The most important of all partial differential operators is the Laplacian

$$\Delta = \sum_1^n \frac{\partial \ell^2}{\partial x_j^2} = -4\pi^2 \sum_1^n D_j^2 = P(D) \text{ where } P(\xi) = -4\pi^2|\xi|^2$$

The reason for this is that Δ is essentially the only (scalar) differential operator that is invariant under translations and rotations. (If one considers operators on vector-valued functions, there are others, such as the familiar grad, curl, and div of 3-dimensional vector analysis.) More precisely, we have:

Theorem 8.85: 8.51.

A differential operator L satisfies $L(f \circ T) = (Lf) \circ T$ for all translations and rotations T if and only if there is a polynomial P in one variable such that $L = P(\Delta)$.

Proof. Clearly L is translation-invariant if and only if L has constant coefficients, in which case $L = Q(D)$ for some polynomial Q in n variables. Moreover, since $(Lf)^\wedge = Q\widehat{f}$ and the Fourier transform commutes with rotations, L commutes with rotations if and only if Q is rotation-invariant. Let $Q = \sum_0^m Q_j$ where Q_j is homogeneous of degree j; then it is easy to see that Q is rotation-invariant if and only if each Q_j is rotation-invariant. (Use induction on j and the fact that $Q_j(\xi) = \lim_{r\to 0} r^{-j} \sum_j^m Q_i(r\xi)$.) But this means that $Q_j(\xi)$ depends only on $|\xi|$, so $Q_j(\xi) = c_j|\xi|^j$ by homogeneity. Moreover, $|\xi|^j$ is a polynomial precisely when j is even, so $c_j = 0$ for j odd. Setting $b_k = (-4\pi^2)^{-k} c_{2k}$, then, we have $Q(\xi) = \sum b_k(-4\pi^2|\xi|^2)^k$, that is, $L = \sum b_k \Delta^k$.

One of the basic boundary value problems for the Laplacian is the Dirichlet problem: Given an open set $\Omega \subset \mathbb{R}^n$ and a function f on its boundary $\partial\Omega$, find a function u on $\overline{\Omega}$ such that $\Delta u = 0$ on Ω and $u|\partial\Omega = f$. (This statement of the problem is deliberately a bit imprecise.) We shall solve the Dirichlet problem when Ω is a half-space.

For this purpose it will be convenient to replace n by $n+1$ and to denote the coordinates on $\mathbb{R}^{n+1}$ by $x_1, \ldots, x_n, t$. We continue to use the symbol Δ to denote the Laplacian on $\mathbb{R}^n$, and we set

$$\partial_t = \frac{\partial}{\partial t}$$

so the Laplacian on $\mathbb{R}^{n+1}$ is $\Delta + \partial_t^2$. We take the half-space Ω to be $\mathbb{R}^n \times (0, \infty)$. Thus, given a function f on $\mathbb{R}^n$, satisfying conditions to be made more precise below, we wish to find a function u on $\mathbb{R}^n \times [0, \infty)$ such that $(\Delta + \partial_t^2)u = 0$ and $u(x, 0) = f(x)$.

The idea is to apply the Fourier transform on $\mathbb{R}^n$, thus converting the partial differential equation $(\Delta + \partial_t^2)u = 0$ into the simple ordinary differential equation $(-4\pi^2|\xi|^2 + \partial_t^2)\widehat{u} = 0$. The general solution of this equation is

$$\widehat{u}(\xi, t) = c_1(\xi)e^{-2\pi t|\xi|} + c_2(\xi)e^{2\pi t|\xi|}$$

and we require that $\widehat{u}(\xi, 0) = \widehat{f}(\xi)$. We therefore obtain a solution to our problem by taking $c_1(\xi) = \widehat{f}(\xi), c_2(\xi) = 0$ (more about the reasons for this choice below); this gives $\widehat{u}(\xi, t) = \widehat{f}(\xi)e^{-2\pi t|\xi|}$, or $u(x, t) = (f * P_t)(x)$ where $P_t = (e^{-2\pi t|\xi|})^\vee$ is the Poisson kernel introduced in Folland Section 8.4. As we calculated in Folland Exercise 8.26,

$$P_t(x) = \frac{\Gamma\left(\frac{1}{2}(n+1)\right)}{\pi^{(n+1)/2}} \frac{t}{(t^2 + |x|^2)^{-(n+1)/2}}$$

So far this is all formal, since we have not specified conditions on f to ensure that these manipulations are justified. We now give a precise result.

Theorem 8.86: 8.53.

Suppose $f \in L^p(\mathbb{R}^n)(1 \leqslant p \leqslant \infty)$. Then the function $u(x, t) = (f * P_t)(x)$ satisfies $(\Delta + \partial_t^2)u = 0$ on $\mathbb{R}^n \times (0, \infty)$, and $\lim_{t \to 0} u(x, t) = f(x)$ for a.e. x and for every x at which f is continuous. Moreover, $\lim_{t \to 0} \|u(\cdot, t) - f\|_p = 0$ provided $p < \infty$.

Proof. P_t and all of its derivatives are in $L^q(\mathbb{R}^n)$ for $1 \leqslant q \leqslant \infty$, since a rough calculation shows that $|\partial_x^\alpha P_t(x)| \leqslant C_\alpha |x|^{-n-1-|\alpha|}$ and $|\partial_t^j P_t(x)| \leqslant C_j |x|^{-n-1}$ for large x. Also, $(\Delta + \partial_t^2)P_t(x) = 0$, as can be verified by direct calculation or (more easily) by taking the Fourier transform. Hence $f * P_t$ is well defined and

$$(\Delta + \partial_t^2)(f * P_t) = f * (\Delta + \partial_t^2)P_t = 0$$

Since $P_t(x) = t^{-n}P_1(t^{-1}x)$ and $\int P_1(x)dx = \widehat{P}_1(0) = 1$, the remaining assertions follow from Theorems 8.14 and 8.15.

The function $u(x, t) = (f * P_t)(x)$ is not the only one satisfying the conclusions of Theorem 86; for example, $v(x, t) = u(x, t) + ct$ also works, for any $c \in \mathbb{C}$. For $f \in L^1$, we could also obtain a large family of solutions by taking c_2 in (8.52) to be an arbitrary function in C_c^∞ and $c_1 = \widehat{f} - c_2$. (But there is no nice convolution formula for the resulting function u, because $e^{2\pi t|\xi|}$ is not the Fourier transform of a function or even a distribution.) The solution $u(x, t) = (f * P_t)(x)$ is distinguished, however, by its regularity at infinity; for example, it can be shown that if $f \in BC(\mathbb{C}^n)$, then u is the unique solution in $BC(\mathbb{C}^n \times [0, \infty))$.

The same idea can be used to solve the heat equation

$$(\partial_t - \Delta)u = 0$$

on $\mathbb{R}^n \times (0, \infty)$ subject to the initial condition $u(x, 0) = f(x)$. (Physical interpretation: $u(x, t)$ represents the temperature at position x and time t in a homogeneous isotropic medium, given that the temperature at time 0 is $f(x)$.) Indeed, Fourier transformation leads to the ordinary differential equation $(\partial_t + 4\pi^2|\xi|^2)\widehat{u} = 0$ with initial condition $\widehat{u}(\xi, 0) = \widehat{f}(\xi)$. The unique solution of the latter problem is $\widehat{u}(\xi, t) = \widehat{f}(\xi)e^{-4\pi^2 t|\xi|^2}$. In view of Proposition 33, this yields

$$u(x, t) = f * G_t(x), \quad G_t(x) = (4\pi t)^{-n/2}e^{-|x|^2/4t}$$

Here we have $G_t(x) = t^{-n/2}G_1(t^{-1/2}x)$, so after the change of variable $s = \sqrt{t}$, Theorems 8.14 and 8.15 apply again, and we obtain an exact analogue of Theorem 86 for the initial value problem $(\partial_t - \Delta)u = 0, u(x, 0) = f(x)$. Actually, in the present case the hypotheses on f can be relaxed considerably because $G_t \in \mathcal{S}$; see Folland Exercise 8.44.

Another fundamental equation of mathematical physics is the wave equation

$$(\partial_t^2 - \Delta)u = 0$$

(Physical interpretation: $u(x, t)$ is the amplitude at position x and time t of a wave traveling in a homogeneous isotropic medium, with units chosen so that the speed of propagation is 1.) Here it is appropriate to specify both $u(x, 0)$ and $\partial_t u(x, 0)$:

$$(\partial_t^2 - \Delta)u = 0, \quad u(x, 0) = f(x), \quad \partial_t u(x, 0) = g(x)$$

After applying the Fourier transform, we obtain

$$(\partial_t^2 + 4\pi^2|\xi|^2)\widehat{u}(\xi, t) = 0, \quad \widehat{u}(\xi, 0) = \widehat{f}(\xi), \quad \partial_t\widehat{u}(\xi, 0) = \widehat{g}(\xi)$$

the solution to which is

$$\widehat{u}(\xi, t) = (\cos 2\pi t|\xi|)\widehat{f}(\xi) + \frac{\sin 2\pi t|\xi|}{2\pi|\xi|}\widehat{g}(\xi)$$

Since

$$\cos 2\pi t|\xi| = \frac{\partial}{\partial t}\left[\frac{\sin 2\pi t|\xi|}{2\pi|\xi|}\right]$$

it follows that

$$u(x, t) = f * \partial_t W_t(x) + g * W_t(x), \text{ where } W_t = \left[\frac{\sin 2\pi t|\xi|}{2\pi|\xi|}\right]^\vee$$

But here there is a problem: $(2\pi|\xi|)^{-1}\sin 2\pi t|\xi|$ is the Fourier transform of a function only when $n \leqslant 2$ and the Fourier transform of a measure only when $n \leqslant 3$; for these cases the resulting solution of the wave equation is worked out in Exercises 45-47. To carry out this analysis in higher dimensions requires the theory of distributions, which we shall examine in Chapter 9. (We shall not, however, derive the explicit formula for W_t, which becomes increasingly complicated as n increases.)

Exercise 8.87: Folland Exercise 8.43.

Let $\phi(x) = e^{-|x|/2}$ on $\mathbb{R}$. Use the Fourier transform to derive the solution $u = f * \phi$ of the differential equation $u - u'' = f$, and then check directly that it works. What hypotheses are needed on f?

Exercise 8.88: Folland Exercise 8.44.

Let $G_t(x) = (4\pi t)^{-n/2} e^{-|x|^2/4t}$, and suppose that $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ satisfies $|f(x)| \leqslant C_\varepsilon e^{\varepsilon |x|^2}$ for every $\varepsilon > 0$. Then $u(x,t) = f * G_t(x)$ is well defined for all $x \in \mathbb{R}^n$ and $t > 0$; $(\partial_t - \Delta)u = 0$ on $\mathbb{R}^n \times (0, \infty)$; and $\lim_{t \to 0} u(x,t) = f(x)$ for a.e. x and for every x at which f is continuous. (To show $u(x,t) \to f(x)$ a.e. on a bounded open set V, write $f = \phi f + (1-\phi) f$ where $\phi \in C_c$ and $\phi = 1$ on V, and show that $[(1-\phi)f] * G_t \to 0$ on V.)

Solution.

(i) For $x \in \mathbb{R}^n$ and $t > 0$, and choose $\varepsilon > 0$ such that $1 - 4t\varepsilon > 0$. Then by completing the square in the exponent and applying [Fol99, Proposition 2.53] we obtain

$$\begin{aligned}
|(4\pi t)^{n/2}(f * G_t)(x)| &= \int |f(y)| e^{-|x-y|^2/4t}\, dy \leqslant C_\varepsilon \int e^{\varepsilon |y|^2 - |x-y|^2/4t}\, dy \\
&\leqslant C_\varepsilon \int e^{\varepsilon |y|^2 - \frac{1}{4t}(|x|^2 - 2|x||y| + |y|^2)}\, dy \\
&\leqslant C_\varepsilon e^{\left(\left(\frac{4t}{1-4t\varepsilon}\right)^2 - \frac{1}{4t}\right)|x|^2} \int e^{-\frac{4t}{1-4t\varepsilon}\left(|y|^2 - \frac{8t}{1-4t\varepsilon}|x||y| + \left(\frac{4t}{1-4t\varepsilon}\right)^2 |x|^2\right)}\, dy \\
&\leqslant C_\varepsilon e^{\left(\left(\frac{4t}{1-4t\varepsilon}\right)^2 - \frac{1}{4t}\right)|x|^2} \int e^{-\frac{4t}{1-4t\varepsilon}\left(|y| - \frac{4t}{1-4t\varepsilon}|x|\right)^2}\, dy \\
&\leqslant C_\varepsilon e^{\left(\left(\frac{4t}{1-4t\varepsilon}\right)^2 - \frac{1}{4t}\right)|x|^2} \int e^{-\frac{4t}{1-4t\varepsilon}\left|y - \frac{4t}{1-4t\varepsilon}x\right|^2}\, dy \\
&\leqslant C_\varepsilon \pi^{n/2} (4t)^{-n/2} (1 - 4t\varepsilon)^{n/2} e^{\left(\left(\frac{4t}{1-4t\varepsilon}\right)^2 - \frac{1}{4t}\right)|x|^2} < \infty,
\end{aligned}$$

so $y \mapsto f(y) G_t(x-y)$ is in $L^1(\mathbb{R}^n)$. Thus $f * G_t(x)$ is well-defined for all $x \in \mathbb{R}^n$ and all $t > 0$.

(ii) We claim

$$(\partial_t - \Delta)(f * G_t) = 0. \tag{8.88.1}$$

By [Fol99, Proposition 8.24], the Fourier transform of G_t for $t > 0$ is given by

$$\widehat{G}_t(\xi) = (4\pi t)^{-n/2} (4t)^{n/2} e^{-4\pi t |\xi|^2} = \pi^{-n/2} e^{-4\pi t |\xi|^2}. \tag{8.88.2}$$

Applying the Fourier transform to Equation (8.88.1), we obtain by [Fol99, discussion on p. 273] and [Fol99, Theorem 8.22(c), p. 249]

$$(\partial_t + 4\pi^2 |\xi|^2)(f * G_t)^\wedge(\xi) = \pi^{-n/2} \widehat{f}(\xi) (\partial_t e^{-4\pi t |\xi|^2} + 4\pi^2 |\xi|^2 e^{-4\pi t |\xi|^2})$$

$$= \pi^{-n/2}\widehat{f}(\xi)\big(-4\pi|\xi|^2 e^{-4\pi t|\xi|^2} + 4\pi|\xi|e^{-4\pi t|\xi|^2}\big) = 0,$$

so $(\partial_t - \Delta)(f * G_t) = 0$ on $\mathbb{R}^n \times (0,\infty)$. This proves (ii).

(iii) Fix $t, r > 0$, and again choose $\varepsilon > 0$ such that $1 - 4t\varepsilon > 0$. For $x \in \mathbb{R}^n$, let $B_r(x)$ denote the open ball in $\mathbb{R}^n$ centered at x. Since $|f(x)| \leqslant C_\varepsilon e^{\varepsilon|x|^2}$ and $B_r(x)$ is bounded, $f \in L^p(B_r(x))$ for all $p \in [1,\infty]$. Now choose $\phi \in C_c(\mathbb{R}^n)$ such that $\phi|_{B_r(x)} = 1$. By estimating as in part (i) and noting $1 - \phi = 0$ on $B_r(x)$, we obtain

$$\begin{aligned}
|(1-\phi)f * G_s(x)| &\leqslant C_\varepsilon(4\pi t)^{-n/2}\int |1-\phi(y)|e^{\varepsilon|y|^2 - |x-y|^2/4t}\,\mathrm{d}y \\
&\leqslant (4\pi t)^{-n/2}C_\varepsilon e^{\left(\left(\frac{4t}{1-4t\varepsilon}\right)^2 - \frac{1}{4t}\right)|x|^2}\int |1-\phi(y))|e^{-\frac{4t}{1-4t\varepsilon}\left|y - \frac{4t}{1-4t\varepsilon}x\right|^2}\,\mathrm{d}y \\
&\leqslant (4\pi t)^{-n/2}C_\varepsilon e^{\left(\left(\frac{4t}{1-4t\varepsilon}\right)^2 - \frac{1}{4t}\right)|x|^2}\int_{B_r(x)} |1-\phi(y)|e^{-\frac{4t}{1-4t\varepsilon}\left|y - \frac{4t}{1-4t\varepsilon}x\right|^2}\,\mathrm{d}y \\
&\leqslant (4\pi t)^{-n/2}C_\varepsilon e^{\left(\left(\frac{4t}{1-4t\varepsilon}\right)^2 - \frac{1}{4t}\right)|x|^2}\|1-\phi\|_\infty\int_{B_r(x)} e^{-\frac{4t}{1-4t\varepsilon}\left|y - \frac{4t}{1-4t\varepsilon}x\right|^2}\,\mathrm{d}y \\
&= (4\pi t)^{-n/2}C_\varepsilon \pi^{n/2}(4t)^{-n/2}(1-4t\varepsilon)^{n/2}e^{\left(\left(\frac{4t}{1-4t\varepsilon}\right)^2 - \frac{1}{4t}\right)|x|^2}\|1-\phi\|_\infty,
\end{aligned}$$

and this is finite because $\phi \in C_c(\mathbb{R}^n)$ (so that $\|1-\phi\|_\infty < \infty$). Since the exponential decays to 0 faster than any polynomial as $t \to \infty$ (since $\left(\frac{4t}{1-4t\varepsilon}\right)^2 - \frac{1}{4t} = -1/4t + O(t^2)$), it follows that implies $|(1-\phi)f * G_t(x)| \to 0$ as $t \to 0$.

We claim $(\phi f) * G_t(x) \to f(x)$ as $t \to 0$ for a.e. x in the Lebesgue set of f. Now let $r = t$. Since $\phi = 1$ on $B_r(x)$, we have by taking t small enough so that $B_t(x) \subset \operatorname{supp}\phi$ that

$$\begin{aligned}
|(\phi f) * G_t(x) - f(x)| &= \left|\int f(y)G_t(x-y)\,\mathrm{d}y - f(x)\right| \\
&\leqslant \int_{B_t(x)} |f(y) - f(x)||G_t(x-y)|\,\mathrm{d}y \\
&\leqslant (4\pi t)^{-n/2}\int_{B_t(x)} |f(y) - f(x)|\,\mathrm{d}y, \\
&= C\left(\frac{1}{m(B_t(x))}\int_{B_t(x)} |f(y) - f(x)|\,\mathrm{d}y\right),
\end{aligned}$$

where C is the reciprocal of the constant given explicitly in [Fol99, Corollary 2.55, p. 80]. By Lebesgue's differentiation theorem [Fol99, Theorem 3.21, p. 98], the integral on the right-hand side converges to $f(x)$ for all x in the Lebesgue set of f. In particular, $\lim_{t\to 0} f * G_t(x) = f(x)$ for a.e. x and for all x at which f is continuous (see [Fol99, §3.4, Exercise 24, p. 100]). □

Exercise 8.89: Folland Exercise 8.45.

Let $n = 1$. Use (8.55) and Folland Exercise 8.15(a) to derive d'Alembert's solution to the initial value problem (8.54):

$$u(x,t) = \frac{1}{2}[f(x+t) + f(x-t)] + \frac{1}{2}\int_{x-t}^{x+t} g(s)ds$$

Under what conditions on f and g does this formula actually give a solution?

Exercise 8.90: Folland Exercise 8.46.

Let $n = 3$, and let σ_t denote surface measure on the sphere $|x| = t$. Then

$$\frac{\sin 2\pi t|\xi|}{2\pi|\xi|} = (4\pi t)^{-1}\widehat{\sigma}_t(\xi)$$

(See Folland Exercise 8.22(d).) What is the resulting solution of the initial value problem (8.54), expressed in terms of convolutions? What conditions on f and g ensure its validity?

Exercise 8.91: Folland Exercise 8.47.

Let $n = 2$. If $\xi \in \mathbb{R}^2$, let $\tilde{\xi} = (\xi, 0) \in \mathbb{R}^3$. Rewrite the result of Folland Exercise 8.46,

$$\frac{\sin 2\pi t|\tilde{\xi}|}{2\pi|\tilde{\xi}|} = \frac{1}{4\pi t}\int_{|x|=t} e^{-2\pi i\tilde{\xi}\cdot x} d\sigma_t(x)$$

in terms of an integral over the disc $D_t = \{y \mid |y| \leqslant t\}$ in $\mathbb{R}^2$ by projecting the upper and lower hemispheres of the sphere $|x| = t$ in $\mathbb{R}^3$ onto the equatorial plane. Conclude that $(2\pi|\xi|)^{-1}\sin 2\pi t|\xi|$ is the Fourier transform of

$$W_t(x) = (2\pi)^{-1}(t^2 - |x|^2)^{-1/2}\chi_{D_t}(x)$$

and write out the resulting solution of the initial value problem (8.54).

Exercise 8.92: Folland Exercise 8.48.

Solve the following initial value problems in terms of Fourier series, where f, g, and $u(\cdot, t)$ are periodic functions on $\mathbb{R}$:

(a) $(\partial_t^2 + \partial_x^2)u = 0, u(x,0) = f(x)$. (Cf. the discussion of Abel means in Folland Section 8.4.)

(b) $(\partial_t - \partial_x^2)u = 0, u(x,0) = f(x)$.

(c) $(\partial_t^2 - \partial_x^2)u = 0, u(x,0) = f(x), \partial_t u(x,0) = g(x)$.

Exercise 8.93: Folland Exercise 8.49.

In this exercise we discuss heat flow on an interval.

(a) Solve $(\partial_t - \partial_x^2)u = 0$ on $(a,b)\times(0,\infty)$ with boundary conditions $u(x,0) = f(x)$ for $x \in (a,b), u(a,t) = u(b,t) = 0$ for $t > 0$, in terms of Fourier series. (This describes heat flow on (a,b) when the endpoints are held at a constant temperature. It suffices to assume $a = 0, b = \frac{1}{2}$; extend f to $\mathbb{R}$ by requiring f to be odd and periodic, and use Folland Exercise 8.48(b).)

(b) Solve the same problem with the condition $u(a,t) = u(b,t) = 0$ replaced by $\partial_x u(a,t) = \partial_x u(b,t) = 0$. (This describes heat flow on (a,b) when the endpoints are insulated. This time, extend f to be even and periodic.)

Exercise 8.94: Folland Exercise 8.50.

Solve $(\partial_t^2 - \partial_x^2)u = 0$ on $(a,b)\times(0,\infty)$ with boundary conditions $u(x,0) = f(x)$ and $\partial_t u(x,0) = g(x)$ for $x \in (a,b), u(a,t) = u(b,t) = 0$ for $t > 0$, in terms of Fourier series by the method of Folland Exercise 8.49(a). (This problem describes the motion of a vibrating string that is fixed at the endpoints. It can also be solved by extending f to be odd and periodic and using Folland Exercise 8.45. That form of the solution tells you what you see when you look at a vibrating string; this one tells you what you hear when you listen to it.)

9 Extra section: Rate of decay of Fourier coefficients

The following theorems (and their proofs) are from 4/3–4/12 lectures.

The following is a partial solution to the "inverse Fourier series problem", which asks when a function has a prescribed Fourier series.

Theorem 9.1.

If $\{a_n\}_{n=-\infty}^{\infty} \subset \mathbb{R}$ is a nonnegative even sequence[a] that satisfies the condition[b]

$$a_n \leqslant \frac{1}{2}(a_{n+1} + a_{n-1}) \qquad \forall n \in \mathbb{Z}_{>0},$$

then there exists $f \in L^1(\mathbb{T})$ such that $a_n = \widehat{f}(n)$.

[a] By "even sequence" we mean $a_n = -a_n$ for $n \in \mathbb{Z}_{\geqslant 0}$.

[b] This is informally referred to as a "convexity condition" for reasons you can probably guess,

Corollary 9.2.

The Fourier series coefficients of $f \in L^1(\mathbb{T})$ tend to 0 at an arbitrarily slow rate.

Theorem 9.3.

If $f \in L^1(\mathbb{T})$ and $\widehat{f}$ is an odd function,[a] then

$$\sum_{n \in \mathbb{Z} \setminus \{0\}} \frac{\widehat{f}(n)}{n} < \infty.$$

[a]By "odd function" we mean $\widehat{f}(|n|) = -\widehat{f}(-|n|)$ for all $n \in \mathbb{Z}$.

For $\alpha \in (0,1)$, define

$$\mathrm{Lip}_\alpha(\mathbb{T}) := \{f \in C(\mathbb{T}) \mid \exists C > 0 \text{ such that } \forall x \in \mathbb{T}, |f(x+k) - f(x)| \leqslant C|k|^\alpha\}$$

$f \in \mathrm{Lip}_\alpha(\mathbb{T})$ means that $f \in C(\mathbb{T})$ and there exists $C > 0$ such that $|f(x+k) - f(x)| \leqslant C|k|^\alpha$ for all $x \in \mathbb{T}$.

Theorem 9.4.

If $f \in \mathrm{Lip}_\alpha(\mathbb{T})$, then $\widehat{f}(n) = O(n^{-\alpha})$ as $|n| \to \infty$.

Theorem 9.5.

If $f \in L^1(\mathbb{T})$ and $\widehat{f}(n) = O(1/n)$ as $|n| \to \infty$, then $S_n f(x)$ and $\sigma_n f(x)$, the symmetric partial sums and Cesàro partial sums, respectively, converge for the same values of x and to the same limit. Moreover, if $\sigma_n f(x)$ converges uniformly in a set E, then $S_n f(x)$ converges uniformly on E.

Corollary 9.6.

If $f \in \mathrm{BV}(\mathbb{T})$ then $S_n f(x) \to \lim_{k\to\infty} \frac{1}{2}(f(x+k) + f(x-k))$. If in addition $f \in C(\mathbb{T})$, then the Fourier series of f converges to f everywhere. (This is a consequence of Fejér's theorem and the fact that $f \in \mathrm{BV}(\mathbb{T})$ implies $\widehat{f}(n) = O(1/n)$ as $n \to \infty$.)

9.7 Principle of localization.

Suppose $f, g \in L^1(\mathbb{T})$ and $f(x) = g(x)$ in a neighborhood of y. Then the Fourier series of f and g at x either both converge to the same limit or both diverge.

Note 8. *The local behavior of f can affect the global behavior of its Fourier series. For instance, suppose f is continuous on $\mathbb{T}$ except at some point where f is a jump discontinuity.*

Then its Fourier series does not converge absolutely anywhere, that is, $\widehat{f} \notin \ell^1$. In fact, if $f \in L^1(\mathbb{T})$ and $\widehat{f} \in \ell^1$, then Fourier inversion holds in the sense that the Fourier series of f converges a.e. to f. (Indeed, $\sum_{n\in\mathbb{Z}} \widehat{f}(n)e^{2\pi inx}$ converges absolutely and uniformly to a continuous function g. On the other hand, since $\ell^1(\mathbb{T}) \subset \ell^2(\mathbb{T})$, the Fourier series converges in the L^2 norm to a function $f_0 \in L^2(\mathbb{T}) \subset L^1(\mathbb{T})$, so $\widehat{f_0}(n) = \widehat{f}(x)$ for all $n \in \mathbb{Z}$. Then by the uniqueness theorem for L^1 functions, $f = f_0$ a.e. and thus $f = g$ a.e.)

To show the principle of localization, it suffices to show that if $f \in L^1(\mathbb{T})$ and vanishes on an interval I, then $S_n f(x) \to 0$ as $n \to \infty$ for $x \in I$. In fact, if $f \in L^1(\mathbb{T})$ and

$$\int_{-1/2}^{1/2} \left|\frac{f(t)}{t}\right| \mathrm{d}t < \infty,$$

then $\lim_{n\to\infty} S_n f(0) = 0$.

9.1 Absolute convergence of Fourier series

As before, $C(\mathbb{T})$ denotes 1-periodic functions $f\colon \mathbb{R} \to \mathbb{C}$. Let

$$A(\mathbb{T}) := \left\{ f \in C(\mathbb{T}) \,\middle|\, \sum_{n\in\mathbb{Z}} |\widehat{f}(n)| < \infty \right\}.$$

Thus $A(\mathbb{T})$ is the set of 1-periodic functions $f\colon \mathbb{R} \to \mathbb{C}$ whose Fourier series converge absolutely.

Theorem 9.9: Sergei Bernstein, 1914.

For $\alpha \in (1/2, 1]$, $\mathrm{Lip}_\alpha(\mathbb{T}) \subset A(\mathbb{T})$.

Theorem 9.10: Antoni Zygmund, 1928.

For any $\alpha \in (0, 1)$, $\mathrm{Lip}_\alpha(\mathbb{T}) \cap \mathrm{BV}(\mathbb{T}) \subset A(\mathbb{T})$.

9.1.1 Application of multidimensional Fourier series to random walks

Consider a particle on the d-dimensional lattice $\mathbb{Z}^\ell$ that moves to a neighboring point in the lattice at each unit time interval. Assume each unit step u_n at time n is independent of each other and each possible direction has equal probability. The position at time n is $s_n = u_1 + \cdots + u_n$ and for given unit steps $e_1, \dots, e_n$, $p(u_1 = z_1, u_2 = e_2, \dots, e_n = e_n) = \prod_{j=1}^n p(u_j = e_j) = \left(\frac{1}{2d}\right)^n$.

Question. *Assume the particle starts at the origin. What is the expected number of times that it returns to the origin? (The answer depends on the dimension d).*

Theorem 9.11: Pólya, 1921.

$P(s_n = 0 \text{ infinitely often}) = 1$ when $d = 1$ or $d = 2$, and

$$P\Big(\lim_{n\to\infty}|s_n| = \infty\Big) = 1$$

when $d \in \mathbb{Z}_{\geqslant 1}$.

Theorem 9.12.

Assume $\Phi \in L^1(\mathbb{R}^n) \cap C_0(\mathbb{R}^n)$, $\Phi(0) = 1$, and $\Phi = \widehat{\varphi}$ where $\varphi \in L^1(\mathbb{R}^n)$. For $f \in L^1 + L^2$ and $\varepsilon > 0$, define

$$f_\varepsilon(x) \coloneqq \int_{\mathbb{R}^n} \Phi(\varepsilon\xi)\widehat{f}(\xi)e^{2\pi i x\cdot\xi}\,\mathrm{d}\xi.$$

(a) For $p \in [1,\infty)$, if $f \in L^p$ then $f_\varepsilon \to f$ in L^p.
(b) If f is bounded and uniformly continuous, then $f_\varepsilon \to f$ uniformly (and f_ε is uniformly continuous).
(c) If $|\varphi(x)| \leqslant C(1+|x|)^{-n-\sigma}$ for some $C, \sigma > 0$, then $f_\varepsilon(x) \to f(x)$ at point in the Lebesgue set of f (that is, when $\lim_{r\to 0}\frac{1}{m(B_r(x))}\int_{B_r(x)}|f(y)-f(x)|\,\mathrm{d}y = 0$), and thus $f_\varepsilon \to f$ pointwise a.e.

Proposition 9.13.

Suppose $f, g \in L^2(\mathbb{R}^n)$. Then $(\widehat{f}\widehat{g})^\vee = f * g$.

The discrete analog (for Fourier series) of this theorem is the following.

Theorem 9.14.

Assume $\Phi \in C(\mathbb{R}^n)$, $\Phi(0) = 1$, and $\Phi = \widehat{\varphi}$, where $|\Phi(\xi) \leqslant A(1+|\xi|)^{-n-\sigma}|$ and $|\varphi(x)|l \leqslant A(1+|x|)^{-n-\sigma}$ for some $\sigma > 0$. For $f \in L^1(\mathbb{T}^n)$ and $\varepsilon > 0$, define

$$f_\xi(x) \coloneqq \sum_{k\in\mathbb{Z}^n} \Phi(\varepsilon k)\widehat{f}(k)e^{2\pi i k\cdot x}.$$

(a) If $p \in [1,\infty)$ and $f \in L^p(\mathbb{T}^n)$, then $f_\varepsilon \to f$ in L^p. If $f \in C(\mathbb{T}^n)$, then $f_\varepsilon \to f$ uniformly.
(b) If x is a point in the Lebesgue set of f, then $f_\varepsilon(x) \to f(x)$. In particular, $f_\varepsilon \to f$ a.e.

Corollary 9.15.

For $p \in [1,\infty)$, the ***Riesz means*** $\sum_{k\in\mathbb{Z}^n,|k|\leqslant R}\Big(1-\frac{|k|^2}{R^2}\Big)^\alpha \widehat{f}(k)e^{2\pi i k\cdot x}$ converge to f in L^p and a.e. as $R \to \infty$ when $\alpha > (n-1)/2$.

Proof. See Folland. □

Corollary 15 is false for $\alpha = (n-1)/2$ and $p = 1$.

9 Elements of Distribution Theory

At least as far back as Heaviside in the 1890s, engineers and physicists have found it convenient to consider mathematical objects which, roughly speaking, resemble functions but are more singular than functions. Despite their evident efficacy, such objects were at first received with disdain and perplexity by the pure mathematicians, and one of the most important conceptual advances in modern analysis is the development of methods for dealing with them in a rigorous and systematic way. The method that has proved to be most generally useful is Laurent Schwartz's theory of distributions, based on the idea of linear functionals on test functions. For some purposes, however, it is preferable to use a theory more closely tied to L^2 on which the power of Hilbert space methods and the Plancherel theorem can be brought to bear, namely, the (L^2) Sobolev spaces. In this chapter we present the fundamentals of these theories and some of their applications.

9.1 Distributions

In order to find a fruitful generalization of the notion of function on $\mathbb{R}^n$, it is necessary to get away from the classical definition of function as a map that assigns to each point of $\mathbb{R}^n$ a numerical value. We have already done this to some extent in the theory of L^p spaces: If $f \in L^p$, the pointwise values $f(x)$ are of little significance for the behavior of f as an element of L^p, as f can be modified on any set of measure zero without affecting the latter. What is more to the point is the family of integrals $\int f\phi$ as ϕ ranges over the dual space L^q. Indeed, we know that f is completely determined by its action as a linear functional on L^q; on the other hand, if we take $\phi = \phi_r = m(B_r)^{-1}\chi_{B_r}$ where B_r is the ball of radius r about x, by the Lebesgue differentiation theorem we can recover the pointwise value $f(x)$, for almost every x, as $\lim_{r\to 0}\int f\phi_r$. Thus, we lose nothing by thinking of f as a linear map from $L^q(\mathbb{R}^n)$ to $\mathbb{R}$ rather than as a map from $\mathbb{R}^n$ to $\mathbb{R}$.

Let us modify this idea by allowing f to be merely locally integrable on $\mathbb{R}^n$ but requiring ϕ to lie in C_c^∞. Again the map $\phi \mapsto \int f\phi$ is a well-defined linear functional on C_c^∞, and again the pointwise values of f can be recovered a.e. from it, by an easy extension of Theorem 18. But there are many linear functionals on C_c^∞ that are not of the form $\phi \mapsto \int f\phi$, and these—subject to a mild continuity condition to be specified below—will be our "generalized functions."

Recall that for $E \subset \mathbb{R}^n$ we have defined $C_c^\infty(E)$ to be the set of all C^∞ functions whose support is compact and contained in E. If $U \subset \mathbb{R}^n$ is open, $C_c^\infty(U)$ is the union of the spaces $C_c^\infty(K)$ as K ranges over all compact subsets of U. Each of the latter is a Fréchet space with the topology defined by the norms

$$\phi \mapsto \|\partial^\alpha \phi\|_u \quad (\alpha \in \{0,1,2,\ldots\}^n)$$

in which a sequence $\{\phi_j\}$ converges to ϕ if and only if $\partial^\alpha \phi_j \to \partial^\alpha \phi$ uniformly for all α. (The completeness of $C_c^\infty(K)$ is easily proved by the argument in Folland Exercise 5.9.) With this in mind, we make the following definitions, in which U is an open subset of $\mathbb{R}^n$:

i. A sequence $\{\phi_j\}$ in $C_c^\infty(U)$ converges in C_c^∞ to ϕ if $\{\phi_j\} \subset C_c^\infty(K)$ for some compact set $K \subset U$ and $\phi_j \to \phi$ in the topology of $C_c^\infty(K)$, that is, $\partial^\alpha \phi_j \to \partial^\alpha \phi$ uniformly for all α.

ii. If X is a locally convex topological vector space and $T\colon C_c^\infty(U) \to X$ is a linear map, T is continuous if $T|C_c^\infty(K)$ is continuous for each compact $K \subset U$, that is, if $T\phi_j \to T\phi$ whenever $\phi_j \to \phi$ in $C_c^\infty(K)$ and $K \subset U$ is compact.

iii. A linear map $T\colon C_c^\infty(U) \to C_c^\infty(U')$ is continuous if for each compact $K \subset U$ there is a compact $K' \subset U'$ such that $T(C_c^\infty(K)) \subset C_c^\infty(K')$, and T is continuous from $C_c^\infty(K)$ to $C_c^\infty(K')$.

iv. A distribution on U is a continuous linear functional on $C_c^\infty(U)$. The space of all distributions on U is denoted by $\mathcal{D}'(U)$, and we set $\mathcal{D}' = \mathcal{D}'(\mathbb{R}^n)$. We impose the weak* topology on $\mathcal{D}'(U)$, that is, the topology of pointwise convergence on $C_c^\infty(U)$.

Two remarks: First, the standard notation $\mathcal{D}'$ for the space of distributions comes from Schwartz's notation $\mathcal{D}$ for C_c^∞, which is also quite common. Second, there is a locally convex topology on C_c^∞ with respect to which sequential convergence in C_c^∞ is given by (i) and continuity of linear maps $T\colon C_c^\infty \to X$ and $T\colon C_c^\infty \to C_c^\infty$ is given by (ii) and (iii). However, its definition is rather complicated and of little importance for the elementary theory of distributions, so we shall omit it.

Here are some examples of distributions; more will be presented below. - Every $f \in L^1_{\text{loc}}(U)$—that is, every function f on U such that $\int_K |f| < \infty$ for every compact $K \subset U$—defines a distribution on U, namely, the functional $\phi \to \int f\phi$, and two functions define the same distribution precisely when they are equal a.e. - Every Radon measure μ on U defines a distribution by $\phi \mapsto \int \phi d\mu$. - If $x_0 \in U$ and α is a multi-index, the map $\phi \mapsto \partial^\alpha \phi(x_0)$ is a distribution that does not arise from a function; it arises from a measure μ precisely when $\alpha = 0$, in which case μ is the point mass at x_0.

If $f \in L^1_{\text{loc}}(U)$, we denote the distribution $\phi \mapsto \int f\phi$ also by f, thereby identifying $L^1_{\text{loc}}(U)$ with a subspace of $\mathcal{D}'(U)$. In order to avoid notational confusion between $f(x)$ and $f(\phi) = \int f\phi$, we adopt a different notation for the pairing between $C_c^\infty(U)$ and $\mathcal{D}'(U)$. Namely, if $F \in \mathcal{D}'(U)$ and $\phi \in C_c^\infty(U)$, the value of F at ϕ will be denoted by $\langle F, \phi \rangle$. Observe that the pairing $\langle \cdot, \cdot \rangle$ between $\mathcal{D}'(U)$ and $C_c^\infty(U)$ is linear in each variable; this conflicts with our earlier notation for inner products but will cause no serious confusion. If μ is a measure, we shall also identify μ with the distribution $\phi \mapsto \int \phi d\mu$

Sometimes it is convenient to pretend that a distribution F is a function even when it really is not, and to write $\int F(x)\phi(x)dx$ instead of $\langle F, \phi \rangle$. This is the case especially when the explicit presence of the variable x is notationally helpful.

At this point we set forth two pieces of notation that will be used consistently throughout this chapter. First, we shall use a tilde to denote the reflection of a function in the origin:

$$\tilde{\phi}(x) = \phi(-x).$$

Second, we denote the point mass at the origin, which plays a central role in distribution theory, by δ:

$$\langle \delta, \phi \rangle = \phi(0)$$

As an illustration of the role of δ and the notion of convergence in $\mathcal{D}'$, we record the following important corollary of Theorem 17:

Proposition 9.1: 9.1.

Suppose that $f \in L^1(\mathbb{R}^n)$ and $\int f = a$, and for $t > 0$ let $f_t(x) = t^{-n} f(t^{-1}x)$. Then $f_t \to a\delta$ in $\mathcal{D}'$ as $t \to 0$.

Proof. If $\phi \in C_c^\infty$, by Theorem 17 we have

$$\langle f_t, \phi \rangle = \int f_t \phi = f_t * \tilde{\phi}(0) \to a\tilde{\phi}(0) = a\phi(0) = a\langle \delta, \phi \rangle$$

Although it does not make sense to say that two distributions F and G in $\mathcal{D}'(U)$ agree at a single point, it does make sense to say that they agree on an open set $V \subset U$; namely, $F = G$ on V if and only if $\langle F, \phi \rangle = \langle G, \phi \rangle$ for all $\phi \in C_c^\infty(V)$. (Clearly, if F and G are continuous functions, this condition is equivalent to the pointwise equality of F and G on V; if F and G are merely locally integrable, it means that $F = G$ a.e. on V.) Since a function in $C_c^\infty(V_1 \cup V_2)$ need not be supported in either V_1 or V_2, it is not immediately obvious that if $F = G$ on V_1 and on V_2 then $F = G$ on $V_1 \cup V_2$. However, it is true:

Proposition 9.2: 9.2.

Let $\{V_\alpha\}$ be a collection of open subsets of U and let $V = \bigcup_\alpha V_\alpha$. If $F, G \in \mathcal{D}'(U)$ and $F = G$ on each V_α, then $F = G$ on V.

Proof. If $\phi \in C_c^\infty(V)$, there exist $\alpha_1, \ldots \alpha_m$ such that supp $\phi \subset \bigcup_1^m V_{\alpha_j}$. Pick $\psi_1, \ldots, \psi_m \in C_c^\infty$ such that $\text{supp}(\psi_j) \subset V_{\alpha_j}$ and $\sum_1^m \psi_j = 1$ on $\text{supp}(\phi)$. (That this can be done is the C^∞ analogue of Proposition 125, proved in the same way as that result by using the C^∞ Urysohn lemma.) Then $\langle F, \phi \rangle = \sum \langle F, \psi_j \phi \rangle = \sum \langle G, \psi_j \phi \rangle = \langle G, \phi \rangle$.

According to Proposition 2, if $F \in \mathcal{D}'(U)$, there is a maximal open subset of U on which $F = 0$, namely the union of all the open subsets on which $F = 0$. Its complement in U is called the support of F.

There is a general procedure for extending various linear operations from functions to distributions. Suppose that U and V are open sets in $\mathbb{R}^n$, and T is a linear map from some subspace X of $L^1_{\text{loc}}(U)$ into $L^1_{\text{loc}}(V)$. Suppose that there is another linear map $T' : C_c^\infty(V) \to C_c^\infty(U)$ such that

$$\int (Tf)\phi = \int f(T'\phi) \quad (f \in X, \phi \in C_c^\infty(V))$$

Suppose also that T' is continuous in the sense defined above. Then T can be extended to a map from $\mathcal{D}'(U)$ to $\mathcal{D}'(V)$, still denoted by T, by

$$\langle TF, \phi \rangle = \langle F, T'\phi \rangle \quad (F \in \mathcal{D}'(U), \phi \in C_c^\infty(V))$$

The intervention of the continuous map T' guarantees that the original T, as well as its extension to distributions, is continuous with respect to the weak* topology on distributions: If $F_\alpha \to F \in \mathcal{D}'(U)$, then $TF_\alpha \to TF$ in $\mathcal{D}'(V)$.

Here are the most important instances of this procedure. In each of them, U is an open set in $\mathbb{R}^n$, and the continuity of T' is an easy exercise that we leave to the reader.

i. (Differentiation) Let $Tf = \partial^\alpha f$, defined on $C^{|\alpha|}(U)$. If $\phi \in C_c^\infty(U)$, integration by parts gives $\int(\partial^\alpha f)\phi = (-1)^{|\alpha|} \int f(\partial^\alpha \phi)$; there are no boundary terms since ϕ has compact support. Hence $T' = (-1)^{|\alpha|} T | C_c^\infty(U)$, and we can define the derivative $\partial^\alpha F \in \mathcal{D}'(U)$ of any $F \in \mathcal{D}'(U)$ by

$$\langle \partial^\alpha F, \phi \rangle = (-1)^{|\alpha|} \langle F, \partial^\alpha \phi \rangle.$$

Notice, in particular, that by this procedure we can define derivatives of arbitrary locally integrable functions even when they are not differentiable in the classical sense; this is one of the main reasons for the power of distribution theory. We shall discuss this matter in more detail below. ii. (Multiplication by Smooth Functions) Given $\psi \in C^\infty(U)$, define $Tf = \psi f$. Then $T' = T | C_c^\infty(U)$, so we can define the product $\psi F \in \mathcal{D}'(U)$ for $F \in \mathcal{D}'(U)$ by

$$\langle \psi F, \phi \rangle = \langle F, \psi \phi \rangle$$

Moreover, if $\psi \in C_c^\infty(U)$, this formula makes sense for any $\phi \in C_c^\infty(\mathbb{R}^n)$ and defines ψF as a distribution on $\mathbb{R}^n$.

iii. (Translation) Given $y \in \mathbb{R}^n$, let $V = U + y = \{x + y \mid x \in U\}$ and let $T = \tau_y$. (Recall that we have defined $\tau_y f(x) = f(x-y)$.) Since $\int f(x-y)\phi(x)dx = \int f(x)\phi(x+y)dx$, we have $T' = \tau_{-y} | C_c^\infty(U + y)$. For $F \in \mathcal{D}'(U)$, then, we define the translated distribution $\tau_y F \in \mathcal{D}'(U + y)$ by

$$\langle \tau_y F, \phi \rangle = \langle F, \tau_{-y} \phi \rangle$$

For example, the point mass at y is $\tau_y \delta$.

iv. (Composition with Linear Maps) Given an invertible linear transformation S of $\mathbb{R}^n$, let $V = S^{-1}(U)$ and let $Tf = f \circ S$. Then $T'\phi = |\det S|^{-1} \phi \circ S^{-1}$ by Theorem 87, so for $F \in \mathcal{D}'(U)$ we define $F \circ S \in \mathcal{D}'(S^{-1}(U))$ by

$$\langle F \circ S, \phi \rangle = |\det S|^{-1} \langle F, \phi \circ S^{-1} \rangle.$$

In particular, for $Sx = -x$ we have $f \circ S = \tilde{f}, S^{-1} = S$, and $|\det S| = 1$, so we define the reflection of a distribution in the origin by

$$\langle \widetilde{F}, \phi \rangle = \langle F, \widetilde{\phi} \rangle$$

v. (Convolution, First Method) Given $\psi \in C_c^\infty$, let

$$V = \{x \mid x - y \in U \text{ for } y \in \operatorname{supp}(\psi)\}$$

(V is open but may be empty.) If $f \in L^1_{\text{loc}}(U)$, the integral

$$f * \psi(x) = \int f(x-y)\psi(y)dy = \int f(y)\psi(x-y)dy = \int f\left(\tau_x \widetilde{\psi}\right)$$

is well defined for all $x \in V$. The same definition works for $F \in \mathcal{D}'(U)$: the convolution $F * \psi$ is the function defined on V by

$$F * \psi(x) = \left\langle F, \tau_x \widetilde{\psi} \right\rangle$$

Since $\tau_x \widetilde{\psi} \to \tau_{x_0} \widetilde{\psi}$ in C_c^∞ as $x \to x_0, F * \psi$ is a continuous function (actually C^∞, as we shall soon see) on V. As an example, for any $\psi \in C_c^\infty$ we have

$$\delta * \psi(x) = \left\langle \delta, \tau_x \widetilde{\psi} \right\rangle = \tau_x \widetilde{\psi}(0) = \psi(x)$$

so δ is the multiplicative identity for convolution. vi. (Convolution, Second Method) Let $\psi, \tilde{\psi}$, and V be as in (v). If $f \in L^1_{\text{loc}}(U)$ and $\phi \in C_c^\infty(V)$, we have

$$\int (f * \psi)\phi = \iint f(y)\psi(x-y)\phi(y)dydx = \int f(\phi * \widetilde{\psi})$$

That is, if $Tf = f * \psi$, then T maps $L^1_{\text{loc}}(U)$ into $L^1_{\text{loc}}(V)$ and $T'\phi = \phi * \widetilde{\psi}$. For $F \in \mathcal{D}'(U)$, we can therefore define $F * \psi$ as a distribution on V by

$$\langle F * \psi, \phi \rangle = \langle F, \phi * \widetilde{\psi} \rangle$$

Again, we have $\delta * \psi = \psi$, for

$$\langle \delta * \psi, \phi \rangle = \langle \delta, \phi * \widetilde{\psi} \rangle = \phi * \widetilde{\psi}(0) = \int \phi(x)\psi(x)dx = \langle \psi, \phi \rangle$$

The definitions of convolution in (v) and (vi) are actually equivalent, as we shall now show.

Proposition 9.3: 9.3.

Suppose that U is open in $\mathbb{R}^n$ and $\psi \in C_c^\infty$. Let $V = \{x \mid x - y \in U \text{ for } y \in \operatorname{supp}(\psi)\}$. For $F \in \mathcal{D}'(U)$ and $x \in V$ let $F * \psi(x) = \left\langle F, \tau_x \widetilde{\psi} \right\rangle$. Then

(a) $F * \psi \in C^\infty(V)$.
(b) $\partial^\alpha (F * \psi) = (\partial^\alpha F) * \psi = F * (\partial^\alpha \psi)$.
(c) For any $\phi \in C_c^\infty(V), \int (F * \psi)\phi = \langle F, \phi * \widetilde{\psi} \rangle$.

Proof. Let $e_1, \ldots, e_n$ be the standard basis for $\mathbb{R}^n$. If $x \in V$, there exists $t_0 > 0$ such that $x + te_j \in U$ for $|t| < t_0$, and it is easily verified that

$$t^{-1}\left(\tau_{x+te_j}\widetilde{\psi} - \tau_x\widetilde{\psi}\right) \to \tau_x\widetilde{\partial_j\psi} \text{ in } C_c^\infty(U) \text{ as } t \to 0$$

It follows that $\partial_j(F * \psi)(x)$ exists and equals $F * \partial_j\psi(x)$, so by induction, $F * \psi \in C^\infty(V)$ and $\partial^\alpha(F * \psi) = F * \partial^\alpha\psi$. Moreover, since $\partial^\alpha\widetilde{\psi} = (-1)^{|\alpha|}\widetilde{\partial^\alpha\psi}$ and $\partial^\alpha\tau_x = \tau_x\partial^\alpha$, we have

$$(\partial^\alpha F) * \psi(x) = \left\langle \partial^\alpha F, \tau_x\widetilde{\psi}\right\rangle = (-1)^{|\alpha|}\left\langle F, \partial^\alpha\tau_x\widetilde{\psi}\right\rangle = \left\langle F, \tau_x\widetilde{\partial^\alpha\psi}\right\rangle = F * (\partial^\alpha\psi)(x).$$

Next, if $\phi \in C_c^\infty(V)$, we have

$$\phi * \widetilde{\psi}(x) = \int \phi(y)\psi(y-x)dy = \int \phi(y)\tau_y\widetilde{\psi}(x)dy$$

The integrand here is continuous and supported in a compact subset of U, so the integral can be approximated by Riemann sums. That is, for each (large) $m \in \mathbb{Z}_{\geqslant 1}$ we can approximate $\operatorname{supp}(\phi)$ by a union of cubes of side length 2^{-m} (and volume 2^{-nm}) centered at points $y_1^m, \ldots, y_{k(m)}^m \in \operatorname{supp}(\phi)$; then the corresponding Riemann sums $S^m = 2^{-nm}\sum_j \phi(y_j^m)\tau_{y_j^m}\widetilde{\psi}$ are supported in a common compact subset of U and converge uniformly to $\phi*\widetilde{\psi}$ as $m \to \infty$. Likewise, $\partial^\alpha S^m = 2^{-nm}\sum_j \phi(y_j^m)\tau_{y_j^m}\partial^\alpha\widetilde{\psi}$ converges uniformly to $\phi * \partial^\alpha\widetilde{\psi} = \partial^\alpha(\phi * \widetilde{\psi})$, so $S^m \to \phi * \widetilde{\psi}$ in $C_c^\infty(U)$. Hence,

$$\begin{aligned}\langle F, \phi * \widetilde{\psi}\rangle &= \lim_{m\to\infty}\langle F, S^m\rangle = \lim_{m\to\infty} 2^{-nm}\sum_j \phi(y_j^m)\left\langle F, \tau_{y_j^m}\widetilde{\psi}\right\rangle \\ &= \int \phi(y)\left\langle F, \tau_y\widetilde{\psi}\right\rangle dy = \int \phi(y) F * \psi(y)dy\end{aligned}$$

Next we show that although distributions may be highly singular objects, they can all be approximated in the (weak*) topology of distributions by smooth functions, even by compactly supported ones.

Lemma 9.4: 9.4.

Suppose that $\phi \in C_c^\infty, \psi \in C_c^\infty$, and $\int\psi = 1$, and let $\psi_t(x) = t^{-n}\psi(t^{-1}x)$.

(a) Given any neighborhood U of $\operatorname{supp}(\phi)$, we have $\operatorname{supp}(\phi * \psi_t) \subset U$ for t sufficiently small.

(b) $\phi * \psi_t \to \phi$ in C_c^∞ as $t \to 0$.

Proof. If $\operatorname{supp}(\psi) \subset \{x \mid |x| \leqslant R\}$ then $\operatorname{supp}(\phi * \psi_t)$ is contained in the set of points whose distance from $\operatorname{supp}(\phi)$ is at most tR; this is included in a fixed compact set if $t \leqslant 1$ and is included in U if t is small. Moreover, $\partial^\alpha(\phi * \psi_t) = (\partial^\alpha\phi) * \psi_t \to \partial^\alpha\phi$ uniformly as $t \to 0$, by Theorem 17. The result follows.

Proposition 9.5: 9.5.

For any open $U \subset \mathbb{R}^n, C_c^\infty(U)$ is dense in $\mathcal{D}'(U)$ in the topology of $\mathcal{D}'(U)$.

Proof. Suppose $F \in \mathcal{D}'(U)$. We shall first approximate F by distributions supported in compact subsets of U, then approximate the latter by functions in $C_c^\infty(U)$.

Let $\{V_j\}$ be an increasing sequence of precompact open subsets of U whose union is U, as in Proposition 122. For each j, by the C^∞ Urysohn lemma we can pick $\zeta_j \in C_c^\infty(U)$ such that $\zeta_j = 1$ on $\overline{V}_j$. Given $\phi \in C_c^\infty(U)$, for j sufficiently large we have $\operatorname{supp}(\phi) \subset V_j$ and hence $\langle F, \phi\rangle = \langle F, \zeta_j \phi\rangle = \langle \zeta_j F, \phi\rangle$. Therefore $\zeta_j F \to F$ as $j \to \infty$.

Now, as we noted in defining products of smooth functions and distributions, since $\operatorname{supp}(\zeta_j)$ is compact, $\zeta_j F$ can be regarded as a distribution on $\mathbb{R}^n$. Let ψ, ψ_t be as in Lemma 4, and $\widetilde{\psi}(x) = \psi(-x)$. Then $\int \widetilde{\psi} = 1$ also, so given $\phi \in C_c^\infty$, we have $\phi * \widetilde{\psi}_t \to \phi$ in C_c^∞ by Lemma 4. But then by Proposition 3, we have $(\zeta_j F) * \psi_t \in C^\infty$ and $\langle (\zeta_j F) * \psi_t, \phi\rangle = \left\langle \zeta_j F, \phi * \widetilde{\psi}_t \right\rangle \to \langle \zeta_j F, \phi\rangle$, so $(\zeta_j F) * \psi_t \to \zeta_j F$ in $\mathcal{D}'$. In short, every neighborhood of F in $\mathcal{D}'(U)$ contains the C^∞ functions $(\zeta_j F) * \psi_t$ for j large and t small.

Finally, we observe that $\operatorname{supp}(\zeta_j) \subset V_k$ for some k. If $\operatorname{supp}(\phi) \cap \overline{V}_k = \varnothing$, then for sufficiently small t we have $\operatorname{supp}\left(\phi * \widetilde{\psi}_t\right) \cap \overline{V}_k = \varnothing$ (Lemma 4 again) and hence $\langle (\zeta_j F) * \psi_t, \phi\rangle = \left\langle F, \zeta_j \left(\phi * \widetilde{\psi}_t\right)\right\rangle = 0$. In other words, $\operatorname{supp}((\zeta_j F) * \psi_t) \subset \overline{V}_k \subset U$, so we are done.

We conclude this section with some further remarks and examples concerning differentiation of distributions. To restate the basic facts: Every $F \in \mathcal{D}'(U)$ possesses derivatives $\partial^\alpha F \in \mathcal{D}'(U)$ of all orders; moreover, ∂^α is a continuous linear map of $\mathcal{D}'(U)$ into itself. Let us examine a couple of one-dimensional examples to see what sort of things arise by taking distribution derivatives of functions that are not classically differentiable.

First, differentiating functions with jump discontinuities leads to "delta-functions," that is, distributions given by measures that are point masses. The simplest example is the Heaviside step function $H = \chi_{(0,\infty)}$, for which we have

$$\langle H', \phi\rangle = -\langle H, \phi'\rangle = -\int_0^\infty \phi'(x)dx = \phi(0) = \langle \delta, \phi\rangle$$

so $H' = \delta$. See Exercises 5 and 7 for generalizations.

Second, distribution derivatives can be used to extract "finite parts" from divergent integrals. For example, let $f(x) = x^{-1}\chi_{(0,\infty)}(x).f$ is locally integrable on $\mathbb{R}\backslash\{0\}$ and so defines a distribution there, but $\int f\phi$ diverges whenever $\phi(0) \neq 0$. Nonetheless, there is a distribution on $\mathbb{R}$ that agrees with f on $\mathbb{R}\backslash\{0\}$, namely, the distribution derivative of the locally integrable function $L(x) = (\log x)\chi_{(0,\infty)}(x)$. One way of seeing what is going on here is to consider the functions $L_\varepsilon(x) = (\log x)\chi_{(\varepsilon,\infty)}(x)$. By the dominated convergence theorem we have $\int L\phi = \lim_{\varepsilon\to 0}\int L_\varepsilon \phi$ for any $\phi \in C_c^\infty$, that is, $L_\varepsilon \to L$ in $\mathcal{D}'$; it follows that $L_\varepsilon' \to L'$ in $\mathcal{D}'$. But

$$\langle L_\varepsilon', \phi\rangle = -\langle L_\varepsilon, \phi'\rangle = -\int_\varepsilon^\infty \phi'(x)\log x dx = \int_\varepsilon^\infty \frac{\phi(x)}{x}dx + \phi(\varepsilon)\log\varepsilon$$

As $\varepsilon \to 0$, this last sum converges even though the two terms individually do not. Formally, passage to the limit gives $\langle L', \phi\rangle = \int f\phi + (\log 0)\phi(0)$; that is, L' is obtained from f by subtracting an infinite multiple of δ. (This process is akin to the "renormalizations" used by physicists to remove the divergences from quantum field theory.)

Another way to analyze this situation is to consider smooth approximations to L, such as $L^\varepsilon(x) = L(x)\psi(\varepsilon x)$ where ψ is a smooth function such that $\psi(x) = 0$ for $x \leqslant 1$ and $\psi(x) = 1$ for $x \geqslant 2$. The reader is invited to sketch the graphs of L^ε and $(L^\varepsilon)'$; the latter will look like the graph of f together with a large negative spike near the origin, which turns into " $-\infty \cdot \delta$" as $\varepsilon \to 0$. See also Folland Exercise 9.10,Folland Exercise 9.12.

Finally, we remark that one of the bugbears of advanced calculus, that equality of mixed partials need not hold for functions whose derivatives are not continuous, disappears in the setting of distributions: $\partial_j\partial_k = \partial_k\partial_j$ on C_c^∞; therefore $\partial_j\partial_k = \partial_k\partial_j$ on $\mathcal{D}'$! In the standard counterexample, $f(x,y) = xy(x^2 - y^2)(x^2+y^2)^{-1}$ (with $f(0,0) = 0$), $\partial_x\partial_y f$ and $\partial_y\partial_x f$ are locally integrable functions that agree everywhere except at the origin; hence they are identical as distributions.

Exercise 9.6: Folland Exercise 9.1.

Suppose that $f_1, f_2, \ldots,$ and f are in $L^1_{\mathrm{loc}}(U)$. The conditions in (a) and (b) below imply that $f_n \to f$ in $\mathcal{D}'(U)$, but the condition in (c) does not.

(a) $f_n \in L^p(U) (1 \leqslant p \leqslant \infty)$ and $f_n \to f$ in the L^p norm or weakly in L^p.
(b) For all $n, |f_n| \leqslant g$ for some $g \in L^1_{\mathrm{loc}}(U)$, and $f_n \to f$ a.e.
(c) $f_n \to f$ pointwise.

Exercise 9.7: Folland Exercise 9.2.

The product rule for derivatives is valid for products of smooth functions and distributions.

Exercise 9.8: Folland Exercise 9.3.

On $\mathbb{R}$, if $\psi \in C^\infty$ then $\psi\delta^{(k)} = \sum_0^k (-1)^j \binom{k}{j} \psi^{(j)}(0)\delta^{(k-j)}$, where the superscripts denote derivatives.

Exercise 9.9: Folland Exercise 9.4.

Suppose that U and V are open in $\mathbb{R}^n$ and $\Phi \colon V \to U$ is a C^∞ diffeomorphism. Explain how to define $F \circ \Phi \in \mathcal{D}'(U)$ for any $F \in \mathcal{D}'(V)$.

Exercise 9.10: Folland Exercise 9.5.

Suppose that f is continuously differentiable on $\mathbb{R}$ except at $x_1, \ldots, x_m$, where f has jump discontinuities, and that its pointwise derivative df/dx (defined except at the x_js) is in $L^1_{\text{loc}}(\mathbb{R})$. Then the distribution derivative f' of f is given by $f' = (df/dx) + \sum_1^m [f(x_j+) - f(x_j-)]\tau_{x_j}\delta$.

Exercise 9.11: Folland Exercise 9.6.

If f is absolutely continuous on compact subsets of an interval $U \subset \mathbb{R}$, the distribution derivative $f' \in \mathcal{D}'(U)$ coincides with the pointwise (a.e.-defined) derivative of f.

Exercise 9.12: Folland Exercise 9.7.

Suppose $f \in L^1_{\text{loc}}(\mathbb{R})$. Then the distribution derivative f' is a complex measure on $\mathbb{R}$ if and only if f agrees a.e. with a function $F \in NBV$, in which case $\langle f', \phi \rangle = \int \phi dF$.

Exercise 9.13: Folland Exercise 9.8.

Suppose $f \in L^p(\mathbb{R}^n)$. If the strong L^p derivatives $\partial_j f$ exist in the sense of Folland Exercise 8.8, , they coincide with the partial derivatives of f in the sense of distributions.

Exercise 9.14: Folland Exercise 9.9.

A distribution F on $\mathbb{R}^n$ is called homogeneous of degree λ if $F \circ S_r = r^\lambda F$ for all $r > 0$, where $S_r(x) = rx$.

(a) δ is homogeneous of degree $-n$.
(b) If F is homogeneous of degree λ, then $\partial^\alpha F$ is homogeneous of degree $\lambda - |\alpha|$.
(c) The distribution $(d/dx)[\chi_{(0,\infty)}(x) \log x]$ discussed in the text is not homogeneous, although it agrees on $\mathbb{R}\backslash\{0\}$ with a function that is homogeneous of degree -1.

Exercise 9.15: Folland Exercise 9.10.

Let f be a continuous function on $\mathbb{R}^n\backslash\{0\}$ that is homogeneous of degree $-n$ (i.e., $f(rx) = r^{-n}f(x)$) and has mean zero on the unit sphere (i.e., $\int f d\sigma = 0$ where σ is surface measure on the sphere). Then f is not locally integrable near the origin (unless $f = 0$), but the formula

$$\langle PV(f), \phi \rangle = \lim_{\varepsilon \to 0} \int_{|x| > \varepsilon} f(x)\phi(x)dx \quad (\phi \in C_c^\infty)$$

defines a distribution $PV(f)$– " PV" stands for "principal value"—that agrees with f on $\mathbb{R}^n\backslash\{0\}$ and is homogeneous of degree $-n$ in the sense of Folland Exercise 9.9.

(Hint: For any $a > 0$, the indicated limit equals

$$\int_{|x|\leqslant a} f(x)[\phi(x) - \phi(0)]dx + \int_{|x|>a} f(x)\phi(x)dx$$

and these integrals converge absolutely.)

Exercise 9.16: Folland Exercise 9.11.

Let F be a distribution on $\mathbb{R}^n$ such that $\operatorname{supp}(F) = \{0\}$.

(a) There exist $N \in \mathbb{Z}_{\geqslant 1}, C > 0$ such that for all $\phi \in C_c^\infty$,

$$|\langle F, \phi\rangle| \leqslant C \sum_{|\alpha|\leqslant N} \sup_{|x|\leqslant 1}|\partial^\alpha \phi(x)|$$

(b) Fix $\psi \in C_c^\infty$ with $\psi(x) = 1$ for $|x| \leqslant 1$ and $\psi(x) = 0$ for $|x| \geqslant 2$. If $\phi \in C_c^\infty$, let $\phi_k(x) = \phi(x)[1 - \psi(kx)]$. If $\partial^\alpha\phi(0) = 0$ for $|\alpha| \leqslant N$, then $\partial^\alpha\phi_k \to \partial^\alpha\phi$ uniformly as $k \to \infty$ for $|\alpha| \leqslant N$. (Hint: By Taylor's theorem, $|\partial^\alpha\phi(x)| \leqslant C|x|^{N+1-|\alpha|}$ for $|\alpha| \leqslant N$.)

(c) If $\phi \in C_c^\infty$ and $\partial^\alpha\phi(0) = 0$ for $|\alpha| \leqslant N$, then $\langle F, \phi\rangle = 0$.

(d) There exist constants $c_\alpha(|\alpha| \leqslant N)$ such that $F = \sum_{|\alpha|\leqslant N} c_\alpha \partial^\alpha \delta$.

Exercise 9.17: Folland Exercise 9.12.

Suppose $\lambda > n$; then the function $x \mapsto |x|^{-\lambda}$ on $\mathbb{R}^n$ is not locally integrable near the origin. Here are some ways to make it into a distribution:

(a) If $\phi \in C_c^\infty$, let P_ϕ^k be the Taylor polynomial of ϕ about $x = 0$ of degree k. Given $k > \lambda - n - 1$ and $a > 0$, define

$$\langle F_a^k, \phi\rangle = \int_{|x|\leqslant a} [\phi(x) - P_\phi^k(x)]|x|^{-\lambda}dx + \int_{|x|>a} \phi(x)|x|^{-\lambda}dx$$

Then F_a^k is a distribution on $\mathbb{R}^n$ that agrees with $|x|^{-\lambda}$ on $\mathbb{R}^n\backslash\{0\}$.

(b) If $\lambda \notin \mathbb{Z}$ and we take k to be the greatest integer $\leqslant \lambda - n$, we can let $a \to \infty$ in (a) to obtain another distribution F that agrees with $|x|^{-\lambda}$ on $\mathbb{Z}^n\backslash\{0\}$:

$$\langle F, \phi\rangle = \int [\phi(x) - P_\phi^k(x)]|x|^{-\lambda}dx$$

(c) Let $n = 1$ and let k be the greatest integer $\leqslant \lambda$. Let

$$f(x) = \begin{cases} [(k-\lambda)\cdots(1-\lambda)]^{-1}(\operatorname{sgn} x)^k|x|^{k-\lambda} & \text{if } \lambda > k \\ (-1)^{k-1}[(k-1)!]^{-1}(\operatorname{sgn} x)^k \log|x| & \text{if } \lambda = k \end{cases}$$

Then $f \in L^1_{\text{loc}}(\mathbb{R})$, and the distribution derivative $f^{(k)}$ agrees with $|x|^{-\lambda}$ on $\mathbb{R}\backslash\{0\}$.

(d) According to Folland Exercise 9.11, the difference between any two of the distributions constructed in (a)-(c) is a linear combination of δ and its derivatives. Which one?

Exercise 9.18: Folland Exercise 9.13.

If $F \in \mathcal{D}'$ and $\partial_j F = 0$ for $j = 1, \ldots, n$, then F is a constant function. (Consider $f * \psi_t$ where ψ_t is an approximate identity in C_c^∞.)

Exercise 9.19: Folland Exercise 9.14.

For $n \geqslant 3$, define $F, F^\varepsilon \in L^1_{\text{loc}}(\mathbb{R}^n)$ by

$$F(x) = \frac{|x|^{2-n}}{\omega_n(2-n)}, \quad F^\varepsilon(x) = \frac{(|x|^2 + \varepsilon^2)^{(2-n)/2}}{\omega_n(2-n)}$$

where $\omega_n = 2\pi^{n/2}/\Gamma(n/2)$ is the volume of the unit sphere, and let Δ be the Laplacian.

(a) $\Delta F^\varepsilon(x) = \varepsilon^{-n} g(\varepsilon^{-1}x)$ where $g(x) = n\omega_n^{-1}(|x|^2+1)^{-(n+2)/2}$.

(b) $\int g = 1$. (Use polar coordinates and set $s = r^2/(r^2+1)$.)

(c) $\Delta F = \delta$.($F^\varepsilon \to F$ in $\mathcal{D}'$; use Proposition 1.)

(d) If $\phi \in C_c^\infty$, the function $f = F * \phi$ satisfies $\Delta f = \phi$.

(e) The results of (c) and (d) hold also for $n = 1$ but can be proved more simply there. For $n = 2$, they hold provided F, F^ε are defined by $F(x) = (2\pi)^{-1}\log|x|$ and $F^\varepsilon = (4\pi)^{-1}\log(|x|^2+\varepsilon^2)$.

Exercise 9.20: Folland Exercise 9.15.

Define G on $\mathbb{R}^n \times \mathbb{R}$ by $G(x,t) = (4\pi t)^{-n/2} e^{-|x|^2/4t}\chi_{(0,\infty)}(t)$.

(a) $(\partial_t - \Delta)G = \delta$, where Δ is the Laplacian on $\mathbb{R}^n$. (Let $G^\varepsilon(x,t) = G(x,t)\chi_{(\varepsilon,\infty)}(t)$; then $G^\varepsilon \to G$ in $\mathcal{D}'$. Compute $\langle(\partial_t - \Delta)G^\varepsilon, \phi\rangle$ for $\phi \in C_c^\infty$, recalling the discussion of the heat equation in §8.7.)

(b) If $\phi \in C_c^\infty(\mathbb{R}^n \times \mathbb{R})$, the function $f = G * \phi$ satisfies $(\partial_t - \Delta)f = \phi$.

Solution.

(a) Let $\varepsilon > 0$ and $\phi \in C_c^\infty(\mathbb{R}^n \times \mathbb{R})$. Then $G^\varepsilon \in \mathcal{D}'(\mathbb{R}^n \times \mathbb{R})$, and by Fubini's theorem

$$\langle \phi, G^\varepsilon \rangle = \int_{\mathbb{R}^n \times \mathbb{R}} G^\varepsilon \phi = \int_\varepsilon^\infty \int_{\mathbb{R}^n} G(x,t)\phi(x,t)\,\mathrm{d}x\,\mathrm{d}t$$
$$\xrightarrow{\varepsilon \to 0} \int_0^\infty \int_{\mathbb{R}^n} G(x,t)\phi(x,t)\,\mathrm{d}x\,\mathrm{d}t = \langle \phi, G \rangle,$$

so $G^\varepsilon \to G$ in $\mathcal{D}'(\mathbb{R}^n \times \mathbb{R})$ as $\varepsilon \to 0$. For $x \in \mathbb{R}^n$ and $t \in (\varepsilon, \infty)$, it follows from [Fol99, §8.7, Exercise 44, assertion (ii)] that $(\Delta - \partial_t)G^\varepsilon(x,t) = 0$; thus

$$\Delta G^\varepsilon(x,t) = \partial_t G^\varepsilon(x,t). \tag{9.20.1}$$

In addition (see [Fol99, p. 284]), we have

$$\partial_t \phi(x,t) G^\varepsilon(x,t) = -\phi(x,t)\partial_t G^\varepsilon(x,t), \tag{9.20.2}$$

so

$$\begin{aligned}
\langle \phi, (\partial_t - \Delta)G^\varepsilon \rangle &= \langle (\partial_t + \Delta)\phi, G^\varepsilon \rangle \\
&= \int_{\mathbb{R}^n} \left(\int_\varepsilon^\infty (\phi(x,t)\partial_t G^\varepsilon(x,t) + \Delta\phi(x,t) G^\varepsilon(x,t))\, \mathrm{d}t \right. \\
&\qquad \left. + \int_0^\varepsilon (\phi(x,t)\Delta G^\varepsilon(x,t) + \phi(x,t)\Delta G^\varepsilon(x,t))\, \mathrm{d}t \right) \mathrm{d}x \\
&\overset{(9.20.2)}{=} \int_{\mathbb{R}^n} \left(\int_\varepsilon^\infty (\phi(x,t)\Delta G^\varepsilon(x,t) + \Delta\phi(x,t) G^\varepsilon(x,t))\, \mathrm{d}t \right. \\
&\qquad \left. + \int_0^\varepsilon (\phi(x,t)\partial_t G^\varepsilon(x,t) - \phi(x,t)\Delta G^\varepsilon(x,t))\, \mathrm{d}t \right) \mathrm{d}x \\
&\overset{(9.20.1)}{=} \int_{\mathbb{R}^n} \left(\int_\varepsilon^\infty (\cancel{\phi(x,t)\Delta G^\varepsilon(x,t)} - \cancel{\phi(x,t)\Delta G^\varepsilon(x,t)})\, \mathrm{d}t \right. \\
&\qquad \left. + \int_0^\varepsilon (\phi(x,t)\partial_t G^\varepsilon(x,t) - \phi(x,t)\Delta G^\varepsilon(x,t))\, \mathrm{d}t \right) \mathrm{d}x \\
&= \int_{\mathbb{R}^n} \int_0^\varepsilon \phi(x,t)(\partial_t - \Delta)G^\varepsilon(x,t)\, \mathrm{d}t\, \mathrm{d}x = \int_{\mathbb{R}^n} \phi(x,\varepsilon)G(x,\varepsilon)\, \mathrm{d}x.
\end{aligned}$$

(b) Let $\phi \in C_c^\infty(\mathbb{R}^n \times \mathbb{R})$. Since $G \in \mathcal{D}'(\mathbb{R}^n \times \mathbb{R})$, by [Fol99, Proposition 9.3(b)] we have

$$(\partial_t - \Delta)(G * \phi) = ((\partial_t - \Delta)G) * \phi = \delta * \phi = \phi,$$

where the second equality is by part (a) and the third equality holds since δ is an identity for the convolution product. □

9.2 Compactly Supported, Tempered, and Periodic Distributions

If U is an open set in $\mathbb{R}^n$, the space of all distributions on U whose support is a compact subset of U is denoted by $\mathcal{E}'(U)$; as usual, we set $\mathcal{E}' = \mathcal{E}'(\mathbb{R}^n)$.$\mathcal{E}'(U)$ turns out to be a dual space in its own right, as we shall now show.

The space $C^\infty(U)$ of C^∞ functions on U is a Fréchet space with the C^∞ topology—that is, the topology of uniform convergence of functions, together with all their derivatives, on compact subsets of U. This topology can be defined by a countable family of seminorms as follows. Let $\{V_m\}_1^\infty$ be an increasing sequence of precompact open subsets of U whose union is U, as in Proposition 122; then for each $m \in \mathbb{Z}_{\geqslant 1}$ and each multi-index α we have the seminorm

$$\|f\|_{[m,\alpha]} = \sup_{x \in \overline{V}_m} |\partial^\alpha f(x)|$$

Clearly $\partial^\alpha f_j \to \partial^\alpha f$ uniformly on compact sets for all α if and only if $\|f_j - f\|_{[m,\alpha]} \to 0$ for all m, α; a different choice of sets V_m would yield an equivalent family of seminorms.

Proposition 9.21: 9.7.

$C_c^\infty(U)$ is dense in $C^\infty(U)$.

Proof. Let $\{V_m\}_1^\infty$ be as in (9.6). For each m, by the C^∞ Urysohn lemma we can pick $\psi_m \in C_c^\infty(U)$ with $\psi_m = 1$ on $\overline{V}_m$. If $\phi \in C^\infty(U)$, clearly $\|\psi_m\phi - \phi\|_{[m_0,\alpha]} = 0$ provided $m \geqslant m_0$; thus $\psi_m\phi \to \phi$ in the C^∞ topology.

Theorem 9.22: 9.8.

$\mathcal{E}'(U)$ is the dual space of $C^\infty(U)$. More precisely: If $F \in \mathcal{E}'(U)$, then F extends uniquely to a continuous linear functional on $C^\infty(U)$; and if G is a continuous linear functional on $C^\infty(U)$, then $G|C_c^\infty(U) \in \mathcal{E}'(U)$.

Proof. If $F \in \mathcal{E}'(U)$, choose $\psi \in C_c^\infty(U)$ with $\psi = 1$ on $\mathrm{supp}(F)$, and define the linear functional G on $C^\infty(U)$ by $\langle G, \phi\rangle = \langle F, \psi\phi\rangle$. Since F is continuous on $C_c^\infty(\mathrm{supp}(\psi))$, and the topology of the latter is defined by the norms $\phi \mapsto \|\partial^\alpha\phi\|_u$, by Proposition 87 there exist $N \in \mathbb{Z}_{\geqslant 1}$ and $C > 0$ such that $|\langle G, \phi\rangle| \leqslant C\sum_{|\alpha|\leqslant N}\|\partial^\alpha(\psi\phi)\|_u$ for $\phi \in C^\infty(U)$. By the product rule, if we choose m large enough so that $\mathrm{supp}(\psi) \subset V_m$, this implies that

$$|\langle G, \phi\rangle| \leqslant C'\sum_{|\alpha|\leqslant N}\sup_{x\in\mathrm{supp}(\psi)}|\partial^\alpha\phi(x)| \leqslant C'\sum_{|\alpha|\leqslant N}\|\phi\|_{[m,\alpha]}$$

so that G is continuous on $C^\infty(U)$. That G is the unique continuous extension of F follows from Proposition 21.

On the other hand, if G is a continuous linear functional on $C^\infty(U)$, by Proposition 87 there exist C, m, N such that $|\langle G, \phi\rangle| \leqslant C\sum_{|\alpha|\leqslant N}\|\phi\|_{[m,\alpha]}$ for all $\phi \in C^\infty(U)$. Since $\|\phi\|_{[m,\alpha]} \leqslant \|\partial^\alpha\phi\|_u$, this implies that G is continuous on $C_c^\infty(K)$ for each compact $K \subset U$, so $G|C_c^\infty(U) \in \mathcal{D}'(U)$. Moreover, if $[\mathrm{supp}(\phi)] \cap \overline{V}_m = \varnothing$, then $\langle G, \phi\rangle = 0$; hence $\mathrm{supp}(G) \subset \overline{V}_m$ and $G|C_c^\infty(U) \in \mathcal{D}'(U)$.

The operations of differentiation, multiplication by C^∞ functions, translation, and composition by linear maps discussed in Folland Section 9.1 all preserve the class $\mathcal{E}'$. As for convolution, there is more to be said.

First, if $F \in \mathcal{E}'$ and $\phi \in C_c^\infty$ then $F * \phi \in C_c^\infty$, as Proposition 11d remains valid in this setting. Second, if $F \in \mathcal{E}'$ and $\psi \in C^\infty$, $F * \psi$ can be defined as a C^∞ function or as a distribution just as before:

$$F * \psi(x) = \left\langle F, \tau_x\widetilde{\psi}\right\rangle, \quad \langle F * \psi, \phi\rangle = \langle F, \phi * \widetilde{\psi}\rangle \quad (\phi \in C_c^\infty)$$

(see Folland Exercise 9.16). Finally, a further dualization allows us to define convolutions of arbitrary distributions with compactly supported distributions. To wit, if $F \in \mathcal{D}'$ and $G \in \mathcal{D}'$, we can define $F * G \in \mathcal{D}'$ and $G * F \in \mathcal{D}'$ as follows:

$$\langle F * G, \phi\rangle = \langle F, \widetilde{G} * \phi\rangle, \quad \langle G * F, \phi\rangle = \langle G, \widetilde{F} * \phi\rangle \quad (\phi \in C_c^\infty)$$

and likewise for $\widetilde{F}$. The proof that $F * G$ and $G * F$ are indeed distributions (i.e., that they are continuous on C_c^∞) and that $F * G = G * F$ requires a closer examination of the continuity of the maps involved. We shall not pursue this matter here; however, see Exercises 20 and 21.

A notable omission from our list of operations that can be extended from functions to distributions is the Fourier transform $\mathcal{F}$. The trouble is that $\mathcal{F}$ does not map C_c^∞ into itself; in fact, if $\phi \in C_c^\infty$, then $\widehat{\phi}$ cannot vanish on any nonempty open set unless $\phi = 0$. To see this, suppose $\widehat{\phi} = 0$ on a neighborhood of ξ_0. Replacing ϕ by $e^{-2\pi i \xi_0 \cdot x}\phi$, we may assume that $\xi_0 = 0$. Since ϕ has compact support, we can expand $e^{-2\pi i \xi \cdot x}$ in its Maclaurin series and integrate term by term to obtain

$$\widehat{\phi}(\xi) = \sum_{k=0}^{\infty} \frac{1}{k!} \int (-2\pi i \xi \cdot x)^k \phi(x) dx = \sum_\alpha \frac{1}{\alpha!} \xi^\alpha \int (-2\pi i x)^\alpha \phi(x) dx$$

(see Folland Exercise 8.2(a) in Folland Section 8.1). But $\int (-2\pi i x)^\alpha \phi(x) dx = \partial^\alpha \widehat{\phi}(0)$ for all α by Theorem 31d. These derivatives all vanish by assumption, so $\widehat{\phi} = 0$ and hence $\phi = 0$.

However, we do have available a slightly larger space of smooth functions that is mapped into itself by $\mathcal{F}$, namely, the Schwartz class $\mathcal{S}$. We recall that $\mathcal{F}$ is a Fréchet space with the topology defined by the norms

$$\|\phi\|_{(N,\alpha)} = \sup_{x \in \mathbb{R}^n} (1 + |x|)^N |\partial^\alpha \phi(x)|$$

Proposition 9.23: 9.9.

Suppose $\psi \in C_c^\infty$ and $\psi(0) = 1$, and let $\psi^\varepsilon(x) = \psi(\varepsilon x)$. Then for any $\phi \in \mathcal{S}, \psi^\varepsilon \phi \to \phi$ in $\mathcal{S}$ as $\varepsilon \to 0$. In particular, C_c^∞ is dense in $\mathcal{S}$.

Proof. Given $N \in \mathbb{Z}_{\geqslant 1}$, for any $\eta > 0$ we can choose a compact set K such that $(1 + |x|)^N |\phi(x)| < \eta$ for $x \notin K$. Since $\psi(\varepsilon x) \to 1$ uniformly for $x \in K$ as $\varepsilon \to 0$, it follows easily that $\|\psi^\varepsilon \phi - \phi\|_{(N,0)} \to 0$ for every N. For the norms involving derivatives, we observe that by the product rule,

$$(1 + |x|)^N \partial^\alpha (\psi^\varepsilon \phi - \phi) = (1 + |x|)^N (\psi^\varepsilon \partial^\alpha \phi - \partial^\alpha \phi) + E_\varepsilon(x)$$

where E_ε is a sum of terms involving derivative of ψ^ε. Since

$$|\partial^\beta \psi^\varepsilon(x)| = \varepsilon^{|\beta|} |\partial^\beta \psi(\varepsilon x)| \leqslant C_\beta \varepsilon^{|\beta|}$$

we have $\|E_\varepsilon\|_u \leqslant C\varepsilon \to 0$ as $\varepsilon \to 0$. The preceding argument then shows that $\|\psi^\varepsilon \phi - \phi\|_{(N,\alpha)} \to 0$.

A tempered distribution is a continuous linear functional on $\mathcal{S}$. The space of tempered distributions is denoted by $\mathcal{S}'$; it comes equipped with the weak* topology, that is, the topology of pointwise convergence on $\mathcal{S}$. If $F \in \mathcal{S}'$, then $F|C_c^\infty$ is clearly a distribution, since convergence in C_c^∞ implies convergence in $\mathcal{S}$, and $F|C_c^\infty$ determines F uniquely by Proposition 23. Thus we may, and shall, identify $\mathcal{S}'$ with the set of distributions that extend continuously from C_c^∞ to $\mathcal{S}$. We say that a locally integrable function is tempered if it is tempered as a distribution.

The condition that a distribution be tempered means, roughly speaking, that it does not grow too fast at infinity. Here are a few examples: - Every compactly supported distribution is tempered. - If $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ and $\int(1+|x|)^N|f(x)|dx < \infty$ for some N, then f is tempered, for $\left|\int f\phi\right| \leqslant C\|\phi\|_{(0,N)}$. - The function $f(x) = e^{ax}$ on $\mathbb{R}$ is tempered if and only if a is purely imaginary. Indeed, suppose $a = b + ic$ with b, c real. If $b = 0$, then f is bounded and hence tempered by (ii). If $b \neq 0$, choose a function $\psi \in C_c^\infty$ such that $\int \psi = 1$, and let $\phi_j(x) = e^{-ax}\psi(x-j)$. It is easily verified that $\phi_j \to 0$ in $\mathcal{S}$ as $j \to +\infty$ (if $b > 0$) or $j \to -\infty$ (if $b < 0$), but $\int f\phi_j = \int \psi = 1$ for all j. - On the other hand, the function $f(x) = e^x \cos e^x$ on $\mathbb{R}$ is tempered, because it is the derivative of the bounded function $\sin e^x$. Indeed, if $\phi \in \mathcal{S}$, integration by parts yields

$$\left|\int f\phi\right| = \left|-\int \phi'(x)\sin e^x dx\right| \leqslant C\|\phi\|_{(2,1)}$$

Intuitively, $f(x)$ is not too large "on average" when x is large, because of its rapid oscillations.

We turn to the consideration of the basic linear operations on tempered distributions. The operations of differentiation, translation, and composition with linear transformations work just the same way for tempered distributions as for plain distributions; these operations all map $\mathcal{S}$ and $\mathcal{S}'$ into themselves. The same is not true of multiplication by arbitrary smooth functions, however. The proper requirement on $\psi \in C^\infty$ in order for the map $F \to \psi F$ to preserve $\mathcal{S}$ and $\mathcal{S}'$ is that ψ and all its derivatives should have at most polynomial growth at infinity:

$$|\partial^\alpha \psi(x)| \leqslant C_\alpha(1+|x|)^{N(\alpha)} \text{ for all } \alpha$$

Such C^∞ functions are called slowly increasing. For example, every polynomial is slowly increasing; so are the functions $(1+|x|^2)^s (s \in \mathbb{R})$, which will play an important role in the next section.

As for convolutions, for any $F \in \mathcal{S}'$ and $\psi \in \mathcal{S}$ we can define the convolution $F * \psi$ by $F * \psi(x) = \left\langle F, \tau_x\widetilde{\psi}\right\rangle$, as before, and we have an analogue of Proposition 3:

Proposition 9.24: 9.10.

If $F \in \mathcal{S}'$ and $\psi \in \mathcal{S}$, then $F * \psi$ is a slowly increasing C^∞ function, and for any $\phi \in \mathcal{S}$ we have $\int(F * \psi)\phi = \langle F, \phi * \widetilde{\psi}\rangle$.

Proof. That $F * \psi \in C^\infty$ is established as in Proposition 3. By Proposition 87, the continuity of F implies that there exist m, N, C such that

$$|\langle F, \phi\rangle| \leqslant C\sum_{|\alpha| \leqslant N} \|\phi\|_{(m,\alpha)} \quad (\phi \in \mathcal{S})$$

and hence by (8.12),

$$\begin{aligned}|F * \psi(x)| &\leqslant C \sum_{|\alpha|\leqslant N} \sup_y (1+|y|)^m |\partial^\alpha \psi(x-y)| \\ &\leqslant C(1+|x|)^m \sum_{|\alpha|\leqslant N} \sup_y (1+|x-y|)^m |\partial^\alpha \psi(x-y)| \\ &\leqslant C(1+|x|)^m \sum_{|\alpha|\leqslant N} \|\psi\|_{(m,\alpha)}.\end{aligned}$$

The same reasoning applies with ψ replaced by $\partial^\beta \psi$, so $F * \psi$ is slowly increasing. Next, by Proposition 3 we know that the equation $\int (F * \psi)\phi = \langle F, \phi * \widetilde{\psi}\rangle$ holds when $\phi, \psi \in C_c^\infty$. By Proposition 23, if $\phi, \psi \in \mathcal{S}$ we can find sequences $\{\phi_j\}$ and $\{\psi_j\}$ in C_c^∞ that converge to ϕ and ψ in $\mathcal{S}$. Then $\phi_j * \widetilde{\psi}_j \to \phi * \widetilde{\psi}$ in $\mathcal{S}$ by (the proof of) Proposition 16, so $\left\langle F, \phi_j * \widetilde{\psi}_j \right\rangle \to \langle F, \phi * \widetilde{\psi}\rangle$. On the other hand, the preceding estimates show that $|F * \psi_j(x)| \leqslant C(1+|x|)^m$ with C and m independent of j, and likewise $|\phi_j(x)| \leqslant C(1+|x|)^{-m-n-1}$, so $\int (F * \psi_j)\phi_j \to \int (F * \psi)\phi$ by the dominated convergence theorem.

Finally, we come to the principal raison d'être of tempered distributions, the Fourier transform. We recall (Corollary 32) that the Fourier transform maps $\mathcal{S}$ continuously into itself, and that for $f, g \in L^1$ (in particular, for $f, g \in \mathcal{S}$) we have

$$\int \widehat{f}(y) g(y)\, dy = \iint f(x) g(y) e^{-2\pi i x\cdot y}\, dx\, dy = \int f(x)\widehat{g}(x)\, dx$$

We can therefore extend the Fourier transform to a continuous linear map from $\mathcal{S}'$ to itself by defining

$$\langle \widehat{F}, \phi\rangle = \langle F, \widehat{\phi}\rangle \quad (F \in \mathcal{S}', \phi \in \mathcal{S}).$$

This definition clearly agrees with the one in Chapter 8 when $F \in L^1 + L^2$.

The basic properties of the Fourier transform in Theorem 31 continue to hold in this setting. To wit,

$$\begin{gathered}(\tau_y F) = e^{-2\pi i \xi\cdot y}\widehat{F}, \quad \tau_\eta \widehat{F} = \left[e^{2\pi i \eta\cdot x}\widehat{F}\right], \\ \partial^\alpha \widehat{F} = \left[(-2\pi i x)^\alpha \widehat{F}\right], \quad \left(\partial^\alpha F \widehat{F} = (2\pi i \xi)^\alpha \widehat{F}\right. \\ \left(f \circ T\widehat{)} = |\det T|^{-1} \widehat{f} \circ (T^*)^{-1} \quad (T \in GL(n, \mathbb{R})),\right. \\ (F * \psi)\widehat{\psi} = \widehat{\psi}\widehat{F} \quad (\psi \in \mathcal{S}).\end{gathered}$$

(The first four of these formulas involve products of slowly increasing C^∞ functions, specified by their values at a general point x or ξ, and tempered distributions.) The easy verifications of these facts are left to the reader (Folland Exercise 9.17).

Moreover, we can define the inverse transform in the same way:

$$\langle F^\vee, \phi\rangle = \langle F, \phi^\vee\rangle.$$

The Fourier inversion theorem formula $\phi = (\widehat{\phi})^\vee = (\phi^\vee)^{\widehat{T}}$ then extends to $\mathcal{S}'$:

$$\left\langle (\widehat{F})^\vee, \phi \right\rangle = \left\langle \widehat{F}, \phi^\vee \right\rangle = \langle F, (\phi^\vee)\widehat{\ } \rangle = \langle F, \phi \rangle$$

so that $(\widehat{F})^\vee = F$, and likewise $\left(F^\vee\right)\widehat{\ } = F$. Thus the Fourier transform is an isomorphism on S'.

If $F \in \mathcal{E}'$, there is an alternative way to define $\widehat{F}$. Indeed, $\langle F, \phi \rangle$ makes sense for any $\phi \in C^\infty$, and if we take $\phi(x) = e^{-2\pi i \xi \cdot x}$, we obtain a function of ξ that has a strong claim to be called $\widehat{F}(\xi)$. In fact, the two definitions are equivalent:

Proposition 9.25: 9.11.

If $F \in \mathcal{E}'$, then $\widehat{F}$ is a slowly increasing C^∞ function, and it is given by $\widehat{F}(\xi) = \langle F, E_{-\xi} \rangle$ where $E_\xi(x) = e^{2\pi i \xi \cdot x}$.

Proof. Let $g(\xi) = \langle F, E_{-\xi} \rangle$. Consideration of difference quotients of g, as in the proof of Proposition 3, shows that g is a C^∞ function with derivatives given by $\partial^\alpha g(\xi) = \langle F, \partial_\xi^\alpha E_{-\xi} \rangle = (-2\pi i)^{|\alpha|} \langle F, x^\alpha E_{-\xi} \rangle$. Moreover, by Theorem 22 and Proposition 87, there exist m, N, C such that

$$|\partial^\alpha g(\xi)| \leqslant C \sum_{|\beta| \leqslant N} \sup_{|x| \leqslant m} |\partial^\beta [x^\alpha E_{-\xi}(x)]| \leqslant C'(1+m)^{|\alpha|}(1+|\xi|)^N$$

so g is slowly increasing.

It remains to show that $g = \widehat{F}$, and by Proposition 23 it suffices to show that $\int g\phi = \langle F, \widehat{\phi} \rangle$ for $\phi \in C_c^\infty$. In this case $g\phi \in C_c^\infty$, so $\int g\phi$ can be approximated by Riemann sums as in the proof of Proposition 3, say $\sum g(\xi_j)\phi(\xi_j)\Delta\xi_j$. The corresponding sums $\sum \phi(\xi_j) e^{-2\pi i \xi_j \cdot x} \Delta\xi_j$ and their derivatives in x converge uniformly, for x in any compact set, to $\widehat{\phi}(x)$ and its derivatives. Therefore, since F is a continuous functional on C^∞,

$$\int g\phi = \lim \sum \langle F, E_{-\xi_j} \rangle \phi(\xi_j)\Delta\xi_j = \lim \left\langle F, \sum \phi(x_j) E_{-\xi_j} \Delta\xi_j \right\rangle = \langle F, \widehat{\phi} \rangle$$

It is time for some examples. First and foremost, the Fourier transform of the point mass at 0 is the constant function $1 : \langle \delta, E_{-\xi} \rangle = E_{-\xi}(0) = 1$. More generally, for point masses at other points and their derivatives, we have

$$\begin{aligned}\left(\partial^\alpha \tau_y \delta\right)\widehat{\ }(\xi) &= (-1)^{|\alpha|} \langle \delta, \tau_{-y} \partial^\alpha E_{-\xi} \rangle = (-1)^{|\alpha|} \partial_x^\alpha (e^{-2\pi i \xi \cdot (x+y)})|_{x=0} \\ &= (2\pi i \xi)^\alpha e^{-2\pi i \xi \cdot y}\end{aligned}$$

In particular:

Proposition 9.26: 9.12.

The Fourier transforms of the linear combinations of δ and its derivatives are precisely the polynomials.

The Fourier inversion theorem then yields the formulas for the Fourier transforms of polynomials and imaginary exponentials:

$$(x^\alpha)\widehat{\ } = [(-x)^\alpha]^\vee = (-2\pi i)^{-|\alpha|}\partial^\alpha \delta, \quad \widehat{E}_y = (E_{-y})^\vee = \tau_y \delta$$

As an illustration of the heuristics associated to these results, consider the formula

$$\int e^{2\pi i\xi\cdot x} d\xi = \delta(x)$$

Although this is nonsensical as a pointwise equality, it is valid when viewed from the right angle. One the one hand, it expresses the fact that the Fourier transform of the constant function 1 is δ. More interestingly, it is a concise statement of the Fourier inversion theorem. Indeed, if we replace x by $x-y$, integrate both sides against $\phi \in \mathcal{S}$, and reverse the order of integration on the left, we obtain

$$\iint \phi(y) e^{2\pi i\xi\cdot(x-y)} dy dx = \int \delta(x-y)\phi(y) dy$$

The integral on the left is $(\widehat{\phi})^\vee(x)$, and the integral on the right equals $\phi(x)$!

It is an important fact that every distibution is, at least locally, a linear combination of derivatives of continuous functions. The Fourier transform yields an easy proof of this:

Proposition 9.27: 9.14.

(a) If $F \in \mathcal{E}'$, there exist $N \in \mathbb{Z}_{\geqslant 1}$, constants $c_\alpha(|\alpha| \leqslant N)$, and $f \in C_0(\mathbb{Z}^n_{\geqslant 1})$ such that $F = \sum_{|\alpha|\leqslant N} c_\alpha \partial^\alpha f$.

(b) If $F \in \mathcal{D}'(U)$ and V is a precompact open set with $\overline{V} \subset U$, there exist N, c_α, f as above such that $F = \sum_{|\alpha|\leqslant N} c_\alpha \partial^\alpha f$ on V.

Proof. By Proposition 25, if $F \in \mathcal{E}'$ then $\widehat{F}$ is slowly increasing, so the function $g(\xi) = (1+|\xi|^2)^{-M}\widehat{F}(\xi)$ will be in L^1 if the integer M is chosen sufficiently large. Let $f = \widehat{g}$; then $f \in C_0$ and $\widehat{F} = (1+|\xi|^2)^M \widehat{f}$, so $F = (I - (4\pi^2)^{-1}\sum_1^n \partial_j^2)^M f$. This proves (a); for (b), choose $\psi \in C_c^\infty(U)$ such that $\psi = 1$ on V, and apply (a) to ψF.

We conclude this section with a sketch of the theory of periodic distributions; some of the details are fleshed out in Exercises 22-24.

The space $C^\infty(\mathbb{T}^n)$ of smooth periodic functions is a Fréchet space with the topology defined by the seminorms $\phi \mapsto \|\partial^\alpha \phi\|_u$, and a distribution on $\mathbb{T}^n$ is a continuous linear functional on this space; the space of distributions on $\mathbb{T}^n$ is denoted by $\mathcal{D}'(\mathbb{T}^n)$. If $F \in \mathcal{D}'(\mathbb{T}^n)$, its Fourier transform is the function $\widehat{F}$ on $\mathbb{T}^n$ define by $\widehat{f}(\kappa) = \langle F, E_{-\kappa}\rangle$ where $E_\kappa(x) = e^{2\pi i\kappa\cdot x}$. Since F satisfies an estimate of the form $|\langle F, \phi\rangle| \leqslant C\sum_{|\alpha|\leqslant N}\|\partial^\alpha\phi\|_u$, there exist C, N such that

$$|\widehat{F}(\kappa)| \leqslant C(1+|\kappa|)^N$$

and the Fourier transform is an isomorphism from $\mathcal{D}'(\mathbb{T}^n)$ to the space of all functions on $\mathbb{T}^n$ satisfying such an estimate. Moreover, if $F \in \mathcal{D}'(\mathbb{T}^n)$, the Fourier series $\sum_\kappa \widehat{F}(\kappa)E_\kappa$

converges in $\mathcal{D}'(\mathbb{T}^n)$ to F.

Instead of defining periodic distributions as distributions on $\mathbb{T}^n$ (linear functionals on $C^\infty(\mathbb{T}^n)$), one can define them as distributions on $\mathbb{T}^n$ (linear functionals on $C_c^\infty(\mathbb{T}^n)$) that are invariant under the translations $\tau_\kappa, \kappa \in \mathbb{T}^n$. Accordingly, let

$$\mathcal{D}'(\mathbb{R}^n)_{\text{per}} = \{F \in \mathcal{D}'(\mathbb{R}^n) \mid \tau_\kappa F = F \text{ for } \kappa \in \mathbb{R}\}$$

The periodization map $P\phi = \sum_{\kappa \in \mathbb{Z}^n} \tau_\kappa \phi$ used in Theorem 40 is easily seen to map $C_c^\infty(\mathbb{Z}^n)$ continuously into $C^\infty(\mathbb{Z}^n)$, so it induces a map $P' \colon \mathcal{D}'(\mathbb{Z}^n) \to \mathcal{D}'(\mathbb{Z}^n)$ given by $\langle P'F, \phi\rangle = \langle F, P\phi\rangle$. Since $P \circ \tau_\kappa = P$ for $\kappa \in \mathbb{Z}^n$, we have $\tau_\kappa \circ P' = P'$, that is, the range of P' lies in $\mathcal{D}'(\mathbb{Z}^n)_{\text{per.}}$. In fact, $P' \colon \mathcal{D}'(\mathbb{Z}^n) \to \mathcal{D}'(\mathbb{Z}^n)_{\text{per}}$ is a bijection. (The proof is nontrivial; see Folland Exercise 9.24.) Moreover, if $f \in L^1(\mathbb{Z}^n)$, then f and $P'f$ coincide as periodic functions on $\mathbb{Z}$, for if $\phi \in C_c^\infty(\mathbb{Z}^n)$,

$$\begin{aligned}\langle P'f, \phi\rangle = \langle f, P\phi\rangle &= \int_{[0,1)^n} f(x) \sum \phi(x-\kappa) dx \\ &= \sum \int_{[0,1)^n+\kappa} f(x)\phi(x)dx = \int_{\mathbb{R}^n} f(x)\phi(x)dx = \langle f, \phi\rangle.\end{aligned}$$

Thus the two descriptions of periodic distributions are equivalent.

If $F \in \mathcal{D}'(\mathbb{T}^n)$, the Fourier series $\sum \widehat{F}(\kappa)E_\kappa$ converges in $\mathcal{D}'(\mathbb{T}^n)$ to F; on the other hand, it follows easily from (9.15) that it also converges in $S'(\mathbb{T}^n)$, and its sum there is $P'f$. Thus $\mathcal{D}'(\mathbb{T}^n)_{\text{per}} \subset \mathcal{D}'(\mathbb{T}^n)$, and by (9.13) we have

$$(P'F) = \sum \widehat{F}(\kappa)\widehat{E}_\kappa = \sum \widehat{F}(\kappa)\tau_\kappa \delta$$

giving the relation between the $\mathbb{R}^n$—and $\mathbb{R}^n$-Fourier transforms for periodic distributions. In particular, if $F = \delta_{\mathbb{R}^n}$, the point mass at the origin in $\mathbb{R}^n$, then $\widehat{F}(\kappa) = 1$ for all κ; hence $P'F$ and $(P'F)$ are both equal to $\sum \tau_\kappa \delta -$ a restatement of the Poisson summation formula.

Exercise 9.28: Folland Exercise 9.16.

Suppose $F \in \mathcal{E}'$ and $\psi \in C^\infty$. Show that for any $\phi \in C_c^\infty$, $\int \left\langle F, \tau_x \widetilde{\psi} \right\rangle \phi(x) dx = \langle F, \phi * \widetilde{\psi}\rangle$. (The result can be reduced to Proposition 3; given F and ϕ, the indicated expressions depend only on the values of ψ in a compact set.)

Exercise 9.29: Folland Exercise 9.17.

Suppose that $F \in \mathcal{S}'$. Show that

(a) $\widehat{\left(\tau_y F\right)} = e^{-2\pi i \xi \cdot y}\widehat{F}, \tau_\eta \widehat{F} = \widehat{\left[e^{2\pi i \eta \cdot x} F\right]}$.

(b) $\partial^\alpha \widehat{F} = \widehat{\left[(-2\pi i x)^\alpha F\right]}, \widehat{(\partial^\alpha F)} = (2\pi i \xi)^\alpha \widehat{F}$.

(c) $\widehat{\left(F \circ T\right)} = |\det T|^{-1} \widehat{F} \circ (T^*)^{-1}$ for $T \in GL(n, \mathbb{R})$.

(d) $(F * \psi)\widehat{\psi}\widehat{\psi}\widehat{F}$ for $\psi \in \mathcal{S}$.

Exercise 9.30: Folland Exercise 9.18.

If $n = l + m$, let us write $x \in \mathbb{R}^n$ as (y, z) with $y \in \mathbb{R}^l$ and $z \in \mathbb{R}^m$. Let $\mathcal{F}$ denote the Fourier transform on $\mathbb{R}^n$ and $\mathcal{F}_1, \mathcal{F}_2$ the partial Fourier transforms in the first and second sets of variables—i.e., $\mathcal{F}_1 f(\eta, z) = \int f(y, z)e^{-2\pi i\eta\cdot y}dy$ and likewise for $\mathcal{F}_2$. Then $\mathcal{F}_1$ and $\mathcal{F}_2$ are isomorphisms on $\mathcal{F}(\mathbb{R}^n)$ and $\mathcal{F}'(\mathbb{R}^n)$, and $\mathcal{F} = \mathcal{F}_1\mathcal{F}_2 = \mathcal{F}_2\mathcal{F}_1$.

Exercise 9.31: Folland Exercise 9.19.

On $\mathbb{R}$, let $F_0 = PV(1/x)$ as defined in Folland Exercise 9.10. Also, for $\varepsilon > 0$ let $F_\varepsilon(x) = x(x^2 + \varepsilon^2)^{-1}, G_\varepsilon^{\pm}(x) = (x \pm i\varepsilon)^{-1}$, and $S_\varepsilon(x) = e^{-2\pi\varepsilon|x|}\operatorname{sgn} x$.

(a) $\lim_{\varepsilon\to 0} F_\varepsilon = F_0$ in the weak* topology of $\mathcal{S}'$. (Theorem 17, with $a = 0$, may be useful.)

(b) $\lim_{\varepsilon\to 0} G_\varepsilon = F_0 \mp \pi i\delta$. (Hint: $(x \pm i\varepsilon)^{-1} = (x \mp i\varepsilon)(x^2 + \varepsilon^2)^{-1}$.)

(c) $\widehat{S}_\varepsilon = (\pi i)^{-1}F_\varepsilon$ and hence $\widehat{\operatorname{sgn}} = (\pi i)^{-1}F_0$.

(d) From (c) it follows that $\widehat{F}_0 = -\pi i \operatorname{sgn}$. Prove this directly by showing that $F_0 = \lim_{\varepsilon\to 0, N\to\infty} H_{\varepsilon,N}$, where $H_{\varepsilon,N}(x) = x^{-1}$ if $\varepsilon < |x| < N$ and $H_{\varepsilon,N}(x) = 0$ otherwise, and using Folland Exercise 2.59(b).

(e) Compute $\widehat{\chi}_{(0,\infty)}$ (i) by writing $\chi_{(0,\infty)} = \frac{1}{2}\operatorname{sgn} + \frac{1}{2}$ and using (c), (ii) by writing $\chi_{(0,\infty)}(x) = \lim e^{-\varepsilon x}\chi_{(0,\infty)}(x)$ and using (b).

Exercise 9.32: Folland Exercise 9.20.

Suppose that $F \in \mathcal{S}'$ and $G \in \mathcal{S}'$.

(a) $\widehat{F}\widehat{G}$ is well-defined element of $\mathcal{S}'$.

(b) If $\psi \in \mathcal{S}$, then $G * \psi \in \mathcal{S}$.

(c) Let $F * G$ (or $G * F$) be the tempered distribution such that $(F * G)\widehat{F} = \widehat{F}\widehat{\ }$. Then $\langle F * G, \psi\rangle = \langle F, \widetilde{G} * \psi\rangle = \langle G, \widetilde{F} * \psi\rangle$ for $\psi \in \mathcal{S}$.

Exercise 9.33: Folland Exercise 9.21.

Suppose that $F, G, H \in \mathcal{S}'$.

(a) If at most one of F, G, H has noncompact support, then $(F * G) * H = F * (G * H)$, where the convolutions are defined as in Folland Exercise 9.20.

(b) On $\mathbb{R}$, let F be the constant function $1, G = d\delta/dx$, and $H = \chi_{(0,\infty)}$. Then $(F * G) * H$ and $F * (G * H)$ are well defined in $\mathcal{S}'$ but are unequal.

Exercise 9.34: Folland Exercise 9.22.

Let $E_\kappa(x) = e^{2\pi i\kappa\cdot x}$. If $g\colon \mathbb{Z}^n \to \mathbb{Z}$ satisfies $|g(\kappa)| \leqslant C(1+|\kappa|)^N$ for some $C, N > 0$, then the series $\sum_{\kappa\in\mathbb{Z}^n} g(\kappa)E_\kappa$ converges in $\mathcal{D}'(\mathbb{Z}^n)$ to a distribution F that satisfies $\widehat{F} = g$. It also converges in $\mathcal{D}'(\mathbb{Z}^n)$ to a tempered distribution G $(= P'F)$ such that $\tau_\kappa G = G$ for all κ.

Exercise 9.35: Folland Exercise 9.23.

Suppose that $F, G \in \mathcal{D}'(\mathbb{T}^n)$.

(a) There is a unique $F * G \in \mathcal{D}'(\mathbb{T}^n)$ such that $(F*G)\widehat{\ } = \widehat{F}\widehat{G}$. (Use Folland Exercise 9.22.)

(b) If $G \in C^\infty(\mathbb{T}^n)$, then $F * G \in C^\infty(\mathbb{T}^n)$ and $F * G(x) = \left\langle F, \tau_x\widetilde{G}\right\rangle$ as on $\mathbb{T}^n$.

Exercise 9.36: Folland Exercise 9.24.

Let P be the periodization map, $P\phi = \sum_{\kappa\in\mathbb{Z}^n}\tau_\kappa\phi$.

(a) P is a continuous linear map from $C_c^\infty(\mathbb{R}^n)$ to $C^\infty(\mathbb{R}^n)$. (Note that for $\phi \in C_c^\infty$ and x in a compact set, only finitely many terms of the series $\sum \tau_\kappa\phi(x)$ are nonzero.)

(b) Choose $\gamma \in C_c^\infty$ with $\int\gamma = 1$, and let $\omega = \gamma * \chi_{[0,1)^n}$. Then $\omega \in C_c^\infty$ and $P\omega = 1$.

(c) If $\psi \in C^\infty(\mathbb{T}^n)$, then $\psi = P(\omega\psi)$ (where ψ is regarded as a function on $\mathbb{T}^n$ on the left and as a function on $\mathbb{T}^n$ on the right). Consequently, $P\colon C_c^\infty(\mathbb{T}^n) \to C^\infty(\mathbb{T}^n)$ is surjective and the dual map $P'\colon \mathcal{D}'(\mathbb{T}^n) \to \mathcal{D}'(\mathbb{T}^n)_{\text{per}}$ is injective. d. Given $G \in \mathcal{D}'(\mathbb{T}^n)_{\text{per}}$, define $F \in \mathcal{D}'(\mathbb{T}^n)$ by $\langle F, \psi\rangle = \langle G, \omega\psi\rangle$ (with the same understanding as in part (c)). Then $P'F = G$, so P' maps $\mathcal{D}'(\mathbb{T}^n)$ onto $\mathcal{D}'(\mathbb{T}^n)_{\text{per}}$.

Exercise 9.37: Folland Exercise 9.25.

Suppose that P is a polynomial in n variables such that only zero of $P(\xi)$ in $\mathbb{R}^n$ is $\xi = 0$, and let $P(D)$ be as in Folland Section 8.7.

(a) Every tempered distribution F that satifies $P(D)F = 0$ is a polynomial. (Use Proposition 26 and Folland Exercise 9.11.)

(b) Every bounded function f that satisfies $P(D)f = 0$ is a constant. (This result, for the special cases where $P(D)$ is the Laplacian or the Cauchy-Riemann operator $\partial_x + i\partial_y$ on $\mathbb{R}^2$, is known as Liouville's theorem.)

Exercise 9.38: Folland Exercise 9.26.

On $\mathbb{R}^n \times \mathbb{R}$, let $G(x,t) = (4\pi t)^{-n/2}e^{-|x|^2/4t}\chi_{(0,\infty)}(t)$.

(a) $\widehat{G}$ is the tempered function $\widehat{G}(\xi,\tau) = (2\pi i\tau + 4\pi^2|\xi|^2)^{-1}$. (Use Proposition 33 and Folland Exercise 9.18.)

(b) Deduce that $(\partial_t - \Delta)G = \delta$. (Cf. Folland Exercise 9.15.)

Exercise 9.39: Folland Exercise 9.27.

Suppose that $0 < \operatorname{Re}\alpha < n$.

(a) For any $\phi \in \mathcal{S}$,
$$\frac{\Gamma((n-\alpha)/2)}{\pi^{(n-\alpha)/2}} \int |x|^{\alpha-n}\widehat{\phi}(x)dx = \frac{\Gamma(\alpha/2)}{\pi^{\alpha/2}} \int |\xi|^{-\alpha}\phi(\xi)d\xi$$
(Hint: By Proposition 33 and Lemma 34, if $t > 0$ we have
$$\int e^{-\pi t|x|^2}\widehat{\phi}(x)dx = t^{-n/2}\int e^{-\pi|\xi|^2/t}\phi(\xi)d\xi$$
Multiply both sides by $t^{-1+(n-\alpha)/2}dt$ and integrate from 0 to ∞.)

(b) Let $R_\alpha(x) = \Gamma((n-\alpha)/2)[\Gamma(\alpha/2)2^\alpha\pi^{n/2}]^{-1}|x|^{\alpha-n}$. Then R_α is a tempered function and $\widehat{R}_\alpha$ is the tempered function $\widehat{R}_\alpha(\xi) = (2\pi|\xi|)^{-\alpha}$.

(c) If $n > 2$, then $\Delta R_2 = -\delta$. (Cf. Folland Exercise 9.14. See the next exercise for the case $n = 2$.)

Exercise 9.40: Folland Exercise 9.28.

Suppose $n = 2$. For $0 < \operatorname{Re}\alpha < 2$, let $c_\alpha = \Gamma((2-\alpha)/2)[\Gamma(\alpha/2)2^\alpha\pi]^{-1}$ and $Q_\alpha(x) = c_\alpha(|\xi|^{\alpha-2} - 1)$. (Note that Q_α differs by a constant from the R_α in Folland Exercise 9.27.)

(a) $\lim_{\alpha\to 2} Q_\alpha(x) = -(2\pi)^{-1}\log|x|$, pointwise and in $\mathcal{S}'$.

(b) By (a), $\lim_{\alpha\to 2}\widehat{Q}_\alpha$ exists in $\mathcal{S}'$, and by Exercise 27b, $\widehat{Q}_\alpha(\xi) = (2\pi|\xi|)^{-\alpha} - c_\alpha\delta$. Noting that $(2\pi|\xi|)^{-2}$ is not integrable near the origin and that $\lim_{\alpha\to 2} c_\alpha = \infty$, find an explicit formula for $\lim_{\alpha\to 2}\widehat{Q}_\alpha$. (Folland Exercise 9.12 may help.)

Exercise 9.41: Folland Exercise 9.29.

For $1 \leqslant p < \infty$, let $\mathcal{C}_p$ be the set of all $F \in \mathcal{C}'$ for which there exists $C \geqslant 0$ such that $\|F * \phi\|_p \leqslant C\|\phi\|_p$ for all $\phi \in \mathcal{C}$, so that the map $\phi \mapsto F * \phi$ extends to a bounded operator on L^p.

(a) $\mathcal{C}_1 = M(\mathbb{R}^n)$. (If $F \in \mathcal{C}_1$, consider $F * \phi_t$ where $\{\phi_t\}$ is an approximate identity, and apply Alaoglu's theorem.)

(b) $\mathcal{C}_2 = \left\{F \in \mathcal{C}' \,\middle|\, \widehat{F} \in L^\infty\right\}$. (Use the Plancherel theorem.)

(c) If p and q are conjugate exponents, then $\mathcal{C}_p = \mathcal{C}_q$. (Hint: $\langle F * \phi, \psi\rangle = \langle F * \widetilde{\psi}, \widetilde{\phi}\rangle$.)

(d) If $1 \leqslant p \leqslant 2$ and q is the conjugate exponent to p, then $\mathrm{e}_p \subset \mathrm{C}_r$ for all $r \in (p, q)$.

(Use the Riesz-Thorin theorem.)

(e) $\mathcal{C}_1 \subset \mathcal{C}_p \subset \mathcal{C}_2$ for all $p \in (1, \infty)$.

9.3 Sobolev Spaces

One of the most satisfactory ways of measuring smoothness properties of functions and distributions is in terms of L^2 norms. There are two reasons for this: L^2 has the advantage of being a Hilbert space, and the Fourier transform, which converts differentiation into multiplication by the coordinate functions, is an isometry on L^2.

As a first step, suppose $k \in \mathbb{Z}_{\geqslant 1}$ and let H_k be the space of all functions $f \in L^2(\mathbb{Z}^n_{\geqslant 1})$ whose distribution derivatives $\partial^\alpha f$ are L^2 functions for $|\alpha| \leqslant k$. One can make H_k into a Hilbert space by imposing the inner product

$$(f, g) \mapsto \sum_{|\alpha| \leqslant k} \int (\partial^\alpha f)(\overline{\partial^\alpha g})$$

However, it is more convenient to use an equivalent inner product defined in terms of the Fourier transform. Theorem 31e and the Plancherel theorem imply that $f \in H_k$ if and only if $\xi^\alpha \widehat{f} \in L^2$ for $|\alpha| \leqslant k$. A simple modification of the argument in the proof of Proposition 2 shows that there exist $C_1, C_2 > 0$ such that

$$C_1(1 + |\xi|^2)^k \leqslant \sum_{|\alpha| \leqslant k} |\xi^\alpha|^2 \leqslant C_2(1 + |\xi|^2)^k$$

from which it follows that $f \in H_k$ if and only if $(1 + |\xi|^2)^{k/2} \widehat{f} \in L^2$ and that the norms

$$f \mapsto \left(\sum_{|\alpha| \leqslant k} \|\partial^\alpha f\|_2^2 \right)^{1/2} \quad \text{and} \quad f \mapsto \left\| (1 + |\xi|^2)^{k/2} \widehat{f} \right\|_2$$

are equivalent. The latter norm, however, makes sense for any $k \in \mathbb{R}$, and we can use it to extend the definition of H_k to all real k.

We proceed to the formal definitions. For any $s \in \mathbb{R}$ the function $\xi \mapsto (1 + |\xi|^2)^{s/2}$ is C^∞ and slowly increasing (Folland Exercise 9.30), so the map Λ_s defined by

$$\Lambda_s f = \left[(1 + |\xi|^2)^{s/2} \widehat{f} \right]^\vee$$

is a continuous linear operator on $\mathcal{S}'$—actually an isomorphism, since $\Lambda_s^{-1} = \Lambda_{-s}$. If $s \in \mathbb{R}$, we define the Sobolev space H_s to be

$$H_s = \{ f \in \mathcal{S}' \mid \Lambda_s f \in L^2 \}$$

and we define an inner product and norm on H_s by

$$\langle f, g \rangle_{(s)} = \int (\Lambda_s f)(\overline{\Lambda_s g}) = \int \widehat{f}(\xi)(1 + |\xi|^2)^s \overline{g}(\xi)$$

$$\|f\|_{(s)} = \|\Lambda_s f\|_2 = \left[\int |\widehat{f}(\xi)|^2 (1 + |\xi|^2)^s d\xi \right]^{1/2}$$

(The equality of the two formulas for $\langle f, g \rangle_{(s)}$ and for $\|f\|_{(s)}$ follows from the Plancherel theorem.) Note that the inner products $\langle\cdot,\cdot\rangle_{(s)}$ are conjugate linear in the second variable, but we are continuing to use the notation $\langle\cdot,\cdot\rangle$ for the bilinear pairing between $\mathcal{S}'$ and $\mathcal{S}$. This will cause no confusion, since we shall not be using the inner products $\langle\cdot,\cdot\rangle_{(s)}$ explicitly.

The following properties of Sobolev spaces are simple consequences of the definitions and the preceding discussion:

(i) The Fourier transform is a unitary isomorphism from H_s to $L^2(\mathbb{R}^n,\mu_s)$ where $d\mu_s(\xi) = (1+|\xi|^2)^s d\xi$. In particular, H_s is a Hilbert space.

(ii) $\mathcal{S}$ is a dense subspace of H_s for all $s \in \mathbb{R}$. (This follows easily from (i) and Proposition 19.)

(iii) If $t < s$, H_s is a dense subspace of H_t in the topology of H_t, and $\|\cdot\|_{(t)} \leqslant \|\cdot\|_{(s)}$.

(iv) Λ_t is a unitary isomorphism from H_s to H_{s-t} for all $s,t \in \mathbb{R}$.

(v) $H_0 = L^2$ and $\|\cdot\|_{(0)} = \|\cdot\|_2$ (by Plancherel).

(vi) ∂^α is a bounded linear map from H_s to $H_{s-|\alpha|}$ for all s,α (because $|\xi^\alpha| \leqslant (1+|\xi|^2)^{|\alpha|/2}$).

By (iii) and (v), for $s \geqslant 0$ the distributions in H_s are L^2 functions. For $s < 0$ the elements of H_s are generally not functions. For example, the point mass δ is in H_s if and only if $s < -\frac{1}{2}n$, for $\widehat{\delta}$ is the constant function 1, and $\int(1+|\xi|^2)^s d\xi < \infty$ if and only if $s < -\frac{1}{2}n$. Another example: The distribution W_t whose Fourier transform is $(2\pi|\xi|)^{-1}\sin 2\pi t|\xi|$, which arose in the discussion of the wave equation in Folland Section 8.7, is in H_s if and only if $s < 1-\frac{1}{2}n$; it is in $L^1 \cap L^2$ when $n=1$ and in $L^1 \backslash L^2$ for $n=2$, but is not a function for $n \geqslant 3$.

Proposition 9.42: 9.16.

If $s \in \mathbb{R}$, the duality between $\mathcal{S}'$ and S induces a unitary isomorphism from H_{-s} to $(H_s)^*$. More precisely, if $f \in H_{-s}$, the functional $\phi \mapsto \langle f,\phi\rangle$ on $\mathcal{S}$ extends to a continuous linear functional on H_s with operator norm equal to $\|f\|_{(-s)}$, and every element of $(H_s)^*$ arises in this fashion.

Proof. If $f \in H_{-s}$ and $\phi \in \mathcal{S}$,

$$\langle f,\phi\rangle = \left\langle f^\vee, \widehat{\phi}\right\rangle = \int f^\vee(\xi)\widehat{\phi}(\xi)d\xi$$

since $f^\vee(\xi) = \widehat{f}(-\xi)$ is a tempered function. Thus by the Schwarz inequality,

$$\begin{aligned}|\langle f,\phi\rangle| &\leqslant \left[\int |f^\vee(\xi)|^2(1+|\xi|^2)^{-s}d\xi\right]^{1/2}\left[\int |\widehat{\phi}(\xi)|^2(1+|\xi|^2)^s d\xi\right]^{1/2} \\ &= \|f\|_{(-s)}\|\phi\|_{(s)}\end{aligned}$$

so the functional $\phi \mapsto \langle f,\phi\rangle$ extends continuously to H_s, with norm at most $\|f\|_{(-s)}$. In fact, its norm equals $\|f\|_{(-s)}$, since if $g \in \mathcal{S}'$ is the distribution whose Fourier transform is

$\widehat{g}(\xi) = (1+|\xi|^2)^{-s}\overline{\widehat{f}(\xi)}$, we have $g \in H_s$ and

$$\langle f, g\rangle = \int |\widehat{f}(\xi)|^2(1+|\xi|^2)^s d\xi = \|f\|_{(-s)}^2 = \|f\|_{(-s)}\|g\|_{(s)}$$

Finally, if $G \in (H_s)^*$, then $G \circ \mathcal{F}^{-1}$ is a bounded linear functional on $L^2(\mu_s)$ where $d\mu_s(\xi) = (1+|\xi|^2)^s d\xi$, so there exists $g \in L^2(\mu_s)$ such that

$$G(\phi) = \int \widehat{\phi}(\xi) g(\xi)(1+|\xi|^2)^s d\xi$$

But then $G(\phi) = \langle f, \phi\rangle$ where $f^\vee(\xi) = (1+|\xi|^2)^s g(\xi)$, and $f \in H_{-s}$ since

$$\|f\|_{(-s)}^2 = \int |\widehat{f}(\xi)|^2(1+|\xi|^2)^{-s} d\xi = \int |g(\xi)|^2(1+|\xi|^2)^s d\xi$$

For $s > 0$, the elements of H_s are L^2 functions that are " L^2-differentiable up to order s," and it is natural to ask what is the relationship between this notion of smoothness and ordinary differentiability. Of course, if one thinks of elements of H_s as distributions or elements of L^2, there is no distinction among functions that agree almost everywhere; from this perspective, when one says that a function in H_s is of class C^k, one means that it agrees a.e. with a C^k function. With this understanding, the question just posed has a simple and elegant answer. We introduce the notation

$$C_0^k = \{f \in C^k(\mathbb{R}^n) \mid \partial^\alpha f \in C_0 \text{ for } |\alpha| \leqslant k\}$$

C_0^k is a Banach space with the C^k norm $f \mapsto \sum_{|\alpha|\leqslant k}\|\partial^\alpha f\|_u$.

Theorem 9.43: 9.17: The Sobolev Embedding Theorem.

Suppose $s > k + \frac{1}{2}n$.

(a) If $f \in H_s$, then $\left(\partial^\alpha f\right)\widehat{} \in L^1$ and $\left\|\widehat{(\partial^\alpha f)}\right\|_1 \leqslant C\|f\|_{(s)}$ for $|\alpha| \leqslant k$, where C depends only on $k - s$.

(b) $H_s \subset C_0^k$, and the inclusion map is continuous.

Proof. By the Schwarz inequality,

$$\begin{aligned}(2\pi)^{-|\alpha|}\int |\left(\partial^\alpha f\right)\widehat{}(\xi)| d\xi &= \int \left|\xi^\alpha \widehat{f}(\xi)\right| d\xi \leqslant \int (1+|\xi|^2)^{k/2}|\widehat{f}(\xi)| d\xi \\ &\leqslant \left[\int (1+|\xi|^2)^s|\widehat{f}(\xi)|^2 d\xi\right]^{1/2}\left[\int (1+|\xi|^2)^{k-s} d\xi\right]^{1/2}\end{aligned}$$

The first factor on the right is $\|f\|_{(s)}$, and the second one is finite by Corollary 101 since $2(k-s) < -n$. This proves (a), and (b) follows by the Fourier inversion theorem and the Riemann-Lebesgue lemma.

Corollary 9.44: 9.18.

If $f \in H_s$ for all s, then $f \in C^\infty$.

An example may help to elucidate this theorem. Let $f_\lambda(x) = \phi(x)|x|^\lambda$, where $\lambda \in \mathbb{R}$ and $\phi \in C_c^\infty$ with $\phi = 1$ on a neighborhood of 0. Then the (classical) derivative $\partial^\alpha f_\lambda$ is C^∞ except at 0 and is homogeneous of degree $\lambda - |\alpha|$ near 0, so that $|\partial^\alpha f_\lambda| \leqslant C_{\alpha,\lambda}|x|^{\lambda-|\alpha|}$, and in particular $\partial^\alpha f_\lambda \in L^1$ provided $\lambda - |\alpha| > -n$. In this case $\partial^\alpha f_\lambda$, as an L^1 function, is also the distribution derivative of f_λ. (To see this, replace f_λ by the C^∞ function $\phi(x)(|x|^2 + \varepsilon^2)^{\lambda/2}$ and consider the limit as $\varepsilon \to 0$.) Moreover, $\partial^\alpha f_\lambda \in L^2$ if and only if $\lambda - |\alpha| > -\frac{1}{2}n$, so $f \in H_k (k = 0, 1, 2, \ldots)$ if and only if $\lambda > k - \frac{1}{2}n$, whereas $f_\lambda \in C_0^k$ if and only if $\lambda > k$. See also Exercises 33-35 for some related results.

Next, we show that multiplication by suitably smooth functions preserves the H_s spaces. We need a lemma:

Lemma 9.45: 9.19.

For all $\xi, \eta \in \mathbb{R}^n$ and $s \in \mathbb{R}$,

$$(1 + |\xi|^2)^s(1 + |\eta|^2)^{-s} \leqslant 2^{|s|}(1 + |\xi - \eta|^2)^{|s|}$$

Proof. Since $|\xi| \leqslant |\xi - \eta| + |\eta|$, we have $|\xi|^2 \leqslant 2(|\xi - \eta|^2 + |\eta|^2)$ and hence

$$1 + |\xi|^2 \leqslant 2(1 + |\xi - \eta|^2)(1 + |\eta|^2).$$

If $s \geqslant 0$, we have merely to raise both sides to the sth power. If $s < 0$, we interchange ξ and η and replace s by $-s$, obtaining

$$(1 + |\eta|^2)^{-s} \leqslant 2^{-s}(1 + |\xi|^2)^{-s}(1 + |\xi - \eta|^2)^{-s}$$

which is again the desired result.

Theorem 9.46: 9.20.

Suppose that $\phi \in C_0(\mathbb{R}^n)$ and that $\widehat{\phi}$ is a function that satisfies

$$\int (1 + |\xi|^2)^{a/2}|\widehat{\phi}(\xi)|d\xi = C < \infty$$

for some $a > 0$. Then the map $M_\phi(f) = \phi f$ is a bounded operator on H_s for $|s| \leqslant a$.

Proof. Since Λ_s is a unitary map from H_s to $H_0 = L^2$, it is equivalent to show that $\Lambda_s M_\phi \Lambda_{-s}$ is a bounded operator on L^2. But

$$\left(\Lambda_s M_\phi \Lambda_{-s} f \widehat{f}\right)(\xi) = (1 + |\xi|^2)^{s/2}\left[\widehat{\phi} * (\Lambda_{-s} f)\widehat{]}(\xi) = \int K(\xi, \eta)\widehat{f}(\eta)d\eta\right.$$

where

$$K(\xi, \eta) = (1 + |\xi|^2)^{s/2}(1 + |\eta|^2)^{-s/2}\widehat{\phi}(\xi - \eta)$$

By Lemma 45,

$$|K(\xi,\eta)| \leqslant 2^{|s|/2}(1+|\xi-\eta|^2)^{|s|/2}|\widehat{\phi}(\xi-\eta)|$$

so if $|s| \leqslant a$, then $\int |K(\xi,\eta)|d\xi$ and $\int |K(\xi,\eta)|d\eta$ are bounded by $2^{a/2}C$. That $\Lambda_s M_\phi \Lambda_{-s}$ is bounded on L^2 therefore follows from the Plancherel theorem and Theorem 50.

Corollary 9.47: 9.21.

If $\phi \in \mathcal{S}$, then M_ϕ is a bounded operator on H_s for all $s \in \mathbb{R}$.

Our next result is a compactness theorem that is of great importance in the applications of Sobolev spaces.

Theorem 9.48: 9.22: Rellich's Theorem.

Suppose that $\{f_k\}$ is a sequence of distributions in H_s that are all supported in a fixed compact set K and satisfy $\sup_k \|f_k\|_{(s)} < \infty$. Then there is a subsequence $\{f_{k_j}\}$ that converges in H_t for all $t < s$.

Proof. First we observe that by Proposition 25, $\widehat{f}_k$ is a slowly increasing C^∞ function. Pick $\phi \in C_c^\infty$ such that $\phi = 1$ on a neighborhood of K. Then $f_k = \phi f_k$, so $\widehat{f}_k = \widehat{\phi} * \widehat{f}_k$ where the convolution is defined pointwise by an absolutely convergent integral. By Lemma 45 and the Schwarz inequality,

$$\begin{aligned}&(1+|\xi|^2)^{s/2}\left|\widehat{f}_k(\xi)\right| \\ \leqslant& 2^{|s|/2}\int |\widehat{\phi}(\xi-\eta)|(1+|\xi-\eta|^2)^{|s|/2}\left|\widehat{f}_k(\eta)\right|(1+|\eta|^2)^{s/2}d\eta \\ \leqslant& 2^{|s|/2}\|\phi\|_{(|s|)}\|f_k\|_{(s)} \leqslant \text{ constant.}\end{aligned}$$

Likewise, since $\partial_j\left(\widehat{\phi} * \widehat{f}_k\right) = \left(\partial_j\widehat{\phi}\right) * \widehat{f}_k$, we see that $(1+|\xi|^2)^{s/2}\left|\partial_j \widehat{f}_k(\xi)\right|$ is bounded by a constant independent of ξ, j, and k. In particular, the $\widehat{f}_k$s and their first derivatives are uniformly bounded on compact sets, so by the mean value theorem and the Arzelà-Ascoli theorem there is a subsequence $\left\{\widehat{f}_{k_j}\right\}$ that converges uniformly on compact sets.

We claim that $\{f_{k_j}\}$ is Cauchy in H_t for all $t < s$. Indeed, for any $R > 0$ we can write the integral

$$\|f_{k_i} - f_{k_j}\|_{(t)}^2 = \int (1+|\xi|^2)^t \left|\widehat{f}_{k_i} - \widehat{f}_{k_j}\right|^2(\xi)d\xi$$

as the sum of the integrals over the regions $|\xi| \leqslant R$ and $|\xi| > R$. For $|\xi| \leqslant R$ we use the estimate

$$(1+|\xi|^2)^t \leqslant (1+R^2)^{\max(t,0)}$$

and for $|\xi| > R$ we use the estimate

$$(1+|\xi|^2)^t \leqslant (1+R^2)^{t-s}(1+|\xi|^2)^s$$

which yield

$$\|f_{k_2} - f_{k_j}\|_{(t)}^2 \leqslant CR^n(1+R^2)^{\max(t,0)} \sup_{|\xi|\leqslant R}\left|\widehat{f}_{k_i} - \widehat{f}_{k_j}\right|^2(\xi) + (1+R^2)^{t-s}\|f_{k_i} - f_{k_j}\|_{(s)}^2$$

Given $\varepsilon > 0$, the second term will be less than $\frac{1}{2}\varepsilon$ provided R is chosen sufficiently large, since $t - s < 0$; once such an R is fixed, the first term will less than $\frac{1}{2}\varepsilon$ provided i and j are sufficiently large. The proof is therefore complete.

Although the definition of Sobolev spaces in terms of the Fourier transform entails their elements being defined on all of $\mathbb{R}^n$, these spaces can also be used in the study of local smoothness properties of functions. The key definition is as follows: If U is an open set in $\mathbb{R}^n$, the localized Sobolev space $H_s^{\text{loc}}(U)$ is the set of all distributions $f \in \mathcal{D}'(U)$ such that for every precompact open set V with $\overline{V} \subset U$ there exists $g \in H_s$ such that $g = f$ on V.

Proposition 9.49: 9.23.

A distribution $f \in \mathcal{D}'(U)$ is in $H_s^{\text{loc}}(U)$ if and only if $\phi f \in H_s$ for every $\phi \in C_c^\infty(U)$.

Proof. If $f \in H_s^{\text{loc}}(U)$ and $\phi \in C_c^\infty(U)$, then f agrees with some $g \in H_s$ on a neighborhood of $\operatorname{supp}(\phi)$; hence $\phi f = \phi g \in H_s$ by Corollary 47. For the converse, given a precompact open V with $\overline{V} \subset U$, we can choose $\phi \in C_c^\infty(U)$ with $\overline{\phi} = 1$ on a neighborhood of $\overline{V}$ by the C^∞ Urysohn lemma; then $\phi f \in H_s$ and $\phi f = f$ on V. (We have implicitly used Proposition 110 to obtain compact neighborhoods of $\operatorname{supp}\phi$ and $\overline{V}$ in U.)

We conclude this section with one of the classic applications of Sobolev spaces, a regularity theorem for certain partial differential operators.

If $L = \sum_0^m a_j(d/dx)^j$ is an ordinary differential operator with C^∞ coefficients such that a_m never vanishes, it is not hard to show that smooth data give smooth solutions. More precisely, if $Lu = f$ and f is C^k on an open interval I, then u is C^{k+m} on I. No such result holds for partial differential operators in general. For example, for any $f \in L^1_{\text{loc}}(\mathbb{R})$ the function $u(x,t) = f(x-t)$ satisfies the wave equation $(\partial_t^2 - \partial_x^2)u = 0$, but u has only as much smoothness as f. However, there is a large class of differential operators for which a strong regularity theorem holds. We restrict attention to the constant-coefficient case, although the results are valid in greater generality.

Let $P(D) = \sum_{|\alpha|\leqslant m} c_\alpha D^\alpha$ (notation as in Folland Section 8.7) be a constant-coefficient operator. We assume that m is the true order of $P(D)$, i.e., that $c_\alpha \neq 0$ for some α with $|\alpha| = m$. The principal symbol P_m is the sum of the top-order terms in its symbol:

$$P_m(\xi) = \sum_{|\alpha|=m} c_\alpha \xi^\alpha$$

$P(D)$ is called elliptic if $P_m(\xi) \neq 0$ for all nonzero $\xi \in \mathbb{R}^n$. Thus, ellipticity means that, in a formal sense, $P(D)$ is genuinely mth order in all directions. (For example, the Laplacian Δ is elliptic on $\mathbb{R}^n$, whereas the heat and wave operators $\partial_t - \Delta$ and $\partial_t^2 - \Delta$ are not elliptic on $\mathbb{R}^{n+1}$.)

Lemma 9.50: 9.24.

Suppose that $P(D)$ is of order m. Then $P(D)$ is elliptic if and only if there exist $C, R > 0$ such that $|P(\xi)| \geqslant C|\xi|^m$ when $|\xi| \geqslant R$.

Proof. If $P(D)$ is elliptic, let C_1 be the minimum value of the principal symbol P_m on the unit sphere $|\xi| = 1$. Then $C_1 > 0$, and since P_m is homogeneous of degree m, we have $|P_m(\xi)| \geqslant C_1|\xi|^m$ for all ξ. On the other hand, $P - P_m$ is of order $m-1$, so there exists C_2 such that $|P(\xi) - P_m(\xi)| \leqslant C_2|\xi|^{m-1}$. Therefore,

$$|P(\xi)| \geqslant |P_m(\xi)| - |P(\xi) - P_m(\xi)| \geqslant \frac{1}{2}C_1|\xi|^m \text{ for } |\xi| \geqslant 2C_2C_1^{-1}$$

Conversely, if $P(D)$ is not elliptic, say $P_m(\xi_0) = 0$, then $|P(\xi)| \leqslant C|\xi|^{m-1}$ for every scalar multiple ξ of ξ_0.

Lemma 9.51: 9.25.

If $P(D)$ is elliptic of order $m, u \in H_s$, and $P(D)u \in H_s$, then $u \in H_{s+m}$.

Proof. The hypotheses say that $(1 + |\xi|^2)^{s/2}\widehat{u} \in L^2$ and $(1 + |\xi|^2)^{s/2}P\widehat{u} \in L^2$. By Lemma 50, for some $R \geqslant 1$ we have

$$(1 + |\xi|^2)^{m/2} \leqslant 2^m|\xi|^m \leqslant C^{-1}2^m|P(\xi)| \text{ for } |\xi| \geqslant R$$

and $(1 + |\xi|^2)^{m/2} \leqslant (1 + R^2)^{m/2}$ for $|\xi| \leqslant R$. It follows that

$$(1 + |\xi|^2)^{(s+m)/2}|\widehat{u}| \leqslant C'(1 + |\xi|^2)^{s/2}(|P\widehat{u}| + |\widehat{u}|) \in L^2$$

that is, $u \in H_{s+m}$.

Theorem 9.52: 9.26: The Elliptic Regularity Theorem.

Suppose that L is a constant-coefficient elliptic differential operator of order m, Ω is an open set in $\mathbb{R}^n$, and $u \in \mathcal{D}'(\Omega)$. If $Lu \in H_s^{\text{loc}}(\Omega)$ for some $s \in \mathbb{R}$, then $u \in H_{s+m}^{\text{loc}}(\Omega)$; and if $Lu \in C^\infty(\Omega)$, then $u \in C^\infty(\Omega)$.

Proof. The second assertion follows from the first in view of Corollary 44, so by Proposition 49 we must show that if $Lu \in H_s^{\text{loc}}(\Omega)$ and $\phi \in C_c^\infty(\Omega)$, then $\phi u \in H_{s+m}$. Let V be a precompact open set such that $\text{supp}(\phi) \subset V \subset \overline{V} \subset \Omega$, and choose $\psi \in C_c^\infty(\Omega)$ such that $\psi = 1$ on $\overline{V}$. Then $\psi u \in \mathcal{E}'$, so it follows from Proposition 25 that $\psi u \in H_\sigma$ for some $\sigma \in \mathbb{R}$. By decreasing σ we may assume that $s + m - \sigma$ is a positive integer k.

Set $\psi_0 = \psi$ and $\psi_k = \phi$, and choose recursively $\psi_1, \dots, \psi_{k-1} \in C_c^\infty$ such that $\psi_j = 1$ on a neighborhood of $\operatorname{supp}(\phi)$ and $\operatorname{supp}(\psi_j)$ is contained in the set where $\psi_{j-1} = 1$. We shall prove by induction that $\psi_j u \in H_{\sigma+j}$. When $j = k$, we obtain $\phi u = \psi_k u \in H_{\sigma+k} = H_m$, which will complete the proof.

The crucial observation is that for any $\zeta \in C_c^\infty$ the operator $[L, \zeta]$ defined by

$$[L, \zeta]f = L(\zeta f) - \zeta L f$$

is a differential operator of order $m - 1$ whose coefficients are linear combinations of derivatives of ζ; in particular, these coefficients are C^∞ functions that vanish on any open set where ζ is constant. (This follows from the product rule for derivatives.) Thus, if $f \in H_t$, we have $\partial^\alpha f \in H_{t-(m-1)}$ for $|\alpha| \leqslant m - 1$ and hence $[L, \zeta]f \in H_{t-(m-1)}$ by Theorem 46.

For $j = 0$ we have $\psi_0 u \in H_\sigma$ by assumption. Suppose we have established that $\psi_j u \in H_{\sigma+j}$, where $0 \leqslant j < k$. Then by the preceding remarks,

$$\begin{aligned} L(\psi_{j+1}u) = \psi_{j+1}Lu + [L, \psi_{j+1}]u &= \psi_{j+1}Lu + [L, \psi_{j+1}]\psi_j u \\ &\in H_s + H_{\sigma+j-(m-1)} = H_{\sigma+j+1-m} \end{aligned}$$

Since $\psi_{j+1}u = \psi_{j+1}\psi_j u \in H_{\sigma+j}$, Lemma 51 (with $P(D) = L$) implies that $\psi_{j+1}u \in H_{\sigma+j+1}$, and we are done.

Two classical special cases of this theorem are particularly noteworthy. First, every distribution solution of Laplace's equation $\Delta u = 0$ is a C^∞ function. (This fact is known as Weyl's lemma.) Second, if $L = \partial_1 + i\partial_2$ on $\mathbb{R}^2$, the equation $Lu = 0$ is the Cauchy-Riemann equation, whose solutions are the holomorphic (or analytic) functions of $z = x_1 + ix_2$. We thus recover the fact that holomorphic functions are C^∞.

Exercise 9.53: Folland Exercise 9.30.

Let $f_s(\xi) = (1 + |\xi|^2)^{s/2}$. Then $|\partial^\alpha f_s(\xi)| \leqslant C_\alpha(1 + |\xi|)^{s-|\alpha|}$.

Exercise 9.54: Folland Exercise 9.31.

If $k \in \mathbb{Z}_{\geqslant 1}, H_k$ is the space of all $f \in L^2$ that possess strong L^2 derivatives $\partial^\alpha f$, as defined in Folland Exercise 8.8, for $|\alpha| \leqslant k$; and these strong derivatives coincide with the distribution derivatives.

Exercise 9.55: Folland Exercise 9.32.

Suppose $r < s < t$. For any $\varepsilon > 0$ there exists $C > 0$ such that $\|f\|_{(s)} \leqslant \varepsilon\|f\|_{(t)} + C\|f\|_{(r)}$ for all $f \in H_t$.

Exercise 9.56: Folland Exercise 9.33.

(Converse of the Sobolev Theorem) If $H_s \subset C_0^k$, then $s > k + \frac{1}{2}n$. (Use the closed graph theorem to show that the inclusion map $H_s \to C_0^k$ is continuous and hence that $\partial^\alpha \delta \in (H_s)^*$ for $|\alpha| \leqslant k$.)

Exercise 9.57: Folland Exercise 9.34.

(A Sharper Sobolev Theorem) For $0 < \alpha < 1$, let

$$\Lambda_\alpha(\mathbb{R}^n) = \left\{ f \in BC(\mathbb{R}^n) \,\middle|\, \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x-y|^\alpha} < \infty \right\}$$

(a) If $s = \frac{1}{2}n + \alpha$ where $0 < \alpha < 1$, then $\|\tau_x\delta - \tau_y\delta\|_{(-s)} \leqslant C_\alpha |x-y|^\alpha$. (We have $\widehat{\left(\tau_x\delta\right)}(\xi) = e^{-2\pi i \xi \cdot x}$. Write the integral defining $\|\tau_x\delta - \tau_y\delta\|_{(-s)}^2$ as the sum of the integrals over the regions $|\xi| \leqslant R$ and $|\xi| > R$, where $R = |x-y|^{-1}$, and use the mean value theorem to estimate $(\tau_x\delta - \tau_y\delta)$ on the first region.)

(b) If $s = \frac{1}{2}n + \alpha$ where $0 < \alpha < 1$, then $H_s \subset \Lambda_\alpha(\mathbb{R}^n)$.

(c) If $s = \frac{1}{2}n + k + \alpha$ where $k \in \mathbb{Z}_{\geqslant 1}$ and $0 < \alpha < 1$, then

$$H_s \subset \{ f \in C_0^k \mid \partial^\alpha f \in \Lambda_a(\mathbb{R}^n) \text{ for } |\alpha| \leqslant k \}.$$

Exercise 9.58: Folland Exercise 9.35.

The Sobolev theorem says that if $s > \frac{1}{2}n$, it makes sense to evaluate functions in H_s at a point. For $0 \leqslant s \leqslant \frac{1}{2}n$, functions in H_s are only defined a.e., but if $s > \frac{1}{2}k$ with $k < n$, it makes sense to restrict functions in H_s to subspaces of codimension k. More precisely, let us write $\mathbb{R}^n = \mathbb{R}^{n-k} \times \mathbb{R}^k$, $x = (y,z)$, $\xi = (\eta, \zeta)$, and define $R \colon \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^{n-k})$ by $Rf(y) = f(y, 0)$.

(a) $\widehat{(Rf)}(\eta) = \int \widehat{f}(\eta, \zeta) d\zeta$. (See Folland Exercise 8.20.)

(b) If $s > \frac{1}{2}k$,

$$|(\widehat{Rf}(\eta))|^2 \leqslant C_s (1 + |\eta|^2)^{(k/2)-s} \int |\widehat{f}(\eta,\zeta)|^2 (1 + |\eta|^2 + |\zeta|^2)^s d\zeta$$

(c) R extends to a bounded map from $H_s(\mathbb{R}^n)$ to $H_{s-(k/2)}(\mathbb{R}^{n-k})$ provided $s > \frac{1}{2}k$.

Exercise 9.59: Folland Exercise 9.36.

Suppose that $0 \neq \phi \in C_c^\infty$ and $\{a_j\}$ is a sequence in $\mathbb{R}^n$ with $|a_j| \to \infty$, and let $\phi_j(x) = \phi(x - a_j)$. Then $\{\phi_j\}$ is bounded in H_s for every s but has no convergent subsequence in H_t for any t.

Exercise 9.60: Folland Exercise 9.37.

The heat operator $\partial_t - \Delta$ is not elliptic, but a weakened version of Theorem 52 holds for it. Here we are working on $\mathbb{R}^{n+1}$ with coordinates (x, t) and dual coordinates (ξ, τ), and $\partial_t - \Delta = P(D)$ where $P(\xi, \tau) = 2\pi i\tau + 4\pi^2|\xi|^2$.

(a) There exist $C, R > 0$ such that $|\xi||(\xi, \tau)|^{1/2} \leqslant C|P(\xi, \tau)|$ for $|(\xi, \tau)| > R$. (Consider the regions $|\tau| \leqslant |\xi|^2$ and $|\tau| \geqslant |\xi|^2$ separately.)

(b) If $f \in H_s$ and $(\partial_t - \Delta)f \in H_s$, then $f \in H_{s+1}$ and $\partial_{x_i} f \in H_{s+(1/2)}$ for $1 \leqslant i \leqslant n$.

(c) If $\zeta \in C_c^\infty(\mathbb{R}^{n+1})$, we have

$$[\partial_t - \Delta, \zeta]f = (\partial_t\zeta - \Delta\zeta)f - 2\sum(\partial_{x_i}\zeta)(\partial_{x_i} f)$$

(d) If Ω is open in $\mathbb{R}^{n+1}$, $u \in \mathcal{D}'(\Omega)$, and $(\partial_t - \Delta)u \in H_s^{\mathrm{loc}}(\Omega)$, then $u \in H_{s+1}^{\mathrm{loc}}(\Omega)$. (Let ψ_j be as in the proof of Theorem 52. Show inductively that if $\psi_0 u \in H_\sigma$, then $\psi_j u \in H_{\sigma+(j/2)}$ and $\partial_{x_i}(\psi_j u) \in H_{\sigma+(j-1)/2}$ provided $\sigma + \frac{1}{2}j \leqslant s$.)

Exercise 9.61: Folland Exercise 9.38.

Suppose $s_0 \leqslant s_1$ and $t_0 \leqslant t_1$, and for $0 \leqslant \lambda \leqslant 1$ let

$$s_\lambda = (1-\lambda)s_0 + \lambda s_1, \quad t_\lambda = (1-\lambda)t_0 + \lambda t_1$$

If T is a bounded linear map from H_{s_0} to H_{t_0} whose restriction to H_{s_1} is bounded from H_{s_1} to H_{t_1}, then the restriction of T to H_{s_λ} is bounded from H_{s_λ} to H_{t_λ} for $0 \leqslant \lambda \leqslant 1$. (T is bounded from H_s to H_t if and only if $\Lambda_s T \Lambda_{-t}$ is bounded on L^2. Observe that Λ_z is well defined for all $z \in \mathbb{C}$ and Λ_z is unitary on every H_s if Re $z = 0$. Let $s(z) = (1-z)s_0 + zs_1, t(z) = (1-z)t_0 + zt_1$, and for $0 \leqslant \operatorname{Re} z \leqslant 1$ and $\phi, \psi \in \mathcal{S}$ let $F(z) = \int[\Lambda_{t(z)} T \Lambda_{-s(z)}\phi]\psi$. Apply the three lines lemma as in the proof of the Riesz-Thorin theorem.)

Exercise 9.62: Folland Exercise 9.39.

Let Ω be an open set in $\mathbb{R}^n$, and let $G\colon \Omega \to \mathbb{R}^n$ be a C^∞ diffeomorphism. For any $\phi \in C_c^\infty(G(\Omega))$, the map $Tf = (\phi f) \circ G$ is bounded on H_s for all s; consequently, $f \circ G \in H_s^{\mathrm{loc}}(\Omega)$ whenever $f \in H_s^{\mathrm{loc}}(G(\Omega))$. Proceed as follows:

(a) If $s = 0, 1, 2, \ldots$, use the chain rule and the fact that $f \in H_s$ if and only if $\partial^\alpha f \in L^2$ for $|\alpha| \leqslant s$.

(b) Use Folland Exercise 9.38 to obtain the result for all $s > 0$.

(c) For $s < 0$, use Proposition 42 and the fact that the transpose of T is another operator of the same type, namely, $T'f = (\psi f) \circ H$ where $H = G^{-1}$ and $\psi = (J\phi) \circ G$ with $J(x) = |\det D_x G|$.

Exercise 9.63: Folland Exercise 9.40.

State and prove analogues of the results in this section for the periodic Sobolev spaces

$$H_s(\mathbb{T}^n) = \left\{ f \in \mathcal{D}'(\mathbb{T}^n) \;\middle|\; \sum (1 + |\kappa|^2)^s |\widehat{f}(\kappa)|^2 < \infty \right\}$$

Alphabetical Index

References

[Fol99] Gerald B. Folland. *Real analysis. Modern techniques and their applications*. 2nd ed. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, 1999. ISBN: 0-471-31716-0.