An Introduction to Measure Theory

Terence Tao

To Garth Gaudry, who set me on the road;

To my family, for their constant support;

And to the readers of my blog, for their feedback and contributions.

Contents

Preface

In the fall of 2010, I taught an introductory one-quarter course on graduate real analysis, focusing in particular on the basics of measure and integration theory, both in Euclidean spaces and in abstract measure spaces. This text is based on my lecture notes of that course, which are also available online on my blog `terrytao.wordpress.com`, together with some supplementary material, such as a section on problem solving strategies in real analysis (Section 2.1) which evolved from discussions with my students.

This text is intended to form a prequel to my graduate text [**Ta2010**] (henceforth referred to as *An epsilon of room, Vol. I*), which is an introduction to the analysis of Hilbert and Banach spaces (such as L^p and Sobolev spaces), point-set topology, and related topics such as Fourier analysis and the theory of distributions; together, they serve as a text for a complete first-year graduate course in real analysis.

The approach to measure theory here is inspired by the text [**StSk2005**], which was used as a secondary text in my course. In particular, the first half of the course is devoted almost exclusively to measure theory on Euclidean spaces $\mathbf{R}^d$ (starting with the more elementary Jordan-Riemann-Darboux theory, and only then moving on to the more sophisticated Lebesgue theory), deferring the abstract aspects of measure theory to the second half of the course. I found that this approach strengthened the student's intuition in the early stages of the course, and helped provide motivation for more abstract constructions, such as Carathéodory's general construction of a measure from an outer measure.

Most of the material here is self-contained, assuming only an undergraduate knowledge in real analysis (and in particular, on the Heine-Borel theorem, which we will use as the foundation for our construction of Lebesgue measure); a secondary real analysis text can be used in conjunction with this one, but it is not strictly necessary. A small number of exercises however will require some knowledge of point-set topology or of set-theoretic concepts such as cardinals and ordinals.

A large number of exercises are interspersed throughout the text, and it is intended that the reader perform a significant fraction of these exercises while going through the text. Indeed, many of the key results and examples in the subject will in fact be presented through the exercises. In my own course, I used the exercises as the basis for the examination questions, and signalled this well in advance, to encourage the students to attempt as many of the exercises as they could as preparation for the exams.

The core material is contained in Chapter 1, and already comprises a full quarter's worth of material. Section 2.1 is a much more informal section than the rest of the book, focusing on describing problem solving strategies, either specific to real analysis exercises, or more generally applicable to a wider set of mathematical problems; this section evolved from various discussions with students throughout the course. The remaining three sections in Chapter 2 are optional topics, which require understanding of most of the material in Chapter 1 as a prerequisite (although Section 2.3 can be read after completing Section 1.4.

Notation

For reasons of space, we will not be able to define every single mathematical term that we use in this book. If a term is italicised for reasons other than emphasis or for definition, then it denotes a standard mathematical object, result, or concept, which can be easily looked up in any number of references. (In the blog version of the book, many of these terms were linked to their Wikipedia pages, or other on-line reference pages.)

Given a subset E of a space X, the *indicator function* $1_E : X \to \mathbf{R}$ is defined by setting $1_E(x)$ equal to 1 for $x \in E$ and equal to 0 for $x \notin E$.

For any natural number d, we refer to the vector space $\mathbf{R}^d := \{(x_1, \ldots, x_d) : x_1, \ldots, x_d \in \mathbf{R}\}$ as *(d-dimensional) Euclidean space.* A vector $(x_1, \ldots, x_d)$ in $\mathbf{R}^d$ has length

$$|(x_1, \ldots, x_d)| := (x_1^2 + \ldots + x_d^2)^{1/2}$$

and two vectors $(x_1, \ldots, x_d), (y_1, \ldots, y_d)$ have *dot product*

$$(x_1, \ldots, x_d) \cdot (y_1, \ldots, y_d) := x_1 y_1 + \ldots + x_d y_d.$$

The *extended non-negative real axis* $[0, +\infty]$ is the non-negative real axis $[0, +\infty) := \{x \in \mathbf{R} : x \geq 0\}$ with an additional element adjointed to it, which we label $+\infty$; we will need to work with this system because many sets (e.g. $\mathbf{R}^d$) will have infinite measure. Of course, $+\infty$ is not a real number, but we think of it as an *extended* real number. We extend the addition, multiplication, and order structures on $[0, +\infty)$ to $[0, +\infty]$ by declaring

$$+\infty + x = x + +\infty = +\infty$$

for all $x \in [0, +\infty]$,

$$+\infty \cdot x = x \cdot +\infty = +\infty$$

for all non-zero $x \in (0, +\infty]$,

$$+\infty \cdot 0 = 0 \cdot +\infty = 0,$$

and

$$x < +\infty \text{ for all } x \in [0, +\infty).$$

Most of the laws of algebra for addition, multiplication, and order continue to hold in this extended number system; for instance addition and multiplication are commutative and associative, with the latter distributing over the former, and an order relation $x \leq y$ is preserved under addition or multiplication of both sides of that relation by the same quantity. However, we caution that the laws of cancellation do *not* apply once some of the variables are allowed to be infinite; for instance, we cannot deduce $x = y$ from $+\infty + x = +\infty + y$ or from $+\infty \cdot x = +\infty \cdot y$. This is related to the fact that the forms $+\infty - +\infty$ and $+\infty / +\infty$ are indeterminate (one cannot assign a value to them without breaking a lot of the rules of algebra). A general rule of thumb is that if one wishes to use cancellation (or proxies for cancellation, such as subtraction or division), this is only safe if one can guarantee that all quantities involved are finite (and in the case of multiplicative cancellation, the quantity being cancelled also needs to be non-zero, of course). However, as long as one avoids using cancellation and works exclusively with non-negative quantities, there is little danger in working in the extended real number system.

We note also that once one adopts the convention $+\infty \cdot 0 = 0 \cdot +\infty = 0$, then multiplication becomes *upward continuous* (in the sense that whenever $x_n \in [0, +\infty]$ increases to $x \in [0, +\infty]$, and $y_n \in [0, +\infty]$ increases to $y \in [0, +\infty]$, then $x_n y_n$ increases to xy) but not *downward continuous* (e.g. $1/n \to 0$ but $1/n \cdot +\infty \not\to 0 \cdot +\infty$). This asymmetry will ultimately cause us to define integration from below rather than from above, which leads to other asymmetries (e.g. the monotone convergence theorem (Theorem 1.4.44) applies for monotone increasing functions, but not necessarily for monotone decreasing ones).

Remark 0.0.1. Note that there is a tradeoff here: if one wants to keep as many useful laws of algebra as one can, then one can add in infinity, or have negative numbers, but it is difficult to have both at the same time. Because of this tradeoff, we will see two overlapping types of measure and integration theory: the *non-negative* theory, which involves quantities taking values in $[0, +\infty]$, and the *absolutely integrable* theory, which involves quantities taking values in $(-\infty, +\infty)$ or $\mathbf{C}$. For instance, the fundamental convergence theorem for the former theory is the monotone convergence theorem (Theorem 1.4.44), while the fundamental convergence theorem for the latter is the dominated convergence theorem (Theorem 1.4.49). Both branches of the theory are important, and both will be covered in later notes.

One important feature of the extended nonnegative real axis is that all sums are convergent: given any sequence $x_1, x_2, \ldots \in [0, +\infty]$, we can always form the sum

$$\sum_{n=1}^{\infty} x_n \in [0, +\infty]$$

as the limit of the partial sums $\sum_{n=1}^{N} x_n$, which may be either finite or infinite. An equivalent definition of this infinite sum is as the supremum of all finite subsums:

$$\sum_{n=1}^{\infty} x_n = \sup_{F \subset \mathbf{N}, F \text{ finite}} \sum_{n \in F} x_n.$$

Motivated by this, given any collection $(x_\alpha)_{\alpha \in A}$ of numbers $x_\alpha \in [0, +\infty]$ indexed by an arbitrary set A (finite or infinite, countable or uncountable), we can define the sum $\sum_{\alpha \in A} x_\alpha$ by the formula

$$\sum_{\alpha \in A} x_\alpha = \sup_{F \subset A, F \text{ finite}} \sum_{\alpha \in F} x_\alpha. \tag{0.1}$$

Note from this definition that one can relabel the collection in an arbitrary fashion without affecting the sum; more precisely, given any bijection $\phi : B \to A$, one has the change of variables formula

$$\sum_{\alpha \in A} x_\alpha = \sum_{\beta \in B} x_{\phi(\beta)}. \tag{0.2}$$

Note that when dealing with signed sums, the above rearrangement identity can fail when the series is not absolutely convergent (cf. the *Riemann rearrangement theorem*).

Exercise 0.0.1. If $(x_\alpha)_{\alpha \in A}$ is a collection of numbers $x_\alpha \in [0, +\infty]$ such that $\sum_{\alpha \in A} x_\alpha < \infty$, show that $x_\alpha = 0$ for all but at most countably many $\alpha \in A$, even if A itself is uncountable.

We will rely frequently on the following basic fact (a special case of the *Fubini-Tonelli theorem*, Corollary 1.7.23):

Theorem 0.0.2 (Tonelli's theorem for series)**.** *Let $(x_{n,m})_{n,m \in \mathbf{N}}$ be a doubly infinite sequence of extended non-negative reals $x_{n,m} \in [0, +\infty]$. Then*

$$\sum_{(n,m) \in \mathbf{N}^2} x_{n,m} = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} x_{n,m} = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} x_{n,m}.$$

Informally, Tonelli's theorem asserts that we may rearrange infinite series with impunity as long as all summands are non-negative.

Proof. We shall just show the equality of the first and second expressions; the equality of the first and third is proven similarly.

We first show that

$$\sum_{(n,m)\in\mathbf{N}^2} x_{n,m} \leq \sum_{n=1}^\infty \sum_{m=1}^\infty x_{n,m}.$$

Let F be any finite subset of $\mathbf{N}^2$. Then $F \subset \{1,\dots,N\}\times\{1,\dots,N\}$ for some finite N, and thus (by the non-negativity of the $x_{n,m}$)

$$\sum_{(n,m)\in F} x_{n,m} \leq \sum_{(n,m)\in\{1,\dots,N\}\times\{1,\dots,N\}} x_{n,m}.$$

The right-hand side can be rearranged as

$$\sum_{n=1}^N \sum_{m=1}^N x_{n,m},$$

which is clearly at most $\sum_{n=1}^\infty \sum_{m=1}^\infty x_{n,m}$ (again by non-negativity of $x_{n,m}$). This gives

$$\sum_{(n,m)\in F} x_{n,m} \leq \sum_{n=1}^\infty \sum_{m=1}^\infty x_{n,m}.$$

for any finite subset F of $\mathbf{N}^2$, and the claim then follows from (0.1).

It remains to show the reverse inequality

$$\sum_{n=1}^\infty \sum_{m=1}^\infty x_{n,m} \leq \sum_{(n,m)\in\mathbf{N}^2} x_{n,m}.$$

It suffices to show that

$$\sum_{n=1}^N \sum_{m=1}^\infty x_{n,m} \leq \sum_{(n,m)\in\mathbf{N}^2} x_{n,m}$$

for each finite N.

Fix N. As each $\sum_{m=1}^\infty x_{n,m}$ is the limit of $\sum_{m=1}^M x_{n,m}$, the left-hand side is the limit of $\sum_{n=1}^N \sum_{m=1}^M x_{n,m}$ as $M \to \infty$. Thus it suffices to show that

$$\sum_{n=1}^{N}\sum_{m=1}^{M} x_{n,m} \leq \sum_{(n,m)\in\mathbf{N}^2} x_{n,m}$$

for each finite M. But the left-hand side is $\sum_{(n,m)\in\{1,\ldots,N\}\times\{1,\ldots,M\}} x_{n,m}$, and the claim follows. □

Remark 0.0.3. Note how important it was that the $x_{n,m}$ were non-negative in the above argument. In the signed case, one needs an additional assumption of absolute summability of $x_{n,m}$ on $\mathbf{N}^2$ before one is permitted to interchange sums; this is *Fubini's theorem for series*, which we will encounter later in this text. Without absolute summability or non-negativity hypotheses, the theorem can fail (consider for instance the case when $x_{n,m}$ equals $+1$ when $n = m$, -1 when $n = m + 1$, and 0 otherwise).

Exercise 0.0.2 (Tonelli's theorem for series over arbitrary sets)**.** Let A, B be sets (possibly infinite or uncountable), and $(x_{n,m})_{n\in A, m\in B}$ be a doubly infinite sequence of extended non-negative reals $x_{n,m} \in [0, +\infty]$ indexed by A and B. Show that

$$\sum_{(n,m)\in A\times B} x_{n,m} = \sum_{n\in A}\sum_{m\in B} x_{n,m} = \sum_{m\in B}\sum_{n\in A} x_{n,m}.$$

(*Hint:* although not strictly necessary, you may find it convenient to first establish the fact that if $\sum_{n\in A} x_n$ is finite, then x_n is non-zero for at most countably many n.)

Next, we recall the *axiom of choice*, which we shall be assuming throughout the text:

Axiom 0.0.4 (Axiom of choice)**.** *Let $(E_\alpha)_{\alpha\in A}$ be a family of non-empty sets E_α, indexed by an index set A. Then we can find a family $(x_\alpha)_{\alpha\in A}$ of elements x_α of E_α, indexed by the same set A.*

This axiom is trivial when A is a singleton set, and from mathematical induction one can also prove it without difficulty when A is finite. However, when A is infinite, one cannot deduce this axiom from the other axioms of set theory, but must explicitly add it to the list of axioms. We isolate the countable case as a particularly useful corollary (though one which is strictly weaker than the full axiom of choice):

Corollary 0.0.5 (Axiom of countable choice)**.** *Let $E_1, E_2, E_3, \ldots$ be a sequence of non-empty sets. Then one can find a sequence $x_1, x_2, \ldots$ such that $x_n \in E_n$ for all $n = 1, 2, 3, \ldots$.*

Remark 0.0.6. The question of how much of real analysis still survives when one is not permitted to use the axiom of choice is a delicate one, involving a fair amount of logic and descriptive set theory to answer. We will not discuss these matters in this text. We will however note a theorem of Gödel[**Go1938**] that states that any statement that can be phrased in the first-order language of *Peano arithmetic*, and which is proven with the axiom of choice, can also be proven without the axiom of choice. So, roughly speaking, Gödel's theorem tells us that for any "finitary" application of real analysis (which includes most of the "practical" applications of the subject), it is safe to use the axiom of choice; it is only when asking questions about "infinitary" objects that are beyond the scope of Peano arithmetic that one can encounter statements that are provable using the axiom of choice, but are not provable without it.

Acknowledgments

This text was strongly influenced by the real analysis text of Stein and Shakarchi[**StSk2005**], which was used as a secondary text when teaching the course on which these notes were based. In particular, the strategy of focusing first on Lebesgue measure and Lebesgue integration, before moving onwards to abstract measure and integration theory, was directly inspired by the treatment in [**StSk2005**], and the material on differentiation theorems also closely follows that in [**StSk2005**]. On the other hand, our discussion here differs from that in [**StSk2005**] in other respects; for instance, a far greater emphasis is placed on Jordan measure and the Riemann integral as being an elementary precursor to Lebesgue measure and the Lebesgue integral.

I am greatly indebted to my students of the course on which this text was based, as well as many further commenters on my blog, including Maria Alfonseca, Marco Angulo, J. Balachandran,

Farzin Barekat, Marek Bernát, Lewis Bowen, Chris Breeden, Danny Calegari, Yu Cao, Chandrasekhar, David Chang, Nick Cook, Damek Davis, Eric Davis, Marton Eekes, Wenying Gan, Nick Gill, Ulrich Groh, Tim Gowers, Laurens Gunnarsen, Tobias Hagge, Xueping Huang, Bo Jacoby, Apoorva Khare, Shiping Liu, Colin McQuillan, David Milovich, Hossein Naderi, Brent Nelson, Constantin Niculescu, Mircea Petrache, Walt Pohl, Jim Ralston, David Roberts, Mark Schwarzmann, Vladimir Slepnev, David Speyer, Blake Stacey, Tim Sullivan, Jonathan Weinstein, Duke Zhang, Lei Zhang, Pavel Zorin, and several anonymous commenters, for providing corrections and useful commentary on the material here. These comments can be viewed online at
`terrytao.wordpress.com/category/teaching/245a-real-analysis`

The author is supported by a grant from the MacArthur Foundation, by NSF grant DMS-0649473, and by the NSF Waterman award.

Chapter 1

Measure theory

1.1. Prologue: The problem of measure

One of the most fundamental concepts in Euclidean geometry is that of the *measure* $m(E)$ of a solid body E in one or more dimensions. In one, two, and three dimensions, we refer to this measure as the *length*, *area*, or *volume* of E respectively. In the classical approach to geometry, the measure of a body was often computed by partitioning that body into finitely many components, moving around each component by a rigid motion (e.g. a translation or rotation), and then reassembling those components to form a simpler body which presumably has the same area. One could also obtain lower and upper bounds on the measure of a body by computing the measure of some inscribed or circumscribed body; this ancient idea goes all the way back to the work of Archimedes at least. Such arguments can be justified by an appeal to geometric intuition, or simply by postulating the existence of a measure $m(E)$ that can be assigned to all solid bodies E, and which obeys a collection of geometrically reasonable axioms. One can also justify the concept of measure on "physical" or "reductionistic" grounds, viewing the measure of a macroscopic body as the sum of the measures of its microscopic components.

With the advent of *analytic geometry*, however, Euclidean geometry became reinterpreted as the study of Cartesian products $\mathbf{R}^d$ of the real line $\mathbf{R}$. Using this analytic foundation rather than the classical geometrical one, it was no longer intuitively obvious how to define the measure $m(E)$ of a general[1] subset E of $\mathbf{R}^d$; we will refer to this (somewhat vaguely defined) problem of writing down the "correct" definition of measure as the *problem of measure*.

To see why this problem exists at all, let us try to formalise some of the intuition for measure discussed earlier. The physical intuition of defining the measure of a body E to be the sum of the measure of its component "atoms" runs into an immediate problem: a typical solid body would consist of an infinite (and uncountable) number of points, each of which has a measure of zero; and the product $\infty \cdot 0$ is indeterminate. To make matters worse, two bodies that have exactly

[1]One can also pose the problem of measure on other domains than Euclidean space, such as a Riemannian manifold, but we will focus on the Euclidean case here for simplicity, and refer to any text on Riemannian geometry for a treatment of integration on manifolds.

the same number of points, need not have the same measure. For instance, in one dimension, the intervals $A := [0, 1]$ and $B := [0, 2]$ are in one-to-one correspondence (using the bijection $x \mapsto 2x$ from A to B), but of course B is twice as long as A. So one can disassemble A into an uncountable number of points and reassemble them to form a set of twice the length.

Of course, one can point to the infinite (and uncountable) number of components in this disassembly as being the cause of this breakdown of intuition, and restrict attention to just finite partitions. But one still runs into trouble here for a number of reasons, the most striking of which is the *Banach-Tarski paradox*, which shows that the unit ball $B := \{(x, y, z) \in \mathbf{R}^3 : x^2 + y^2 + z^2 \leq 1\}$ in three dimensions[2] can be disassembled into a finite number of pieces (in fact, just five pieces suffice), which can then be reassembled (after translating and rotating each of the pieces) to form two disjoint copies of the ball B.

Here, the problem is that the pieces used in this decomposition are highly pathological in nature; among other things, their construction requires use of the *axiom of choice*. (This is in fact necessary; there are models of set theory without the axiom of choice in which the Banach-Tarski paradox does not occur, thanks to a famous theorem of Solovay[**So1970**].) Such pathological sets almost never come up in practical applications of mathematics. Because of this, the standard solution to the problem of measure has been to abandon the goal of measuring *every* subset E of $\mathbf{R}^d$, and instead to settle for only measuring a certain subclass of "non-pathological" subsets of $\mathbf{R}^d$, which are then referred to as the *measurable sets*. The problem of measure then divides into several subproblems:

(i) What does it mean for a subset E of $\mathbf{R}^d$ to be measurable?

(ii) If a set E is measurable, how does one define its measure?

(iii) What nice properties or axioms does measure (or the concept of measurability) obey?

[2]The paradox only works in three dimensions and higher, for reasons having to do with the group-theoretic property of *amenability*; see §2.2 of *An epsilon of room, Vol. I* for further discussion.

(iv) Are "ordinary" sets such as cubes, balls, polyhedra, etc. measurable?

(v) Does the measure of an "ordinary" set equal the "naive geometric measure" of such sets? (e.g. is the measure of an $a \times b$ rectangle equal to ab?)

These questions are somewhat open-ended in formulation, and there is no unique answer to them; in particular, one can expand the class of measurable sets at the expense of losing one or more nice properties of measure in the process (e.g. finite or countable additivity, translation invariance, or rotation invariance). However, there are two basic answers which, between them, suffice for most applications. The first is the concept of *Jordan measure* (or *Jordan content*) of a Jordan measurable set, which is a concept closely related to that of the *Riemann integral* (or *Darboux integral*). This concept is elementary enough to be systematically studied in an undergraduate analysis course, and suffices for measuring most of the "ordinary" sets (e.g. the area under the graph of a continuous function) in many branches of mathematics. However, when one turns to the type of sets that arise in *analysis*, and in particular those sets that arise as *limits* (in various senses) of other sets, it turns out that the Jordan concept of measurability is not quite adequate, and must be extended to the more general notion of *Lebesgue measurability*, with the corresponding notion of *Lebesgue measure* that extends Jordan measure. With the Lebesgue theory (which can be viewed as a *completion* of the Jordan-Darboux-Riemann theory), one keeps almost all of the desirable properties of Jordan measure, but with the crucial additional property that many features of the Lebesgue theory are preserved under limits (as exemplified in the fundamental *convergence theorems* of the Lebesgue theory, such as the *monotone convergence theorem* (Theorem 1.4.44) and the *dominated convergence theorem* (Theorem 1.4.49), which do not hold in the Jordan-Darboux-Riemann setting).

As such, they are particularly well suited[3] for applications in analysis, where limits of functions or sets arise all the time.

In later sections, we will formally define Lebesgue measure and the Lebesgue integral, as well as the more general concept of an abstract measure space and the associated integration operation. In the rest of the current section, we will discuss the more elementary concepts of Jordan measure and the Riemann integral. This material will eventually be superceded by the more powerful theory to be treated in later sections; but it will serve as motivation for that later material, as well as providing some continuity with the treatment of measure and integration in undergraduate analysis courses.

1.1.1. Elementary measure.

Before we discuss Jordan measure, we discuss the even simpler notion of *elementary measure*, which allows one to measure a very simple class of sets, namely the *elementary sets* (finite unions of boxes).

Definition 1.1.1 (Intervals, boxes, elementary sets)**.** An *interval* is a subset of $\mathbf{R}$ of the form $[a,b] := \{x \in \mathbf{R} : a \leq x \leq b\}$, $[a,b) := \{x \in \mathbf{R} : a \leq x < b\}$, $(a,b] := \{x \in \mathbf{R} : a < x \leq b\}$, or $(a,b) := \{x \in \mathbf{R} : a < x < b\}$, where $a \leq b$ are real numbers. We define the *length*[4] $|I|$ of an interval $I = [a,b], [a,b), (a,b], (a,b)$ to be $|I| := b - a$. A *box* in $\mathbf{R}^d$ is a Cartesian product $B := I_1 \times \ldots \times I_d$ of d intervals $I_1, \ldots, I_d$ (not necessarily of the same length), thus for instance an interval is a one-dimensional box. The *volume* $|B|$ of such a box B is defined as $|B| := |I_1| \times \ldots \times |I_d|$. An *elementary set* is any subset of $\mathbf{R}^d$ which is the union of a finite number of boxes.

Exercise 1.1.1 (Boolean closure)**.** Show that if $E, F \subset \mathbf{R}^d$ are elementary sets, then the union $E \cup F$, the intersection $E \cap F$, and the set theoretic difference $E \backslash F := \{x \in E : x \not\in F\}$, and the symmetric difference $E \Delta F := (E \backslash F) \cup (F \backslash E)$ are also elementary. If $x \in \mathbf{R}^d$, show that the translate $E + x := \{y + x : y \in E\}$ is also an elementary set.

[3] There are other ways to extend Jordan measure and the Riemann integral, see for instance Exercise 1.6.53 or Section 1.7.3, but the Lebesgue approach handles limits and rearrangement better than the other alternatives, and so has become the standard approach in analysis; it is also particularly well suited for providing the rigorous foundations of probability theory, as discussed in Section 2.3.

[4] Note we allow degenerate intervals of zero length.

We now give each elementary set a measure.

Lemma 1.1.2 (Measure of an elementary set)**.** *Let $E \subset \mathbf{R}^d$ be an elementary set.*

(i) *E can be expressed as the finite union of* disjoint *boxes.*

(ii) *If E is partitioned as the finite union $B_1 \cup \ldots \cup B_k$ of disjoint boxes, then the quantity $m(E) := |B_1| + \ldots + |B_k|$ is independent of the partition. In other words, given any other partition $B'_1 \cup \ldots \cup B'_{k'}$ of E, one has $|B_1| + \ldots + |B_k| = |B'_1| + \ldots + |B'_{k'}|$.*

We refer to $m(E)$ as the elementary measure *of E. (We occasionally write $m(E)$ as $m^d(E)$ to emphasise the d-dimensional nature of the measure.) Thus, for example, the elementary measure of $(1,2) \cup [3,6]$ is $(2-1) + (6-3) = 4$.*

Proof. We first prove (i) in the one-dimensional case $d = 1$. Given any finite collection of intervals $I_1, \ldots, I_k$, one can place the $2k$ endpoints of these intervals in increasing order (discarding repetitions). Looking at the open intervals between these endpoints, together with the endpoints themselves (viewed as intervals of length zero), we see that there exists a finite collection of disjoint intervals $J_1, \ldots, J_{k'}$ such that each of the $I_1, \ldots, I_k$ are a union of some subcollection of the $J_1, \ldots, J_{k'}$. This already gives (i) when $d = 1$. To prove the higher dimensional case, we express E as the union $B_1, \ldots, B_k$ of boxes $B_i = I_{i,1} \times \ldots \times I_{i,d}$. For each $j = 1, \ldots, d$, we use the one-dimensional argument to express $I_{1,j}, \ldots, I_{k,j}$ as the union of subcollections of a collection $J_{1,j}, \ldots, J_{k'_j,j}$ of disjoint intervals. Taking Cartesian products, we can express the $B_1, \ldots, B_k$ as finite unions of boxes $J_{i_1,1} \times \ldots \times J_{i_d,d}$, where $1 \leq i_j \leq k'_j$ for all $1 \leq j \leq d$. Such boxes are all disjoint, and the claim follows.

To prove (ii) we use a discretisation argument. Observe (exercise!) that for any interval I, the length of I can be recovered by the limiting formula

$$|I| = \lim_{N \to \infty} \frac{1}{N} \#(I \cap \frac{1}{N}\mathbf{Z})$$

where $\frac{1}{N}\mathbf{Z} := \{\frac{n}{N} : n \in \mathbf{Z}\}$ and $\#A$ denotes the cardinality of a finite set A. Taking Cartesian products, we see that

$$|B| = \lim_{N\to\infty} \frac{1}{N^d} \#(B \cap \frac{1}{N}\mathbf{Z}^d)$$

for any box B, and in particular that

$$|B_1| + \ldots + |B_k| = \lim_{N\to\infty} \frac{1}{N^d} \#(E \cap \frac{1}{N}\mathbf{Z}^d).$$

Denoting the right-hand side as $m(E)$, we obtain the claim (ii). □

Exercise 1.1.2. Give an alternate proof of Lemma 1.1.2(ii) by showing that any two partitions of E into boxes admit a mutual refinement into boxes that arise from taking Cartesian products of elements from finite collections of disjoint intervals.

Remark 1.1.3. One might be tempted to now define the measure $m(E)$ of an arbitrary set $E \subset \mathbf{R}^d$ by the formula

$$m(E) := \lim_{N\to\infty} \frac{1}{N^d} \#(E \cap \frac{1}{N}\mathbf{Z}^d), \tag{1.1}$$

since this worked well for elementary sets. However, this definition is not particularly satisfactory for a number of reasons. Firstly, one can concoct examples in which the limit does not exist (Exercise!). Even when the limit does exist, this concept does not obey reasonable properties such as translation invariance. For instance, if $d = 1$ and $E := \mathbf{Q} \cap [0, 1] := \{x \in \mathbf{Q} : 0 \leq x \leq 1\}$, then this definition would give E a measure of 1, but would give the translate $E + \sqrt{2} := \{x + \sqrt{2} : x \in \mathbf{Q}; 0 \leq x \leq 1\}$ a measure of zero. Nevertheless, the formula (1.1) will be valid for all Jordan measurable sets (see Exercise 1.1.13). It also makes precise an important intuition, namely that the continuous concept of measure can be viewed[5] as a limit of the discrete concept of (normalised) cardinality.

From the definitions, it is clear that $m(E)$ is a non-negative real number for every elementary set E, and that

$$m(E \cup F) = m(E) + m(F)$$

[5]Another way to obtain continuous measure as the limit of discrete measure is via *Monte Carlo integration*, although in order to rigorously introduce the probability theory needed to set up Monte Carlo integration properly, one already needs to develop a large part of measure theory, so this perspective, while intuitive, is not suitable for foundational purposes.

whenever E and F are disjoint elementary sets. We refer to the latter property as *finite additivity*; by induction it also implies that

$$m(E_1 \cup \ldots \cup E_k) = m(E_1) + \ldots + m(E_k)$$

whenever $E_1, \ldots, E_k$ are disjoint elementary sets. We also have the obvious degenerate case

$$m(\emptyset) = 0.$$

Finally, elementary measure clearly extends the notion of volume, in the sense that

$$m(B) = |B|$$

for all boxes B.

From non-negativity and finite additivity (and Exercise 1.1.1) we conclude the *monotonicity property*

$$m(E) \leq m(F)$$

whenever $E \subset F$ are nested elementary sets. From this and finite additivity (and Exercise 1.1.1) we easily obtain the *finite subadditivity property*

$$m(E \cup F) \leq m(E) + m(F)$$

whenever E, F are elementary sets (not necessarily disjoint); by induction one then has

$$m(E_1 \cup \ldots \cup E_k) \leq m(E_1) + \ldots + m(E_k)$$

whenever $E_1, \ldots, E_k$ are elementary sets (not necessarily disjoint).

It is also clear from the definition that we have the *translation invariance*

$$m(E + x) = m(E)$$

for all elementary sets E and $x \in \mathbf{R}^d$.

These properties in fact define elementary measure up to normalisation:

Exercise 1.1.3 (Uniqueness of elementary measure)**.** Let $d \geq 1$. Let $m' \colon \mathcal{E}(\mathbf{R}^d) \to \mathbf{R}^+$ be a map from the collection $\mathcal{E}(\mathbf{R}^d)$ of elementary subsets of $\mathbf{R}^d$ to the nonnegative reals that obeys the non-negativity, finite additivity, and translation invariance properties. Show that there exists a constant $c \in \mathbf{R}^+$ such that $m'(E) = cm(E)$ for all elementary sets E. In particular, if we impose the additional normalisation $m'([0,1)^d) = 1$, then $m' \equiv m$. (*Hint:* Set $c := m'([0,1)^d)$, and then compute $m'([0,\frac{1}{n})^d)$ for any positive integer n.)

Exercise 1.1.4. Let $d_1, d_2 \geq 1$, and let $E_1 \subset \mathbf{R}^{d_1}$, $E_2 \subset \mathbf{R}^{d_2}$ be elementary sets. Show that $E_1 \times E_2 \subset \mathbf{R}^{d_1+d_2}$ is elementary, and $m^{d_1+d_2}(E_1 \times E_2) = m^{d_1}(E_1) \times m^{d_2}(E_2)$.

1.1.2. Jordan measure. We now have a satisfactory notion of measure for elementary sets. But of course, the elementary sets are a very restrictive class of sets, far too small for most applications. For instance, a solid triangle or disk in the plane will not be elementary, or even a rotated box. On the other hand, as essentially observed long ago by Archimedes, such sets E can be *approximated* from within and without by elementary sets $A \subset E \subset B$, and the inscribing elementary set A and the circumscribing elementary set B can be used to give lower and upper bounds on the putative measure of E. As one makes the approximating sets A, B increasingly fine, one can hope that these two bounds eventually match. This gives rise to the following definitions.

Definition 1.1.4 (Jordan measure)**.** Let $E \subset \mathbf{R}^d$ be a bounded set.

- The *Jordan inner measure* $m_{*,(J)}(E)$ of E is defined as
$$m_{*,(J)}(E) := \sup_{A \subset E, A \text{ elementary}} m(A).$$
- The *Jordan outer measure* $m^{*,(J)}(E)$ of E is defined as
$$m^{*,(J)}(E) := \inf_{B \supset E, B \text{ elementary}} m(B).$$
- If $m_{*,(J)}(E) = m^{*,(J)}(E)$, then we say that E is *Jordan measurable*, and call $m(E) := m_{*,(J)}(E) = m^{*,(J)}(E)$ the *Jordan measure* of E. As before, we write $m(E)$ as $m^d(E)$ when we wish to emphasise the dimension d.

By convention, we do not consider unbounded sets to be Jordan measurable (they will be deemed to have infinite Jordan outer measure).

Jordan measurable sets are those sets which are "almost elementary" with respect to Jordan outer measure. More precisely, we have

Exercise 1.1.5 (Characterisation of Jordan measurability)**.** Let $E \subset \mathbf{R}^d$ be bounded. Show that the following are equivalent:

(1) E is Jordan measurable.

(2) For every $\varepsilon > 0$, there exist elementary sets $A \subset E \subset B$ such that $m(B \backslash A) \leq \varepsilon$.

(3) For every $\varepsilon > 0$, there exists an elementary set A such that $m^{*,(J)}(A \Delta E) \leq \varepsilon$.

As one corollary of this exercise, we see that every elementary set E is Jordan measurable, and that Jordan measure and elementary measure coincide for such sets; this justifies the use of $m(E)$ to denote both. In particular, we still have $m(\emptyset) = 0$.

Jordan measurability also inherits many of the properties of elementary measure:

Exercise 1.1.6. Let $E, F \subset \mathbf{R}^d$ be Jordan measurable sets.

(1) (Boolean closure) Show that $E \cup F$, $E \cap F$, $E \backslash F$, and $E \Delta F$ are Jordan measurable.

(2) (Non-negativity) $m(E) \geq 0$.

(3) (Finite additivity) If E, F are disjoint, then $m(E \cup F) = m(E) + m(F)$.

(4) (Monotonicity) If $E \subset F$, then $m(E) \leq m(F)$.

(5) (Finite subadditivity) $m(E \cup F) \leq m(E) + m(F)$.

(6) (Translation invariance) For any $x \in \mathbf{R}^d$, $E + x$ is Jordan measurable, and $m(E + x) = m(E)$.

Now we give some examples of Jordan measurable sets:

Exercise 1.1.7 (Regions under graphs are Jordan measurable)**.** Let B be a closed box in $\mathbf{R}^d$, and let $f \colon B \to \mathbf{R}$ be a continuous function.

(1) Show that the graph $\{(x, f(x)) : x \in B\} \subset \mathbf{R}^{d+1}$ is Jordan measurable in $\mathbf{R}^{d+1}$ with Jordan measure zero. (*Hint:* on a compact metric space, continuous functions are uniformly continuous.)

(2) Show that the set $\{(x, t) : x \in B; 0 \leq t \leq f(x)\} \subset \mathbf{R}^{d+1}$ is Jordan measurable.

Exercise 1.1.8. Let A, B, C be three points in $\mathbf{R}^2$.

(1) Show that the solid triangle with vertices A, B, C is Jordan measurable.

(2) Show that the Jordan measure of the solid triangle is equal to $\frac{1}{2}|(B-A) \wedge (C-A)|$, where $|(a,b) \wedge (c,d)| := |ad - bc|$.

(*Hint:* It may help to first do the case when one of the edges, say AB, is horizontal.)

Exercise 1.1.9. Show that every compact convex polytope[6] in $\mathbf{R}^d$ is Jordan measurable.

Exercise 1.1.10. (1) Show that all open and closed Euclidean balls $B(x,r) := \{y \in \mathbf{R}^d : |y-x| < r\}$, $\overline{B(x,r)} := \{y \in \mathbf{R}^d : |y-x| \leq r\}$ in $\mathbf{R}^d$ are Jordan measurable, with Jordan measure $c_d r^d$ for some constant $c_d > 0$ depending only on d.

(2) Establish the crude bounds

$$\left(\frac{2}{\sqrt{d}}\right)^d \leq c_d \leq 2^d.$$

(An exact formula for c_d is $c_d = \frac{1}{d}\omega_d$, where $\omega_d := \frac{2\pi^{d/2}}{\Gamma(d/2)}$ is the volume of the unit sphere $S^{d-1} \subset \mathbf{R}^d$ and Γ is the *Gamma function*, but we will not derive this formula here.)

Exercise 1.1.11. This exercise assumes familiarity with linear algebra. Let $L\colon \mathbf{R}^d \to \mathbf{R}^d$ be a linear transformation.

(1) Show that there exists a non-negative real number D such that $m(L(E)) = Dm(E)$ for every elementary set E (note from previous exercises that $L(E)$ is Jordan measurable). (*Hint:* apply Exercise 1.1.3 to the map $E \mapsto m(L(E))$.)

(2) Show that if E is Jordan measurable, then $L(E)$ is also, and $m(L(E)) = Dm(E)$.

---

[6]A *closed convex polytope* is a subset of $\mathbf{R}^d$ formed by intersecting together finitely many closed half-spaces of the form $\{x \in \mathbf{R}^d : x \cdot v \leq c\}$, where $v \in \mathbf{R}^d$, $c \in \mathbf{R}$, and $\cdot$ denotes the usual dot product on $\mathbf{R}^d$. A *compact convex polytope* is a closed convex polytope which is also bounded.

(3) Show that $D = |\det L|$. (*Hint:* Work first with the case when L is an elementary transformation, using Gaussian elimination. Alternatively, work with the cases when L is a diagonal transformation or an orthogonal transformation, using the unit ball in the latter case, and use the polar decomposition.)

Exercise 1.1.12. Define a *Jordan null set* to be a Jordan measurable set of Jordan measure zero. Show that any subset of a Jordan null set is a Jordan null set.

Exercise 1.1.13. Show that (1.1) holds for all Jordan measurable $E \subset \mathbf{R}^d$.

Exercise 1.1.14 (Metric entropy formulation of Jordan measurability)**.** Define a *dyadic cube* to be a half-open box of the form

$$\left[\frac{i_1}{2^n}, \frac{i_1+1}{2^n}\right) \times \ldots \times \left[\frac{i_d}{2^n}, \frac{i_d+1}{2^n}\right)$$

for some integers $n, i_1, \ldots, i_d$. Let $E \subset \mathbf{R}^d$ be a bounded set. For each integer n, let $\mathcal{E}_*(E, 2^{-n})$ denote the number of dyadic cubes of sidelength 2^{-n} that are contained in E, and let $\mathcal{E}^*(E, 2^{-n})$ be the number of dyadic cubes[7] of sidelength 2^{-n} that intersect E. Show that E is Jordan measurable if and only if

$$\lim_{n\to\infty} 2^{-dn}(\mathcal{E}^*(E, 2^{-n}) - \mathcal{E}_*(E, 2^{-n})) = 0,$$

in which case one has

$$m(E) = \lim_{n\to\infty} 2^{-dn}\mathcal{E}_*(E, 2^{-n}) = \lim_{n\to\infty} 2^{-dn}\mathcal{E}^*(E, 2^{-n}).$$

Exercise 1.1.15 (Uniqueness of Jordan measure)**.** Let $d \geq 1$. Let $m' \colon \mathcal{J}(\mathbf{R}^d) \to \mathbf{R}^+$ be a map from the collection $\mathcal{J}(\mathbf{R}^d)$ of Jordan-measurable subsets of $\mathbf{R}^d$ to the nonnegative reals that obeys the non-negativity, finite additivity, and translation invariance properties. Show that there exists a constant $c \in \mathbf{R}^+$ such that $m'(E) = cm(E)$ for all Jordan measurable sets E. In particular, if we impose the additional normalisation $m'([0,1)^d) = 1$, then $m' \equiv m$.

[7] This quantity could be called the (dyadic) *metric entropy* of E at scale 2^{-n}.

Exercise 1.1.16. Let $d_1, d_2 \geq 1$, and let $E_1 \subset \mathbf{R}^{d_1}$, $E_2 \subset \mathbf{R}^{d_2}$ be Jordan measurable sets. Show that $E_1 \times E_2 \subset \mathbf{R}^{d_1+d_2}$ is Jordan measurable, and $m^{d_1+d_2}(E_1 \times E_2) = m^{d_1}(E_1) \times m^{d_2}(E_2)$.

Exercise 1.1.17. Let P, Q be two polytopes in $\mathbf{R}^d$. Suppose that P can be partitioned into finitely many sub-polytopes $P_1, \ldots, P_n$ which, after being rotated and translated, form new polytopes $Q_1, \ldots, Q_n$ which are an *almost disjoint cover* of Q, which means that $Q = Q_1 \cup \ldots \cup Q_n$, and for any $1 \leq i < j \leq n$, Q_i and Q_j only intersect at the boundary (i.e. the interior of Q_i is disjoint from the interior of Q_j). Conclude that P and Q have the same Jordan measure. The converse statement is true in one and two dimensions $d = 1, 2$ (this is the *Bolyai-Gerwien theorem*), but false in higher dimensions (this was Dehn's negative answer[**De1901**] to *Hilbert's third problem*).

The above exercises give a fairly large class of Jordan measurable sets. However, not every subset of $\mathbf{R}^d$ is Jordan measurable. First of all, the unbounded sets are not Jordan measurable, by construction. But there are also bounded sets that are not Jordan measurable:

Exercise 1.1.18. Let $E \subset \mathbf{R}^d$ be a bounded set.

(1) Show that E and the closure $\overline{E}$ of E have the same Jordan outer measure.

(2) Show that E and the interior E° of E have the same Jordan inner measure.

(3) Show that E is Jordan measurable if and only if the topological boundary ∂E of E has Jordan outer measure zero.

(4) Show that the *bullet-riddled square* $[0,1]^2 \backslash \mathbf{Q}^2$, and set of bullets $[0,1]^2 \cap Q^2$, both have Jordan inner measure zero and Jordan outer measure one. In particular, both sets are not Jordan measurable.

Informally, any set with a lot of "holes", or a very "fractal" boundary, is unlikely to be Jordan measurable. In order to measure such sets we will need to develop *Lebesgue measure*, which is done in the next set of notes.

Exercise 1.1.19 (Carathéodory type property)**.** Let $E \subset \mathbf{R}^d$ be a bounded set, and $F \subset \mathbf{R}^d$ be an elementary set. Show that $m^{*,(J)}(E) = m^{*,(J)}(E \cap F) + m^{*,(J)}(E \backslash F)$.

1.1.3. Connection with the Riemann integral. To conclude this section, we briefly discuss the relationship between Jordan measure and the *Riemann integral* (or the equivalent *Darboux integral*). For simplicity we will only discuss the classical one-dimensional Riemann integral on an interval $[a, b]$, though one can extend the Riemann theory without much difficulty to higher-dimensional integrals on Jordan measurable sets. (In later sections, this Riemann integral will be superceded by the *Lebesgue integral*.)

Definition 1.1.5 (Riemann integrability)**.** Let $[a, b]$ be an interval of positive length, and $f : [a, b] \to \mathbf{R}$ be a function. A *tagged partition* $\mathcal{P} = ((x_0, x_1, \ldots, x_n), (x_1^*, \ldots, x_n^*))$ of $[a, b]$ is a finite sequence of real numbers $a = x_0 < x_1 < \ldots < x_n = b$, together with additional numbers $x_{i-1} \leq x_i^* \leq x_i$ for each $i = 1, \ldots, n$. We abbreviate $x_i - x_{i-1}$ as δx_i. The quantity $\Delta(\mathcal{P}) := \sup_{1 \leq i \leq n} \delta x_i$ will be called the *norm* of the tagged partition. The *Riemann sum* $\mathcal{R}(f, \mathcal{P})$ of f with respect to the tagged partition $\mathcal{P}$ is defined as

$$\mathcal{R}(f, \mathcal{P}) := \sum_{i=1}^{n} f(x_i^*) \delta x_i.$$

We say that f is *Riemann integrable* on $[a, b]$ if there exists a real number, denoted $\int_a^b f(x)\ dx$ and referred to as the *Riemann integral* of f on $[a, b]$, for which we have

$$\int_a^b f(x)\ dx = \lim_{\Delta(\mathcal{P}) \to 0} \mathcal{R}(f, \mathcal{P})$$

by which we mean that for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|\mathcal{R}(f, \mathcal{P}) - \int_a^b f(x)\ dx| \leq \varepsilon$ for every tagged partition $\mathcal{P}$ with $\Delta(\mathcal{P}) \leq \delta$.

If $[a, b]$ is an interval of zero length, we adopt the convention that every function $f : [a, b] \to \mathbf{R}$ is Riemann integrable, with a Riemann integral of zero.

Note that unbounded functions cannot be Riemann integrable (why?).

The above definition, while geometrically natural, can be awkward to use in practice. A more convenient formulation of the Riemann integral can be formulated using some additional machinery.

Exercise 1.1.20 (Piecewise constant functions)**.** Let $[a,b]$ be an interval. A *piecewise constant function* $f\colon [a,b] \to \mathbf{R}$ is a function for which there exists a partition of $[a,b]$ into finitely many intervals $I_1, \ldots, I_n$, such that f is equal to a constant c_i on each of the intervals I_i. If f is piecewise constant, show that the expression

$$\sum_{i=1}^n c_i |I_i|$$

is independent of the choice of partition used to demonstrate the piecewise constant nature of f. We will denote this quantity by $\mathrm{p.c.}\int_a^b f(x)\ dx$, and refer to it as the *piecewise constant integral* of f on $[a,b]$.

Exercise 1.1.21 (Basic properties of the piecewise constant integral)**.** Let $[a,b]$ be an interval, and let $f, g\colon [a,b] \to \mathbf{R}$ be piecewise constant functions. Establish the following statements:

(1) (Linearity) For any real number c, cf and $f+g$ are piecewise constant, with $\mathrm{p.c.}\int_a^b cf(x)\ dx = c\mathrm{p.c.}\int_a^b f(x)\ dx$ and $\mathrm{p.c.}\int_a^b f(x)+g(x)\ dx = \mathrm{p.c.}\int_a^b f(x)\ dx + \mathrm{p.c.}\int_a^b g(x)\ dx$.

(2) (Monotonicity) If $f \leq g$ pointwise (i.e. $f(x) \leq g(x)$ for all $x \in [a,b]$) then $\mathrm{p.c.}\int_a^b f(x)\ dx \leq \mathrm{p.c.}\int_a^b g(x)\ dx$.

(3) (Indicator) If E is an elementary subset of $[a,b]$, then the indicator function $1_E\colon [a,b] \to \mathbf{R}$ (defined by setting $1_E(x) := 1$ when $x \in E$ and $1_E(x) := 0$ otherwise) is piecewise constant, and $\mathrm{p.c.}\int_a^b 1_E(x)\ dx = m(E)$.

Definition 1.1.6 (Darboux integral)**.** Let $[a,b]$ be an interval, and $f\colon [a,b] \to \mathbf{R}$ be a bounded function. The *lower Darboux integral* $\underline{\int_a^b} f(x)\ dx$ of f on $[a,b]$ is defined as

$$\underline{\int_a^b} f(x)\ dx := \sup_{g \leq f,\text{ piecewise constant}} \mathrm{p.c.}\int_a^b g(x)\ dx,$$

where g ranges over all piecewise constant functions that are pointwise bounded above by f. (The hypothesis that f is bounded ensures that the supremum is over a non-empty set.) Similarly, we define the *upper Darboux integral* $\overline{\int_a^b} f(x)\, dx$ of f on $[a,b]$ by the formula

$$\overline{\int_a^b} f(x)\, dx := \inf_{h \geq f,\text{ piecewise constant}} \text{p.c.} \int_a^b h(x)\, dx.$$

Clearly $\underline{\int_a^b} f(x)\, dx \leq \overline{\int_a^b} f(x)\, dx$. If these two quantities are equal, we say that f is *Darboux integrable*, and refer to this quantity as the *Darboux integral* of f on $[a,b]$.

Note that the upper and lower Darboux integrals are related by the reflection identity

$$\overline{\int_a^b} -f(x)\, dx = -\underline{\int_a^b} f(x)\, dx.$$

Exercise 1.1.22. Let $[a,b]$ be an interval, and $f\colon [a,b] \to \mathbf{R}$ be a bounded function. Show that f is Riemann integrable if and only if it is Darboux integrable, in which case the Riemann integral and Darboux integrals are equal.

Exercise 1.1.23. Show that any continuous function $f\colon [a,b] \to \mathbf{R}$ is Riemann integrable. More generally, show that any bounded, piecewise continuous[8] function $f\colon [a,b] \to \mathbf{R}$ is Riemann integrable.

Now we connect the Riemann integral to Jordan measure in two ways. First, we connect the Riemann integral to one-dimensional Jordan measure:

Exercise 1.1.24 (Basic properties of the Riemann integral)**.** Let $[a,b]$ be an interval, and let $f, g\colon [a,b] \to \mathbf{R}$ be Riemann integrable. Establish the following statements:

(1) (Linearity) For any real number c, cf and $f+g$ are Riemann integrable, with $\int_a^b cf(x)\, dx = c \cdot \int_a^b f(x)\, dx$ and $\int_a^b f(x) + g(x)\, dx = \int_a^b f(x)\, dx + \int_a^b g(x)\, dx$.

(2) (Monotonicity) If $f \leq g$ pointwise (i.e. $f(x) \leq g(x)$ for all $x \in [a,b]$) then $\int_a^b f(x)\, dx \leq \int_a^b g(x)\, dx$.

[8] A function $f\colon [a,b] \to \mathbf{R}$ is *piecewise continuous* if one can partition $[a,b]$ into finitely many intervals, such that f is continuous on each interval.

(3) (Indicator) If E is a Jordan measurable of $[a,b]$, then the indicator function $1_E\colon [a,b] \to \mathbf{R}$ (defined by setting $1_E(x) := 1$ when $x \in E$ and $1_E(x) := 0$ otherwise) is Riemann integrable, and $\int_a^b 1_E(x)\ dx = m(E)$.

Finally, show that these properties uniquely define the Riemann integral, in the sense that the functional $f \mapsto \int_a^b f(x)\ dx$ is the only map from the space of Riemann integrable functions on $[a,b]$ to $\mathbf{R}$ which obeys all three of the above properties.

Next, we connect the integral to two-dimensional Jordan measure:

Exercise 1.1.25 (Area interpretation of the Riemann integral)**.** Let $[a,b]$ be an interval, and let $f\colon [a,b] \to \mathbf{R}$ be a bounded function. Show that f is Riemann integrable if and only if the sets $E_+ := \{(x,t) : x \in [a,b]; 0 \le t \le f(x)\}$ and $E_- := \{(x,t) : x \in [a,b]; f(x) \le t \le 0\}$ are both Jordan measurable in $\mathbf{R}^2$, in which case one has

$$\int_a^b f(x)\ dx = m^2(E_+) - m^2(E_-),$$

where m^2 denotes two-dimensional Jordan measure. (*Hint:* First establish this in the case when f is non-negative.)

Exercise 1.1.26. Extend the definition of the Riemann and Darboux integrals to higher dimensions, in such a way that analogues of all the previous results hold.

1.2. Lebesgue measure

In Section 1.1, we recalled the classical theory of *Jordan measure* on Euclidean spaces $\mathbf{R}^d$. This theory proceeded in the following stages:

(i) First, one defined the notion of a box B and its volume $|B|$.

(ii) Using this, one defined the notion of an elementary set E (a finite union of boxes), and defines the elementary measure $m(E)$ of such sets.

(iii) From this, one defined the inner and Jordan outer measures $m_{*,(J)}(E), m^{*,(J)}(E)$ of an arbitrary bounded set $E \subset \mathbf{R}^d$. If those measures match, we say that E is *Jordan measurable*, and call $m(E) = m_{*,(J)}(E) = m^{*,(J)}(E)$ the *Jordan measure* of E.

As long as one is lucky enough to only have to deal with Jordan measurable sets, the theory of Jordan measure works well enough. However, as noted previously, not all sets are Jordan measurable, even if one restricts attention to bounded sets. In fact, we shall see later in these notes that there even exist bounded open sets, or compact sets, which are not Jordan measurable, so the Jordan theory does not cover many classes of sets of interest. Another class that it fails to cover is countable unions or intersections of sets that are already known to be measurable:

Exercise 1.2.1. Show that the countable union $\bigcup_{n=1}^\infty E_n$ or countable intersection $\bigcap_{n=1}^\infty E_n$ of Jordan measurable sets $E_1, E_2, \ldots \subset \mathbf{R}$ need not be Jordan measurable, even when bounded.

This creates problems with Riemann integrability (which, as we saw in Section 1.1, was closely related to Jordan measure) and pointwise limits:

Exercise 1.2.2. Give an example of a sequence of uniformly bounded, Riemann integrable functions $f_n\colon [0,1] \to \mathbf{R}$ for $n = 1, 2, \ldots$ that converge pointwise to a bounded function $f\colon [0,1] \to \mathbf{R}$ that is *not* Riemann integrable. What happens if we replace pointwise convergence with uniform convergence?

These issues can be rectified by using a more powerful notion of measure than Jordan measure, namely *Lebesgue measure*. To define this measure, we first tinker with the notion of the Jordan outer measure

$$m^{*,(J)}(E) := \inf_{B \supset E; B \text{ elementary}} m(B)$$

of a set $E \subset \mathbf{R}^d$ (we adopt the convention that $m^{*,(J)}(E) = +\infty$ if E is unbounded, thus $m^{*,(J)}$ now takes values in the extended non-negative reals $[0, +\infty]$, whose properties we will briefly review below). Observe from the finite additivity and subadditivity of elementary measure that we can also write the Jordan outer measure as

$$m^{*,(J)}(E) := \inf_{B_1 \cup \ldots \cup B_k \supset E; B_1, \ldots, B_k \text{ boxes}} |B_1| + \ldots + |B_k|,$$

i.e. the Jordan outer measure is the infimal cost required to cover E by a finite union of boxes. (The natural number k is allowed to vary freely in the above infimum.) We now modify this by replacing the finite union of boxes by a countable union of boxes, leading to the *Lebesgue outer measure*[9] $m^*(E)$ of E:

$$m^*(E) := \inf_{\bigcup_{n=1}^\infty B_n \supset E; B_1, B_2, \ldots \text{ boxes}} \sum_{n=1}^\infty |B_n|,$$

thus the Lebesgue outer measure is the infimal cost required to cover E by a *countable* union of boxes. Note that the countable sum $\sum_{n=1}^\infty |B_n|$ may be infinite, and so the Lebesgue outer measure $m^*(E)$ could well equal $+\infty$.

Clearly, we always have $m^*(E) \leq m^{*,(J)}(E)$ (since we can always pad out a finite union of boxes into an infinite union by adding an infinite number of empty boxes). But $m^*(E)$ can be a lot smaller:

Example 1.2.1. Let $E = \{x_1, x_2, x_3, \ldots\} \subset \mathbf{R}^d$ be a countable set. We know that the Jordan outer measure of E can be quite large; for instance, in one dimension, $m^{*,(J)}(\mathbf{Q})$ is infinite, and $m^{*,(J)}(\mathbf{Q} \cap [-R, R]) = m^{*,(J)}([-R, R]) = 2R$ since $\mathbf{Q} \cap [-R, R]$ has $[-R, R]$ as its closure (see Exercise 1.1.18). On the other hand, all countable sets E have Lebesgue outer measure zero. Indeed, one simply covers E by the degenerate boxes $\{x_1\}, \{x_2\}, \ldots$ of sidelength and volume zero.

Alternatively, if one does not like degenerate boxes, one can cover each x_n by a cube B_n of sidelength $\varepsilon/2^n$ (say) for some arbitrary $\varepsilon > 0$, leading to a total cost of $\sum_{n=1}^\infty (\varepsilon/2^n)^d$, which converges to $C_d \varepsilon^d$ for some absolute constant C_d. As ε can be arbitrarily small, we see that the Lebesgue outer measure must be zero. We will refer to this type of trick as the $\varepsilon/2^n$ *trick*; it will be used many further times in this text.

From this example we see in particular that a set may be unbounded while still having Lebesgue outer measure zero, in contrast to Jordan outer measure.

[9]Lebesgue outer measure is also denoted $m_*(E)$ in some texts.

As we shall see in Section 1.7, Lebesgue outer measure (also known as *Lebesgue exterior measure*) is a special case of a more general concept known as an *outer measure*.

In analogy with the Jordan theory, we would also like to define a concept of "Lebesgue inner measure" to complement that of outer measure. Here, there is an asymmetry (which ultimately arises from the fact that elementary measure is subadditive rather than superadditive): one does not gain any increase in power in the Jordan inner measure by replacing finite unions of boxes with countable ones. But one can get a sort of Lebesgue inner measure by taking complements; see Exercise 1.2.18. This leads to one possible definition for Lebesgue measurability, namely the *Carathéodory criterion* for Lebesgue measurability, see Exercise 1.2.17. However, this is not the most intuitive formulation of this concept to work with, and we will instead use a different (but logically equivalent) definition of Lebesgue measurability. The starting point is the observation (see Exercise 1.1.13) that Jordan measurable sets can be efficiently contained in elementary sets, with an error that has small Jordan outer measure. In a similar vein, we will define Lebesgue measurable sets to be sets that can be efficiently contained in *open* sets, with an error that has small *Lebesgue* outer measure:

Definition 1.2.2 (Lebesgue measurability)**.** A set $E \subset \mathbf{R}^d$ is said to be *Lebesgue measurable* if, for every $\varepsilon > 0$, there exists an open set $U \subset \mathbf{R}^d$ containing E such that $m^*(U \backslash E) \leq \varepsilon$. If E is Lebesgue measurable, we refer to $m(E) := m^*(E)$ as the *Lebesgue measure* of E (note that this quantity may be equal to $+\infty$). We also write $m(E)$ as $m^d(E)$ when we wish to emphasise the dimension d.

Remark 1.2.3. The intuition that measurable sets are almost open is also known as *Littlewood's first principle*, this principle is a triviality with our current choice of definitions, though less so if one uses other, equivalent, definitions of Lebesgue measurability. See Section 1.3.5 for a further discussion of Littlewood's principles.

As we shall see later, Lebesgue measure extends Jordan measure, in the sense that every Jordan measurable set is Lebesgue measurable, and the Lebesgue measure and Jordan measure of a Jordan measurable set are always equal. We will also see a few other equivalent descriptions of the concept of Lebesgue measurability.

In the notes below we will establish the basic properties of Lebesgue measure. Broadly speaking, this concept obeys all the intuitive properties one would ask of measure, so long as one restricts attention to countable operations rather than uncountable ones, and as long as one restricts attention to Lebesgue measurable sets. The latter is not a serious restriction in practice, as almost every set one actually encounters in analysis will be measurable (the main exceptions being some pathological sets that are constructed using the axiom of choice). In the next set of notes we will use Lebesgue measure to set up the Lebesgue integral, which extends the Riemann integral in the same way that Lebesgue measure extends Jordan measure; and the many pleasant properties of Lebesgue measure will be reflected in analogous pleasant properties of the Lebesgue integral (most notably the convergence theorems).

We will treat all dimensions $d = 1, 2, \ldots$ equally here, but for the purposes of drawing pictures, we recommend to the reader that one sets d equal to 2. However, for this topic at least, no additional mathematical difficulties will be encountered in the higher-dimensional case (though of course there are significant *visual* difficulties once d exceeds 3).

1.2.1. Properties of Lebesgue outer measure. We begin by studying the Lebesgue outer measure m^*, which was defined earlier, and takes values in the extended non-negative real axis $[0, +\infty]$. We first record three easy properties of Lebesgue outer measure, which we will use repeatedly in the sequel without further comment:

Exercise 1.2.3 (The outer measure axioms)**.**

(i) (Empty set) $m^*(\emptyset) = 0$.

(ii) (Monotonicity) If $E \subset F \subset \mathbf{R}^d$, then $m^*(E) \leq m^*(F)$.

(iii) (Countable subadditivity) If $E_1, E_2, \ldots \subset \mathbf{R}^d$ is a countable sequence of sets, then $m^*(\bigcup_{n=1}^\infty E_n) \leq \sum_{n=1}^\infty m^*(E_n)$. (*Hint:* Use the axiom of countable choice, Tonelli's theorem for series, and the $\varepsilon/2^n$ trick used previously to show that countable sets had outer measure zero.)

Note that countable subadditivity, when combined with the empty set axiom, gives as a corollary the finite subadditivity property

$$m^*(E_1 \cup \ldots \cup E_k) \leq m^*(E_1) + \ldots + m^*(E_k)$$

for any $k \geq 0$. These subadditivity properties will be useful in establishing *upper* bounds on Lebesgue outer measure. Establishing *lower* bounds will often be a bit trickier. (More generally, when dealing with a quantity that is defined using an infimum, it is usually easier to obtain upper bounds on that quantity than lower bounds, because the former requires one to bound just one element of the infimum, whereas the latter requires one to bound *all* elements.)

Remark 1.2.4. Later on in this text, when we study abstract measure theory on a general set X, we will define the concept of an *outer measure* on X, which is an assigment $E \mapsto m^*(E)$ of element of $[0, +\infty]$ to arbitrary subsets E of a space X that obeys the above three axioms of the empty set, monotonicity, and countable subadditivity; thus Lebesgue outer measure is a model example of an abstract outer measure. On the other hand (and somewhat confusingly), Jordan outer measure will not be an abstract outer measure (even after adopting the convention that unbounded sets have Jordan outer measure $+\infty$): it obeys the empty set and monotonicity axioms, but is only finitely subadditive rather than countably subadditive. (For instance, the rationals $\mathbf{Q}$ have infinite Jordan outer measure, despite being the countable union of points, each of which have a Jordan outer measure of zero.) Thus we already see a major benefit of allowing countable unions of boxes in the definition of Lebesgue outer measure, in contrast to the finite unions of boxes in the definition of Jordan outer measure, in that finite subadditivity is upgraded to countable subadditivity.

Of course, one cannot hope to upgrade countable subadditivity to uncountable subadditivity: $\mathbf{R}^d$ is an uncountable union of points, each of which has Lebesgue outer measure zero, but (as we shall shortly see), $\mathbf{R}^d$ has infinite Lebesgue outer measure.

It is natural to ask whether Lebesgue outer measure has the finite additivity property, that is to say that $m^*(E \cup F) = m^*(E) + m^*(F)$ whenever $E, F \subset \mathbf{R}^d$ are disjoint. The answer to this question is somewhat subtle: as we shall see later, we have finite additivity (and even countable additivity) when all sets involved are Lebesgue measurable, but that finite additivity (and hence also countable additivity) can break down in the non-measurable case. The difficulty here (which, incidentally, also appears in the theory of Jordan outer measure) is that if E and F are sufficiently "entangled" with each other, it is not always possible to take a countable cover of $E \cup F$ by boxes and split the total volume of that cover into separate covers of E and F without some duplication. However, we can at least recover finite additivity if the sets E, F are separated by some positive distance:

Lemma 1.2.5 (Finite additivity for separated sets)**.** *Let $E, F \subset \mathbf{R}^d$ be such that* $\operatorname{dist}(E, F) > 0$*, where*

$$\operatorname{dist}(E, F) := \inf\{|x - y| : x \in E, y \in F\}$$

is the distance[10] *between E and F. Then $m^*(E \cup F) = m^*(E) + m^*(F)$.*

Proof. From subadditivity one has $m^*(E \cup F) \leq m^*(E) + m^*(F)$, so it suffices to prove the other direction $m^*(E) + m^*(F) \leq m^*(E \cup F)$. This is trivial if $E \cup F$ has infinite Lebesgue outer measure, so we may assume that it has finite Lebesgue outer measure (and then the same is true for E and F, by monotonicity).

We use the standard "give yourself an epsilon of room" trick (see Section 2.7 of *An epsilon of room, Vol I.*). Let $\varepsilon > 0$. By definition of Lebesgue outer measure, we can cover $E \cup F$ by a countable family $B_1, B_2, \ldots$ of boxes such that

$$\sum_{n=1}^{\infty} |B_n| \leq m^*(E \cup F) + \varepsilon.$$

Suppose it was the case that each box intersected at most one of E and F. Then we could divide this family into two subfamilies $B_1', B_2', \ldots$

[10]Recall from the preface that we use the usual Euclidean metric $|(x_1, \ldots, x_d)| := \sqrt{x_1^2 + \ldots + x_d^2}$ on $\mathbf{R}^d$.

and $B_1'', B_2'', B_3'', \ldots$, the first of which covered E, and the second of which covered F. From definition of Lebesgue outer measure, we have

$$m^*(E) \leq \sum_{n=1}^{\infty} |B_n'|$$

and

$$m^*(F) \leq \sum_{n=1}^{\infty} |B_n''|;$$

summing, we obtain

$$m^*(E) + m^*(F) \leq \sum_{n=1}^{\infty} |B_n|$$

and thus

$$m^*(E) + m^*(F) \leq m^*(E \cup F) + \varepsilon.$$

Since ε was arbitrary, this gives $m^*(E) + m^*(F) \leq m^*(E \cup F)$ as required.

Of course, it is quite possible for some of the boxes B_n to intersect both E and F, particularly if the boxes are big, in which case the above argument does not work because that box would be double-counted. However, observe that given any $r > 0$, one can always partition a large box B_n into a finite number of smaller boxes, each of which has diameter[11] at most r, with the total volume of these sub-boxes equal to the volume of the original box. Applying this observation to each of the boxes B_n, we see that given any $r > 0$, we may assume without loss of generality that the boxes $B_1, B_2, \ldots$ covering $E \cup F$ have diameter at most r. In particular, we may assume that all such boxes have diameter strictly less than $\mathrm{dist}(E, F)$. Once we do this, then it is no longer possible for any box to intersect both E and F, and then the previous argument now applies. □

In general, disjoint sets E, F need not have a positive separation from each other (e.g. $E = [0, 1)$ and $F = [1, 2]$). But the situation improves when E, F are closed, and at least one of E, F is compact:

[11]The *diameter* of a set B is defined as $\sup\{|x - y| : x, y \in B\}$.

Exercise 1.2.4. Let $E, F \subset \mathbf{R}^d$ be disjoint closed sets, with at least one of E, F being compact. Show that $\operatorname{dist}(E, F) > 0$. Give a counterexample to show that this claim fails when the compactness hypothesis is dropped.

We already know that countable sets have Lebesgue outer measure zero. Now we start computing the outer measure of some other sets. We begin with elementary sets:

Lemma 1.2.6 (Outer measure of elementary sets)**.** *Let E be an elementary set. Then the Lebesgue outer measure $m^*(E)$ of E is equal to the elementary measure $m(E)$ of E: $m^*(E) = m(E)$.*

Remark 1.2.7. Since countable sets have zero outer measure, we note that we have managed to give a proof of *Cantor's theorem* that $\mathbf{R}^d$ is uncountable. Of course, much quicker proofs of this theorem are available. However, this observation shows that the proof this lemma must somehow use some crucial fact about the real line which is not true for countable subfields of $\mathbf{R}$, such as the rationals $\mathbf{Q}$. In the proof we give here, the key fact about the real line we use is the Heine-Borel theorem, which ultimately exploits the important fact that the reals are *complete*. In the one-dimensional case $d = 1$, it is also possible to exploit the fact that the reals are *connected* as a substitute for completeness (note that proper subfields of the reals are neither connected nor complete).

Proof. We already know that $m^*(E) \leq m^{*,(J)}(E) = m(E)$, so it suffices to show that $m(E) \leq m^*(E)$.

We first establish this in the case when the elementary set E is closed. As the elementary set E is also bounded, this allows us to use the powerful *Heine-Borel theorem*, which asserts that every open cover of E has a finite subcover (or in other words, E is *compact*).

We again use the epsilon of room strategy. Let $\varepsilon > 0$ be arbitrary, then we can find a countable family $B_1, B_2, \ldots$ of boxes that cover E,

$$E \subset \bigcup_{n=1}^{\infty} B_n,$$

and such that
$$\sum_{n=1}^\infty |B_n| \leq m^*(E) + \varepsilon.$$
We would like to use the Heine-Borel theorem, but the boxes B_n need not be open. But this is not a serious problem, as one can spend another epsilon to enlarge the boxes to be open. More precisely, for each box B_n one can find an open box B'_n containing B_n such that $|B'_n| \leq |B_n| + \varepsilon/2^n$ (say). The B'_n still cover E, and we have
$$\sum_{n=1}^\infty |B'_n| \leq \sum_{n=1}^\infty (|B_n| + \varepsilon/2^n) = (\sum_{n=1}^\infty |B_n|) + \varepsilon \leq m^*(E) + 2\varepsilon.$$
As the B'_n are open, we may apply the Heine-Borel theorem and conclude that
$$E \subset \bigcup_{n=1}^N B'_n$$
for some finite N. Using the finite subadditivity of elementary measure, we conclude that
$$m(E) \leq \sum_{n=1}^N |B'_n|$$
and thus
$$m(E) \leq m^*(E) + 2\varepsilon.$$
Since $\varepsilon > 0$ was arbitrary, the claim follows.

Now we consider the case when the elementary E is not closed. Then we can write E as the finite union $Q_1 \cup \ldots \cup Q_k$ of disjoint boxes, which need not be closed. But, similarly to before, we can use the epsilon of room strategy: for every $\varepsilon > 0$ and every $1 \leq j \leq k$, one can find a closed sub-box Q'_j of Q_j such that $|Q'_j| \geq |Q_j| - \varepsilon/k$ (say); then E contains the finite union of $Q'_1 \cup \ldots \cup Q'_k$ disjoint *closed* boxes, which is a closed elementary set. By the previous discussion and the finite additivity of elementary measure, we have
$$\begin{aligned} m^*(Q'_1 \cup \ldots \cup Q'_k) &= m(Q'_1 \cup \ldots \cup Q'_k) \\ &= m(Q'_1) + \ldots + m(Q'_k) \\ &\geq m(Q_1) + \ldots + m(Q_k) - \varepsilon \\ &= m(E) - \varepsilon. \end{aligned}$$

Applying by monotonicity of Lebesgue outer measure, we conclude that

$$m^*(E) \geq m(E) - \varepsilon$$

for every $\varepsilon > 0$. Since $\varepsilon > 0$ was arbitrary, the claim follows. $\square$

The above lemma allows us to compute the Lebesgue outer measure of a finite union of boxes. From this and monotonicity we conclude that the Lebesgue outer measure of any set is bounded below by its Jordan inner measure. As it is also bounded above by the Jordan outer measure, we have

$$m_{*,(J)}(E) \leq m^*(E) \leq m^{*,(J)}(E) \tag{1.2}$$

for every $E \subset \mathbf{R}^d$.

Remark 1.2.8. We are now able to explain why not every bounded open set or compact set is Jordan measurable. Consider the countable set $\mathbf{Q} \cap [0,1]$, which we enumerate as $\{q_1, q_2, q_3, \ldots\}$, let $\varepsilon > 0$ be a small number, and consider the set

$$U := \bigcup_{n=1}^\infty (q_n - \varepsilon/2^n, q_n + \varepsilon/2^n).$$

This is the union of open sets and is thus open. On the other hand, by countable subadditivity, one has

$$m^*(U) \leq \sum_{n=1}^\infty 2\varepsilon/2^n = 2\varepsilon.$$

Finally, as U is dense in $[0,1]$ (i.e. $\overline{U}$ contains $[0,1]$), we have

$$m^{*,(J)}(U) = m^{*,(J)}(\overline{U}) \geq m^{*,(J)}([0,1]) = 1.$$

For ε small enough (e.g. $\varepsilon := 1/3$), we see that the Lebesgue outer measure and Jordan outer measure of U disagree. Using (1.2), we conclude that the bounded open set U is not Jordan measurable. This in turn implies that the complement of U in, say, $[-2,2]$, is also not Jordan measurable, despite being a compact set.

Now we turn to countable unions of boxes. It is convenient to introduce the following notion: two boxes are *almost disjoint* if their interiors are disjoint, thus for instance $[0, 1]$ and $[1, 2]$ are almost disjoint. As a box has the same elementary measure as its interior, we see that the finite additivity property

$$m(B_1 \cup \ldots \cup B_k) = |B_1| + \ldots + |B_k| \tag{1.3}$$

holds for *almost* disjoint boxes $B_1, \ldots, B_k$, and not just for disjoint boxes. This (and Lemma 1.2.6) has the following consequence:

Lemma 1.2.9 (Outer measure of countable unions of almost disjoint boxes)**.** *Let $E = \bigcup_{n=1}^\infty B_n$ be a countable union of almost disjoint boxes $B_1, B_2, \ldots$. Then*

$$m^*(E) = \sum_{n=1}^\infty |B_n|.$$

Thus, for instance, $\mathbf{R}^d$ itself has an infinite outer measure.

Proof. From countable subadditivity and Lemma 1.2.6 we have

$$m^*(E) \leq \sum_{n=1}^\infty m^*(B_n) = \sum_{n=1}^\infty |B_n|,$$

so it suffices to show that

$$\sum_{n=1}^\infty |B_n| \leq m^*(E).$$

But for each natural number N, E contains the elementary set $B_1 \cup \ldots \cup B_N$, so by monotonicity and Lemma 1.2.6,

$$\begin{aligned} m^*(E) &\geq m^*(B_1 \cup \ldots \cup B_N) \\ &= m(B_1 \cup \ldots \cup B_N) \end{aligned}$$

and thus by (1.3), one has

$$\sum_{n=1}^N |B_n| \leq m^*(E).$$

Letting $N \to \infty$ we obtain the claim. □

Remark 1.2.10. The above lemma has the following immediate corollary: if $E = \bigcup_{n=1}^\infty B_n = \bigcup_{n=1}^\infty B_n'$ can be decomposed in two different ways as the countable union of almost disjoint boxes, then

$\sum_{n=1}^\infty |B_n| = \sum_{n=1}^\infty |B'_n|$. Although this statement is intuitively obvious and does not explicitly use the concepts of Lebesgue outer measure or Lebesgue measure, it is remarkably difficult to prove this statement rigorously without essentially using one of these two concepts. (Try it!)

Exercise 1.2.5. Show that if a set $E \subset \mathbf{R}^d$ is expressible as the countable union of almost disjoint boxes, then the Lebesgue outer measure of E is equal to the Jordan inner measure: $m^*(E) = m_{*,(J)}(E)$, where we extend the definition of Jordan inner measure to unbounded sets in the obvious manner.

Not every set can be expressed as the countable union of almost disjoint boxes (consider for instance the irrationals $\mathbf{R}\backslash\mathbf{Q}$, which contain no boxes other than the singleton sets). However, there is an important class of sets of this form, namely the *open sets*:

Lemma 1.2.11. *Let $E \subset \mathbf{R}^d$ be an open set. Then E can be expressed as the countable union of almost disjoint boxes (and, in fact, as the countable union of almost disjoint closed cubes).*

Proof. We will use the *dyadic mesh* structure of the Euclidean space $\mathbf{R}^d$, which is a convenient tool for "discretising" certain aspects of real analysis.

Define a *closed dyadic cube* to be a cube Q of the form

$$Q = \left[\frac{i_1}{2^n}, \frac{i_1+1}{2^n}\right] \times \ldots \times \left[\frac{i_d}{2^n}, \frac{i_d+1}{2^n}\right]$$

for some integers $n, i_1, \ldots, i_d$. To avoid some technical issues we shall restrict attention here to "small" cubes of sidelength at most 1, thus we restrict n to the non-negative integers, and we will completely ignore "large" cubes of sidelength greater than one. Observe that the closed dyadic cubes of a fixed sidelength 2^{-n} are almost disjoint, and cover all of $\mathbf{R}^d$. Also observe that each dyadic cube of sidelength 2^{-n} is contained in exactly one "parent" cube of sidelength 2^{-n+1} (which, conversely, has 2^d "children" of sidelength 2^{-n}), giving the dyadic cubes a structure analogous to that of a binary tree (or more precisely, an infinite forest of 2^d-ary trees). As a consequence of these facts, we also obtain the important *dyadic nesting property*: given any two closed dyadic cubes (possibly of different sidelength), either they are almost disjoint, or one of them is contained in the other.

If E is open, and $x \in E$, then by definition there is an open ball centered at x that is contained in E, and it is easy to conclude that there is also a closed dyadic cube containing x that is contained in E. Thus, if we let $\mathcal{Q}$ be the collection of all the dyadic cubes Q that are contained in E, we see that the union $\bigcup_{Q \in \mathcal{Q}} Q$ of all these cubes is exactly equal to E.

As there are only countably many dyadic cubes, $\mathcal{Q}$ is at most countable. But we are not done yet, because these cubes are not almost disjoint (for instance, any cube Q in $\mathcal{Q}$ will of course overlap with its child cubes). But we can deal with this by exploiting the dyadic nesting property. Let $\mathcal{Q}^*$ denote those cubes in $\mathcal{Q}$ which are *maximal* with respect to set inclusion, which means that they are not contained in any other cube in $\mathcal{Q}$. From the nesting property (and the fact that we have capped the maximum size of our cubes) we see that every cube in $\mathcal{Q}$ is contained in exactly one maximal cube in $\mathcal{Q}^*$, and that any two such maximal cubes in $\mathcal{Q}^*$ are almost disjoint. Thus, we see that E is the union $E = \bigcup_{Q \in \mathcal{Q}^*} Q$ of almost disjoint cubes. As $\mathcal{Q}^*$ is at most countable, the claim follows (adding empty boxes if necessary to pad out the cardinality). □

We now have a formula for the Lebesgue outer measure of any open set: it is exactly equal to the Jordan inner measure of that set, or of the total volume of any partitioning of that set into almost disjoint boxes. Finally, we have a formula for the Lebesgue outer measure of an arbitrary set:

Lemma 1.2.12 (Outer regularity)**.** *Let $E \subset \mathbf{R}^d$ be an arbitrary set. Then one has*

$$m^*(E) = \inf_{E \subset U, U \text{ } open} m^*(U).$$

Proof. From monotonicity one trivially has

$$m^*(E) \leq \inf_{E \subset U, U \text{ open}} m^*(U)$$

so it suffices to show that

$$\inf_{E \subset U, U \text{ open}} m^*(U) \leq m^*(E).$$

This is trivial for $m^*(E)$ infinite, so we may assume that $m^*(E)$ is finite.

Let $\varepsilon > 0$. By definition of outer measure, there exists a countable family $B_1, B_2, \ldots$ of boxes covering E such that

$$\sum_{n=1}^{\infty} |B_n| \leq m^*(E) + \varepsilon.$$

We use the $\varepsilon/2^n$ trick again. We can enlarge each of these boxes B_n to an *open* box B'_n such that $|B'_n| \leq |B_n| + \varepsilon/2^n$. Then the set $\bigcup_{n=1}^{\infty} B'_n$, being a union of open sets, is itself open, and contains E; and

$$\sum_{n=1}^{\infty} |B'_n| \leq m^*(E) + \varepsilon + \sum_{n=1}^{\infty} \varepsilon/2^n = m^*(E) + 2\varepsilon.$$

By countable subadditivity, this implies that

$$m^*(\bigcup_{n=1}^{\infty} B'_n) \leq m^*(E) + 2\varepsilon$$

and thus

$$\inf_{E \subset U, U \text{ open}} m^*(U) \leq m^*(E) + 2\varepsilon.$$

As $\varepsilon > 0$ was arbitrary, we obtain the claim. □

Exercise 1.2.6. Give an example to show that the reverse statement

$$m^*(E) = \sup_{U \subset E, U \text{ open}} m^*(U)$$

is false. (For the corrected version of this statement, see Exercise 1.2.15.)

1.2.2. Lebesgue measurability. We now define the notion of a Lebesgue measurable set as one which can be efficiently contained in open sets in the sense of Definition 1.2.2, and set out their basic properties.

First, we show that there are plenty of Lebesgue measurable sets.

Lemma 1.2.13 (Existence of Lebesgue measurable sets)**.**

(i) *Every open set is Lebesgue measurable.*

(ii) *Every closed set is Lebesgue measurable.*

(iii) *Every set of Lebesgue outer measure zero is measurable. (Such sets are called* null sets*.)*

(iv) *The empty set $\emptyset$ is Lebesgue measurable.*

(v) *If $E \subset \mathbf{R}^d$ is Lebesgue measurable, then so is its complement $\mathbf{R}^d \backslash E$.*

(vi) *If $E_1, E_2, E_3, \ldots \subset \mathbf{R}^d$ are a sequence of Lebesgue measurable sets, then the union $\bigcup_{n=1}^\infty E_n$ is Lebesgue measurable.*

(vii) *If $E_1, E_2, E_3, \ldots \subset \mathbf{R}^d$ are a sequence of Lebesgue measurable sets, then the intersection $\bigcap_{n=1}^\infty E_n$ is Lebesgue measurable.*

Proof. Claim (i) is obvious from definition, as are Claims (iii) and (iv).

To prove Claim (vi), we use the $\varepsilon/2^n$ trick. Let $\varepsilon > 0$ be arbitrary. By hypothesis, each E_n is contained in an open set U_n whose difference $U_n \backslash E_n$ has Lebesgue outer measure at most $\varepsilon/2^n$. By countable subadditivity, this implies that $\bigcup_{n=1}^\infty E_n$ is contained in $\bigcup_{n=1}^\infty U_n$, and the difference $(\bigcup_{n=1}^\infty U_n) \backslash (\bigcup_{n=1}^\infty E_n)$ has Lebesgue outer measure at most ε. The set $\bigcup_{n=1}^\infty U_n$, being a union of open sets, is itself open, and the claim follows.

Now we establish Claim (ii). Every closed set E is the countable union of closed and bounded sets (by intersecting E with, say, the closed balls $\overline{B(0,n)}$ of radius n for $n = 1, 2, 3, \ldots$), so by (vi), it suffices to verify the claim when E is closed and bounded, hence compact by the Heine-Borel theorem. Note that the boundedness of E implies that $m^*(E)$ is finite.

Let $\varepsilon > 0$. By outer regularity (Lemma 1.2.12), we can find an open set U containing E such that $m^*(U) \leq m^*(E) + \varepsilon$. It suffices to show that $m^*(U \backslash E) \leq \varepsilon$.

The set $U \backslash E$ is open, and so by Lemma 1.2.11 is the countable union $\bigcup_{n=1}^\infty Q_n$ of almost disjoint closed cubes. By Lemma 1.2.9, $m^*(U \backslash E) = \sum_{n=1}^\infty |Q_n|$. So it will suffice to show that $\sum_{n=1}^N |Q_n| \leq \varepsilon$ for every finite N.

The set $\bigcup_{n=1}^N Q_n$ is a finite union of closed cubes and is thus closed. It is disjoint from the compact set E, so by Exercise 1.2.4 followed by Lemma 1.2.5 one has

$$m^*(E \cup \bigcup_{n=1}^N Q_n) = m^*(E) + m^*(\bigcup_{n=1}^N Q_n).$$

By monotonicity, the left-hand side is at most $m^*(U)$, which is in turn at most $m^*(E) + \varepsilon$. Since $m^*(E)$ is finite, we may cancel it and conclude that $m^*(\bigcup_{n=1}^N Q_n) \leq \varepsilon$, as required.

Next, we establish Claim (v). If E is Lebesgue measurable, then for every n we can find an open set U_n containing E such that $m^*(U_n \backslash E) \leq 1/n$. Letting F_n be the complement of U_n, we conclude that the complement $\mathbf{R}^d \backslash E$ of E contains all of the F_n, and that $m^*((\mathbf{R}^d \backslash E) \backslash F_n) \leq 1/n$. If we let $F := \bigcup_{n=1}^\infty F_n$, then $\mathbf{R}^d \backslash E$ contains F, and from monotonicity $m^*((\mathbf{R}^d \backslash E) \backslash F) = 0$, thus $\mathbf{R}^d \backslash E$ is the union of F and a set of Lebesgue outer measure zero. But F is in turn the union of countably many closed sets F_n. The claim now follows from (ii), (iii), (iv).

Finally, Claim (vii) follows from (v), (vi), and *de Morgan's laws* $(\bigcap_{\alpha \in A} E_\alpha)^c = \bigcup_{\alpha \in A} E_\alpha^c$, $(\bigcup_{\alpha \in A} E_\alpha)^c = \bigcap_{\alpha \in A} E_\alpha^c$, (which work for infinite unions and intersections without any difficulty). □

Informally, the above lemma asserts (among other things) that if one starts with such basic subsets of $\mathbf{R}^d$ as open or closed sets and then takes at most countably many boolean operations, one will always end up with a Lebesgue measurable set. This is already enough to ensure that the majority of sets that one actually encounters in real analysis will be Lebesgue measurable. (Nevertheless, using the axiom of choice one can construct sets that are not Lebesgue measurable; we will see an example of this later. As a consequence, we cannot generalise the countable closure properties here to uncountable closure properties.)

Remark 1.2.14. The properties (iv), (v), (vi) of Lemma 1.2.13 assert that the collection of Lebesgue measurable subsets of $\mathbf{R}^d$ form a σ-*algebra*, which is a strengthening of the more classical concept of a *boolean algebra*. We will study abstract σ-algebras in more detail in Section 1.4.

Note how this Lemma 1.2.13 is significantly stronger than the counterpart for Jordan measurability (Exercise 1.1.6), in particular by allowing countably many boolean operations instead of just finitely many. This is one of the main reasons why we use Lebesgue measure instead of Jordan measure.

Exercise 1.2.7 (Criteria for measurability)**.** Let $E \subset \mathbf{R}^d$. Show that the following are equivalent:

(i) E is Lebesgue measurable.

(ii) (Outer approximation by open) For every $\varepsilon > 0$, one can contain E in an open set U with $m^*(U \backslash E) \leq \varepsilon$.

(iii) (Almost open) For every $\varepsilon > 0$, one can find an open set U such that $m^*(U \Delta E) \leq \varepsilon$. (In other words, E differs from an open set by a set of outer measure at most ε.)

(iv) (Inner approximation by closed) For every $\varepsilon > 0$, one can find a closed set F contained in E with $m^*(E \backslash F) \leq \varepsilon$.

(v) (Almost closed) For every $\varepsilon > 0$, one can find a closed set F such that $m^*(F \Delta E) \leq \varepsilon$. (In other words, E differs from a closed set by a set of outer measure at most ε.)

(vi) (Almost measurable) For every $\varepsilon > 0$, one can find a Lebesgue measurable set E_ε such that $m^*(E_\varepsilon \Delta E) \leq \varepsilon$. (In other words, E differs from a measurable set by a set of outer measure at most ε.)

(*Hint:* Some of these deductions are either trivial or very easy. To deduce (i) from (vi), use the $\varepsilon/2^n$ trick to show that E is contained in a Lebesgue measurable set E'_ε with $m^*(E'_\varepsilon \Delta E) \leq \varepsilon$, and then take countable intersections to show that E differs from a Lebesgue measurable set by a null set.)

Exercise 1.2.8. Show that every Jordan measurable set is Lebesgue measurable.

Exercise 1.2.9 (Middle thirds Cantor set)**.** Let $I_0 := [0, 1]$ be the unit interval, let $I_1 := [0, 1/3] \cup [2/3, 1]$ be I_0 with the interior of the middle third interval removed, let $I_2 := [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$ be I_1 with the interior of the middle third of each of the two intervals of I_1 removed, and so forth. More formally, write

$$I_n := \bigcup_{a_1,\dots,a_n \in \{0,2\}} [\sum_{i=1}^n \frac{a_i}{3^i}, \sum_{i=1}^n \frac{a_i}{3^i} + \frac{1}{3^n}].$$

Let $C := \bigcap_{n=1}^\infty I_n$ be the intersection of all the elementary sets I_n. Show that C is compact, uncountable, and a null set.

Exercise 1.2.10. (This exercise presumes some familiarity with point-set topology.) Show that the half-open interval $[0, 1)$ cannot be expressed as the countable union of disjoint closed intervals. (*Hint:* It is easy to prevent $[0, 1)$ from being expressed as the finite union of disjoint closed intervals. Next, assume for sake of contradiction that $[0, 1)$ is the union of infinitely many closed intervals, and conclude that $[0, 1)$ is homeomorphic to the middle thirds Cantor set, which is absurd. It is also possible to proceed using the *Baire category theorem* (§1.7 of *An epsilon of room*, Vol. I.) For an additional challenge, show that $[0, 1)$ cannot be expressed as the countable union of disjoint closed *sets*.

Now we look at the Lebesgue measure $m(E)$ of a Lebesgue measurable set E, which is defined to equal its Lebesgue outer measure $m^*(E)$. If E is Jordan measurable, we see from (1.2) that the Lebesgue measure and the Jordan measure of E coincide, thus Lebesgue measure extends Jordan measure. This justifies the use of the notation $m(E)$ to denote both Lebesgue measure of a Lebesgue measurable set, and Jordan measure of a Jordan measurable set (as well as elementary measure of an elementary set).

Lebesgue measure obeys significantly better properties than Lebesgue outer measure, when restricted to Lebesgue measurable sets:

Lemma 1.2.15 (The measure axioms)**.**

(i) *(Empty set)* $m(\emptyset) = 0$.

(ii) *(Countable additivity) If $E_1, E_2, \ldots \subset \mathbf{R}^d$ is a countable sequence of disjoint Lebesgue measurable sets, then* $m(\bigcup_{n=1}^\infty E_n) = \sum_{n=1}^\infty m(E_n)$.

Proof. The first claim is trivial, so we focus on the second. We deal with an easy case when all of the E_n are compact. By repeated use of Lemma 1.2.5 and Exercise 1.2.4, we have

$$m(\bigcup_{n=1}^N E_n) = \sum_{n=1}^N m(E_n).$$

Using monotonicity, we conclude that

$$m(\bigcup_{n=1}^\infty E_n) \geq \sum_{n=1}^N m(E_n).$$

(We can use m instead of m^* throughout this argument, thanks to Lemma 1.2.13). Sending $N \to \infty$ we obtain

$$m(\bigcup_{n=1}^\infty E_n) \geq \sum_{n=1}^\infty m(E_n).$$

On the other hand, from countable subadditivity one has

$$m(\bigcup_{n=1}^\infty E_n) \leq \sum_{n=1}^\infty m(E_n),$$

and the claim follows.

Next, we handle the case when the E_n are bounded but not necessarily compact. We use the $\varepsilon/2^n$ trick. Let $\varepsilon > 0$. Applying Exercise 1.2.7, we know that each E_n is the union of a compact set K_n and a set of outer measure at most $\varepsilon/2^n$. Thus

$$m(E_n) \leq m(K_n) + \varepsilon/2^n$$

and hence

$$\sum_{n=1}^\infty m(E_n) \leq (\sum_{n=1}^\infty m(K_n)) + \varepsilon.$$

Finally, from the compact case of this lemma we already know that

$$m(\bigcup_{n=1}^\infty K_n) = \sum_{n=1}^\infty m(K_n)$$

while from monotonicity

$$m(\bigcup_{n=1}^\infty K_n) \leq m(\bigcup_{n=1}^\infty E_n).$$

Putting all this together we see that

$$\sum_{n=1}^{\infty} m(E_n) \leq m(\bigcup_{n=1}^{\infty} E_n) + \varepsilon$$

for every $\varepsilon > 0$, while from countable subadditivity we have

$$m(\bigcup_{n=1}^{\infty} E_n) \leq \sum_{n=1}^{\infty} m(E_n).$$

The claim follows.

Finally, we handle the case when the E_n are not assumed to be bounded or closed. Here, the basic idea is to decompose each E_n as a countable disjoint union of bounded Lebesgue measurable sets. First, decompose $\mathbf{R}^d$ as the countable disjoint union $\mathbf{R}^d = \bigcup_{m=1}^{\infty} A_m$ of bounded measurable sets A_m; for instance one could take the annuli $A_m := \{x \in \mathbf{R}^d : m-1 \leq |x| < m\}$. Then each E_n is the countable disjoint union of the bounded measurable sets $E_n \cap A_m$ for $m = 1, 2, \ldots$, and thus

$$m(E_n) = \sum_{m=1}^{\infty} m(E_n \cap A_m)$$

by the previous arguments. In a similar vein, $\bigcup_{n=1}^{\infty} E_n$ is the countable disjoint union of the bounded measurable sets $E_n \cap A_m$ for $n, m = 1, 2, \ldots$, and thus

$$m(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} m(E_n \cap A_m),$$

and the claim follows. □

From Lemma 1.2.15 one of course can conclude finite additivity

$$m(E_1 \cup \ldots \cup E_k) = m(E_1) + \ldots + m(E_k)$$

whenever $E_1, \ldots, E_k \subset \mathbf{R}^d$ are Lebesgue measurable sets. We also have another important result:

Exercise 1.2.11 (Monotone convergence theorem for measurable sets)**.**

(i) (Upward monotone convergence) Let $E_1 \subset E_2 \subset \ldots \subset \mathbf{R}^n$ be a countable non-decreasing sequence of Lebesgue measurable sets. Show that $m(\bigcup_{n=1}^\infty E_n) = \lim_{n\to\infty} m(E_n)$. (*Hint:* Express $\bigcup_{n=1}^\infty E_n$ as the countable union of the lacunae $E_n \backslash \bigcup_{n'=1}^{n-1} E_{n'}$.)

(ii) (Downward monotone convergence) Let $\mathbf{R}^d \supset E_1 \supset E_2 \supset \ldots$ be a countable non-increasing sequence of Lebesgue measurable sets. If at least one of the $m(E_n)$ is finite, show that $m(\bigcap_{n=1}^\infty E_n) = \lim_{n\to\infty} m(E_n)$.

(iii) Give a counterexample to show that the hypothesis that at least one of the $m(E_n)$ is finite in the downward monotone convergence theorem cannot be dropped.

Exercise 1.2.12. Show that any map $E \mapsto m(E)$ from Lebesgue measurable sets to elements of $[0, +\infty]$ that obeys the above empty set and countable additivity axioms will also obey the monotonicity and countable subadditivity axioms from Exercise 1.2.3, when restricted to Lebesgue measurable sets of course.

Exercise 1.2.13. We say that a sequence E_n of sets in $\mathbf{R}^d$ *converges pointwise* to another set E in $\mathbf{R}^d$ if the *indicator functions* 1_{E_n} converge pointwise to 1_E.

(i) Show that if the E_n are all Lebesgue measurable, and converge pointwise to E, then E is Lebesgue measurable also. (*Hint:* use the identity $1_E(x) = \liminf_{n\to\infty} 1_{E_n}(x)$ or $1_E(x) = \limsup_{n\to\infty} 1_{E_n}(x)$ to write E in terms of countable unions and intersections of the E_n.)

(ii) (Dominated convergence theorem) Suppose that the E_n are all contained in another Lebesgue measurable set F of finite measure. Show that $m(E_n)$ converges to $m(E)$. (*Hint:* use the upward and downward monotone convergence theorems, Exercise 1.2.11.)

(iii) Give a counterexample to show that the dominated convergence theorem fails if the E_n are not contained in a set of finite measure, even if we assume that the $m(E_n)$ are all uniformly bounded.

In later sections we will generalise the monotone and dominated convergence theorems to measurable functions instead of measurable sets; see Theorem 1.4.44 and Theorem 1.4.49.

Exercise 1.2.14. Let $E \subset \mathbf{R}^d$. Show that E is contained in a Lebesgue measurable set of measure exactly equal to $m^*(E)$.

Exercise 1.2.15 (Inner regularity)**.** Let $E \subset \mathbf{R}^d$ be Lebesgue measurable. Show that

$$m(E) = \sup_{K \subset E, K \text{ compact}} m(K).$$

Remark 1.2.16. The inner and outer regularity properties of measure can be used to define the concept of a *Radon measure* (see §1.10 of *An epsilon of room, Vol. I.*).

Exercise 1.2.16 (Criteria for finite measure)**.** Let $E \subset \mathbf{R}^d$. Show that the following are equivalent:

(i) E is Lebesgue measurable with finite measure.

(ii) (Outer approximation by open) For every $\varepsilon > 0$, one can contain E in an open set U of finite measure with $m^*(U \backslash E) \leq \varepsilon$.

(iii) (Almost open bounded) E differs from a bounded open set by a set of arbitrarily small Lebesgue outer measure. (In other words, for every $\varepsilon > 0$ there exists a bounded open set U such that $m^*(E \Delta U) \leq \varepsilon$.)

(iv) (Inner approximation by compact) For every $\varepsilon > 0$, one can find a compact set F contained in E with $m^*(E \backslash F) \leq \varepsilon$.

(v) (Almost compact) E differs from a compact set by a set of arbitrarily small Lebesgue outer measure.

(vi) (Almost bounded measurable) E differs from a bounded Lebesgue measurable set by a set of arbitrarily small Lebesgue outer measure.

(vii) (Almost finite measure) E differs from a Lebesgue measurable set with finite measure by a set of arbitrarily small Lebesgue outer measure.

(viii) (Almost elementary) E differs from an elementary set by a set of arbitrarily small Lebesgue outer measure.

(ix) (Almost dyadically elementary) For every $\varepsilon > 0$, there exists an integer n and a finite union F of closed dyadic cubes of sidelength 2^{-n} such that $m^*(E \Delta F) \leq \varepsilon$.

One can interpret the equivalence of (i) and (ix) in the above exercise as asserting that Lebesgue measurable sets are those which look (locally) "pixelated" at sufficiently fine scales. This will be formalised in later sections with the *Lebesgue differentiation theorem* (Exercise 1.6.24).

Exercise 1.2.17 (Carathéodory criterion, one direction)**.** Let $E \subset \mathbf{R}^d$. Show that the following are equivalent:

(i) E is Lebesgue measurable.

(ii) For every elementary set A, one has $m(A) = m^*(A \cap E) + m^*(A \backslash E)$.

(iii) For every box B, one has $|B| = m^*(B \cap E) + m^*(B \backslash E)$.

Exercise 1.2.18 (Inner measure)**.** Let $E \subset \mathbf{R}^d$ be a bounded set. Define the *Lebesgue inner measure* $m_*(E)$ of E by the formula

$$m_*(E) := m(A) - m^*(A \backslash E)$$

for any elementary set A containing E.

(i) Show that this definition is well defined, i.e. that if A, A' are two elementary sets containing E, that $m(A) - m^*(A \backslash E)$ is equal to $m(A') - m^*(A' \backslash E)$.

(ii) Show that $m_*(E) \leq m^*(E)$, and that equality holds if and only if E is Lebesgue measurable.

Define a G_δ *set* to be a countable intersection $\bigcap_{n=1}^\infty U_n$ of open sets, and an F_σ *set* to be a countable union $\bigcup_{n=1}^\infty F_n$ of closed sets.

Exercise 1.2.19. Let $E \subset \mathbf{R}^d$. Show that the following are equivalent:

(i) E is Lebesgue measurable.

(ii) E is a G_δ set with a null set removed.

(iii) E is the union of a F_σ set and a null set.

Remark 1.2.17. From the above exercises, we see that when describing what it means for a set to be Lebesgue measurable, there is a tradeoff between the type of approximation one is willing to bear, and the type of things one can say about the approximation. If one is only willing to approximate to within a null set, then one can only say that a measurable set is approximated by a G_δ or a F_σ set, which is a fairly weak amount of structure. If one is willing to add on an epsilon of error (as measured in the Lebesgue outer measure), one can make a measurable set open; dually, if one is willing to take away an epsilon of error, one can make a measurable set closed. Finally, if one is willing to both add and subtract an epsilon of error, then one can make a measurable set (of finite measure) elementary, or even a finite union of dyadic cubes.

Exercise 1.2.20 (Translation invariance)**.** If $E \subset \mathbf{R}^d$ is Lebesgue measurable, show that $E + x$ is Lebesgue measurable for any $x \in \mathbf{R}^d$, and that $m(E + x) = m(E)$.

Exercise 1.2.21 (Change of variables)**.** If $E \subset \mathbf{R}^d$ is Lebesgue measurable, and $T\colon \mathbf{R}^d \to \mathbf{R}^d$ is a linear transformation, show that $T(E)$ is Lebesgue measurable, and that $m(T(E)) = |\det T| m(E)$. We caution that if $T\colon \mathbf{R}^d \to \mathbf{R}^{d'}$ is a linear map to a space $\mathbf{R}^{d'}$ of strictly smaller dimension than $\mathbf{R}^d$, then $T(E)$ need not be Lebesgue measurable; see Exercise 1.2.27.

Exercise 1.2.22. Let $d, d' \geq 1$ be natural numbers.

(i) If $E \subset \mathbf{R}^d$ and $F \subset \mathbf{R}^{d'}$, show that $(m^{d+d'})^*(E \times F) \leq (m^d)^*(E)(m^{d'})^*(F)$, where $(m^d)^*$ denotes d-dimensional Lebesgue measure, etc.

(ii) Let $E \subset \mathbf{R}^d$, $F \subset \mathbf{R}^{d'}$ be Lebesgue measurable sets. Show that $E \times F \subset \mathbf{R}^{d+d'}$ is Lebesgue measurable, with $m^{d+d'}(E \times F) = m^d(E) \cdot m^{d'}(F)$. (Note that we allow E or F to have infinite measure, and so one may have to divide into cases or take advantage of the monotone convergence theorem for Lebesgue measure, Exercise 1.2.11.)

Exercise 1.2.23 (Uniqueness of Lebesgue measure)**.** Show that Lebesgue measure $E \mapsto m(E)$ is the only map from Lebesgue measurable sets to $[0, +\infty]$ that obeys the following axioms:

(i) (Empty set) $m(\emptyset) = 0$.

(ii) (Countable additivity) If $E_1, E_2, \ldots \subset \mathbf{R}^d$ is a countable sequence of disjoint Lebesgue measurable sets, then $m(\bigcup_{n=1}^\infty E_n) = \sum_{n=1}^\infty m(E_n)$.

(iii) (Translation invariance) If E is Lebesgue measurable and $x \in \mathbf{R}^d$, then $m(E + x) = m(E)$.

(iv) (Normalisation) $m([0,1]^d) = 1$.

Hint: First show that m must match elementary measure on elementary sets, then show that m is bounded by outer measure.

Exercise 1.2.24 (Lebesgue measure as the completion of elementary measure)**.** The purpose of the following exercise is to indicate how Lebesgue measure can be viewed as a metric completion of elementary measure in some sense. To avoid some technicalities we will not work in all of $\mathbf{R}^d$, but in some fixed elementary set A (e.g. $A = [0,1]^d$).

(i) Let $2^A := \{E : E \subset A\}$ be the power set of A. We say that two sets $E, F \in 2^A$ are *equivalent* if $E \Delta F$ is a null set. Show that this is a equivalence relation.

(ii) Let $2^A / \sim$ be the set of equivalence classes $[E] := \{F \in 2^A : E \sim F\}$ of 2^A with respect to the above equivalence relation. Define a distance $d \colon 2^A / \sim \times 2^A / \sim \to \mathbf{R}^+$ between two equivalence classes $[E], [E']$ by defining $d([E], [E']) := m^*(E \Delta E')$. Show that this distance is well-defined (in the sense that $m(E \Delta E') = m(F \Delta F')$ whenever $[E] = [F]$ and $[E'] = [F']$) and gives $2^A / \sim$ the structure of a complete metric space.

(iii) Let $\mathcal{E} \subset 2^A$ be the elementary subsets of A, and let $\mathcal{L} \subset 2^A$ be the Lebesgue measurable subsets of A. Show that $\mathcal{L} / \sim$ is the closure of $\mathcal{E} / \sim$ with respect to the metric defined above. In particular, $\mathcal{L} / \sim$ is a complete metric space that contains $\mathcal{E} / \sim$ as a dense subset; in other words, $\mathcal{L} / \sim$ is a *metric completion* of $\mathcal{E} / \sim$.

(iv) Show that Lebesgue measure $m\colon \mathcal{L} \to \mathbf{R}^+$ descends to a continuous function $m\colon \mathcal{L}/\sim \to \mathbf{R}^+$, which by abuse of notation we shall still call m. Show that $m\colon \mathcal{L}/\sim \to \mathbf{R}^+$ is the unique continuous extension of the analogous elementary measure function $m\colon \mathcal{E}/\sim \to \mathbf{R}^+$ to $\mathcal{L}/\sim$.

For a further discussion of how measures can be viewed as completions of elementary measures, see §2.1 of *An epsilon of room, Vol. I.*

Exercise 1.2.25. Define a *continuously differentiable curve* in $\mathbf{R}^d$ to be a set of the form $\{\gamma(t) : a \leq t \leq b\}$ where $[a,b]$ is a closed interval and $\gamma\colon [a,b] \to \mathbf{R}^d$ is a continuously differentiable function.

(i) If $d \geq 2$, show that every continuously differentiable curve has Lebesgue measure zero. (Why is the condition $d \geq 2$ necessary?)

(ii) Conclude that if $d \geq 2$, then the unit cube $[0,1]^d$ cannot be covered by countably many continuously differentiable curves.

We remark that if the curve is only assumed to be continuous, rather than continuously differentiable, then these claims fail, thanks to the existence of *space-filling curves.*

1.2.3. Non-measurable sets. In the previous section we have set out a rich theory of Lebesgue measure, which enjoys many nice properties when applied to Lebesgue measurable sets.

Thus far, we have not ruled out the possibility that every single set is Lebesgue measurable. There is good reason for this: a famous theorem of Solovay[**So1970**] asserts that, if one is willing to drop the axiom of choice, there exist models of set theory in which all subsets of $\mathbf{R}^d$ are measurable. So any demonstration of the existence of non-measurable sets must use the axiom of choice in some essential way.

That said, we can give an informal (and highly non-rigorous) motivation as to why non-measurable sets should exist, using intuition from probability theory rather than from set theory. The starting point is the observation that Lebesgue sets of finite measure (and in particular, bounded Lebesgue sets) have to be "almost elementary", in the sense of Exercise 1.2.16. So all we need to do to build a non-measurable set is to exhibit a bounded set which is not almost elementary. Intuitively, we want to build a set which has oscillatory structure even at arbitrarily fine scales.

We will non-rigorously do this as follows. We will work inside the unit interval $[0,1]$. For each $x \in [0,1]$, we imagine that we flip a coin to give either heads or tails (with an independent coin flip for each x), and let $E \subset [0,1]$ be the set of all the $x \in [0,1]$ for which the coin flip came up heads. We suppose for contradiction that E is Lebesgue measurable. Intuitively, since each x had a 50% chance of being heads, E should occupy about "half" of $[0,1]$; applying the *law of large numbers* (see e.g. [**Ta2009**, §1.4]) in an extremely non-rigorous fashion, we thus expect $m(E)$ to equal $1/2$.

Moreover, given any subinterval $[a,b]$ of $[0,1]$, the same reasoning leads us to expect that $E \cap [a,b]$ should occupy about half of $[a,b]$, so that $m(E \cap [a,b])$ should be $|[a,b]|/2$. More generally, given any elementary set F in $[0,1]$, we should have $m(E \cap F) = m(F)/2$. This makes it very hard for E to be approximated by an elementary set; indeed, a little algebra then shows that $m(E \Delta F) = 1/2$ for any elementary $F \subset [0,1]$. Thus E is not Lebesgue measurable.

Unfortunately, the above argument is terribly non-rigorous for a number of reasons, not the least of which is that it uses an uncountable number of coin flips, and the rigorous probabilistic theory that one would have to use to model such a system of random variables is too weak[12] to be able to assign meaningful probabilities to such events as "E is Lebesgue measurable". So we now turn to more rigorous arguments that establish the existence of non-measurable sets. The arguments will be fairly simple, but the sets constructed are somewhat artificial in nature.

Proposition 1.2.18. *There exists a subset $E \subset [0,1]$ which is not Lebesgue measurable.*

Proof. We use the fact that the rationals $\mathbf{Q}$ are an additive subgroup of the reals $\mathbf{R}$, and so partition the reals $\mathbf{R}$ into disjoint cosets $x + \mathbf{Q}$. This creates a quotient group $\mathbf{R}/\mathbf{Q} := \{x + \mathbf{Q} : x \in \mathbf{R}\}$. Each coset C of $\mathbf{R}/\mathbf{Q}$ is dense in $\mathbf{R}$, and so has a non-empty intersection

[12] For some further discussion of this point, see [**Ta2009**, §1.10].

with $[0,1]$. Applying the axiom of choice, we may thus find an element $x_C \in C \cap [0,1]$ for each $C \in \mathbf{R}/\mathbf{Q}$. We then let $E := \{x_C : C \in \mathbf{R}/\mathbf{Q}\}$ be the collection of all these coset representatives. By construction, $E \subset [0,1]$.

Let y be any element of $[0,1]$. Then it must lie in some coset C of $\mathbf{R}/\mathbf{Q}$, and thus differs from x_C by some rational number in $[-1,1]$. In other words, we have

$$[0,1] \subset \bigcup_{q \in \mathbf{Q} \cap [-1,1]} (E+q). \tag{1.4}$$

On the other hand, we clearly have

$$\bigcup_{q \in \mathbf{Q} \cap [-1,1]} (E+q) \subset [-1,2]. \tag{1.5}$$

Also, the different translates $E+q$ are disjoint, because E contains only one element from each coset of $\mathbf{Q}$.

We claim that E is not Lebesgue measurable. To see this, suppose for contradiction that E was Lebesgue measurable. Then the translates $E+q$ would also be Lebesgue measurable. By countable additivity, we thus have

$$m(\bigcup_{q \in \mathbf{Q} \cap [-1,1]} (E+q)) = \sum_{q \in \mathbf{Q} \cap [-1,1]} m(E+q),$$

and thus by translation invariance and (1.4), (1.5)

$$1 \leq \sum_{q \in \mathbf{Q} \cap [-1,1]} m(E) \leq 3.$$

On the other hand, the sum $\sum_{q \in \mathbf{Q} \cap [-1,1]} m(E)$ is either zero (if $m(E) = 0$) or infinite (if $m(E) > 0$), leading to the desired contradiction. $\square$

Exercise 1.2.26 (Outer measure is not finitely additive)**.** Show that there exists disjoint bounded subsets E, F of the real line such that $m^*(E \cup F) \neq m^*(E) + m^*(F)$. (*Hint:* Show that the set constructed in the proof of the above proposition has positive outer measure.)

Exercise 1.2.27 (Projections of measurable sets need not be measurable)**.** Let $\pi \colon \mathbf{R}^2 \to \mathbf{R}$ be the coordinate projection $\pi(x,y) := x$. Show that there exists a measurable subset E of $\mathbf{R}^2$ such that $\pi(E)$ is not measurable.

Remark 1.2.19. The above discussion shows that, in the presence of the axiom of choice, one cannot hope to extend Lebesgue measure to arbitrary subsets of $\mathbf{R}$ while retaining both the countable additivity and the translation invariance properties. If one drops the translation invariant requirement, then this question concerns the theory of *measurable cardinals*, and is not decidable from the standard ZFC axioms. On the other hand, one can construct *finitely additive* translation invariant extensions of Lebesgue measure to the power set of $\mathbf{R}$ by use of the *Hahn-Banach theorem* (§1.5 of *An epsilon of room, Vol. I*) to extend the integration functional, though we will not do so here.

1.3. The Lebesgue integral

In Section 1.2, we defined the *Lebesgue measure* $m(E)$ of a Lebesgue measurable set $E \subset \mathbf{R}^d$, and set out the basic properties of this measure. In this set of notes, we use Lebesgue measure to define the *Lebesgue integral*

$$\int_{\mathbf{R}^d} f(x)\ dx$$

of functions $f\colon \mathbf{R}^d \to \mathbf{C} \cup \{\infty\}$. Just as not every set can be measured by Lebesgue measure, not every function can be integrated by the Lebesgue integral; the function will need to be *Lebesgue measurable*. Furthermore, the function will either need to be unsigned (taking values on $[0, +\infty]$), or absolutely integrable.

To motivate the Lebesgue integral, let us first briefly review two simpler integration concepts. The first is that of an infinite summation

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

of a sequence of numbers c_n, which can be viewed as a discrete analogue of the Lebesgue integral. Actually, there are two overlapping, but different, notions of summation that we wish to recall here. The first is that of the *unsigned infinite sum*, when the c_n lie in the extended non-negative real axis $[0, +\infty]$. In this case, the infinite sum can be defined as the limit of the partial sums

$$\sum_{n=1}^{\infty} c_n = \lim_{N\to\infty} \sum_{n=1}^{N} c_n \tag{1.6}$$

or equivalently as a supremum of arbitrary finite partial sums:

$$\sum_{n=1}^{\infty} c_n = \sup_{A \subset \mathbf{N}, A \text{ finite}} \sum_{n\in A} c_n. \tag{1.7}$$

The unsigned infinite sum $\sum_{n=1}^{\infty} c_n$ always exists, but its value may be infinite, even when each term is individually finite (consider e.g. $\sum_{n=1}^{\infty} 1$).

The second notion of a summation is the *absolutely summable infinite sum*, in which the c_n lie in the complex plane $\mathbf{C}$ and obey the absolute summability condition

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

where the left-hand side is of course an unsigned infinite sum. When this occurs, one can show that the partial sums $\sum_{n=1}^{N} c_n$ converge to a limit, and we can then define the infinite sum by the same formula (1.6) as in the unsigned case, though now the sum takes values in $\mathbf{C}$ rather than $[0, +\infty]$. The absolute summability condition confers a number of useful properties that are not obeyed by sums that are merely conditionally convergent; most notably, the value of an absolutely convergent sum is unchanged if one rearranges the terms in the series in an arbitrary fashion. Note also that the absolutely summable infinite sums can be defined in terms of the unsigned infinite sums by taking advantage of the formulae

$$\sum_{n=1}^{\infty} c_n = (\sum_{n=1}^{\infty} \mathrm{Re}(c_n)) + i(\sum_{n=1}^{\infty} \mathrm{Im}(c_n))$$

for complex absolutely summable c_n, and

$$\sum_{n=1}^{\infty} c_n = \sum_{n=1}^{\infty} c_n^+ - \sum_{n=1}^{\infty} c_n^-$$

for real absolutely summable c_n, where $c_n^+ := \max(c_n, 0)$ and $c_n^- := \max(-c_n, 0)$ are the (magnitudes of the) positive and negative parts of c_n.

In an analogous spirit, we will first define an *unsigned Lebesgue integral* $\int_{\mathbf{R}^d} f(x)\ dx$ of (measurable) unsigned functions $f\colon \mathbf{R}^d \to [0, +\infty]$, and then use that to define the *absolutely convergent Lebesgue integral* $\int_{\mathbf{R}^d} f(x)\ dx$ of *absolutely integrable* functions $f\colon \mathbf{R}^d \to \mathbf{C} \cup \{\infty\}$. (In contrast to absolutely summable series, which cannot have any infinite terms, absolutely integrable functions will be allowed to occasionally become infinite. However, as we will see, this can only happen on a set of Lebesgue measure zero.)

To define the unsigned Lebesgue integral, we now turn to another more basic notion of integration, namely the $\int_a^b f(x)\ dx$ of a Riemann integrable function $f\colon [a, b] \to \mathbf{R}$. Recall from Section 1.1 that this integral is equal to the *lower Darboux integral*

$$\int_a^b f(x) = \underline{\int_a^b} f(x)\ dx := \sup_{g \leq f; g \text{ piecewise constant}} \text{p.c.} \int_a^b g(x)\ dx.$$

(It is also equal to the upper Darboux integral; but much as the theory of Lebesgue measure is easiest to define by relying solely on outer measure and not on inner measure, the theory of the unsigned Lebesgue integral is easiest to define by relying solely on lower integrals rather than upper ones; the upper integral is somewhat problematic when dealing with "improper" integrals of functions that are unbounded or are supported on sets of infinite measure.) Compare this formula also with (1.7). The integral p.c. $\int_a^b g(x)\ dx$ is a *piecewise constant* integral, formed by breaking up the piecewise constant functions g, h into finite linear combinations of *indicator functions* 1_I of intervals I, and then measuring the length of each interval.

It turns out that virtually the same definition allows us to define a *lower Lebesgue integral* $\underline{\int_{\mathbf{R}^d}} f(x)\ dx$ of any unsigned function $f\colon \mathbf{R}^d \to [0, +\infty]$, simply by replacing intervals with the more general class of Lebesgue measurable sets (and thus replacing piecewise constant functions with the more general class of *simple functions*). If the function is *Lebesgue measurable* (a concept that we will define presently), then we refer to the lower Lebesgue integral simply as the

Lebesgue integral. As we shall see, it obeys all the basic properties one expects of an integral, such as monotonicity and additivity; in subsequent notes we will also see that it behaves quite well with respect to limits, as we shall see by establishing the two basic convergence theorems of the unsigned Lebesgue integral, namely *Fatou's lemma* (Corollary 1.4.47) and the *monotone convergence theorem* (Theorem 1.4.44).

Once we have the theory of the unsigned Lebesgue integral, we will then be able to define the absolutely convergent Lebesgue integral, similarly to how the absolutely convergent infinite sum can be defined using the unsigned infinite sum. This integral also obeys all the basic properties one expects, such as linearity and compatibility with the more classical Riemann integral; in subsequent notes we will see that it also obeys a fundamentally important convergence theorem, the *dominated convergence theorem* (Theorem 1.4.49). This convergence theorem makes the Lebesgue integral (and its abstract generalisations to other measure spaces than $\mathbf{R}^d$) particularly suitable for analysis, as well as allied fields that rely heavily on limits of functions, such as PDE, probability, and ergodic theory.

Remark 1.3.1. This is not the only route to setting up the unsigned and absolutely convergent Lebesgue integrals. For instance, one can proceed with the unsigned integral but then making an auxiliary stop at integration of functions that are bounded and are supported on a set of finite measure, before going to the absolutely convergent Lebesgue integral; see e.g. [**StSk2005**]. Another approach (which will not be discussed here) is to take the metric completion of the Riemann integral with respect to the L^1 metric.

The Lebesgue integral and Lebesgue measure can be viewed as *completions* of the Riemann integral and Jordan measure respectively. This means three things. Firstly, the Lebesgue theory *extends* the Riemann theory: every Jordan measurable set is Lebesgue measurable, and every Riemann integrable function is Lebesgue measurable, with the measures and integrals from the two theories being compatible. Conversely, the Lebesgue theory can be *approximated* by the Riemann theory; as we saw in Section 1.2, every Lebesgue measurable set can be approximated (in various senses) by simpler sets, such as open sets or elementary sets, and in a similar fashion, Lebesgue measurable functions can be approximated by nicer functions, such as Riemann integrable or continuous functions. Finally, the Lebesgue theory is *complete* in various ways; this is formalised in §1.3 of *An epsilon of room, Vol. I*, but the convergence theorems mentioned above already hint at this completeness. A related fact, known as *Egorov's theorem*, asserts that a pointwise converging sequence of functions can be approximated as a (locally) uniformly converging sequence of functions. The facts listed here manifestations of *Littlewood's three principles of real analysis* (Section 1.3.5), which capture much of the essence of the Lebesgue theory.

1.3.1. Integration of simple functions. Much as the Riemann integral was set up by first using the integral for piecewise constant functions, the Lebesgue integral is set up using the integral for simple functions.

Definition 1.3.2 (Simple function)**.** A (complex-valued) *simple function* $f\colon \mathbf{R}^d \to \mathbf{C}$ is a finite linear combination

$$f = c_1 1_{E_1} + \ldots + c_k 1_{E_k} \tag{1.8}$$

of indicator functions 1_{E_i} of Lebesgue measurable sets $E_i \subset \mathbf{R}^d$ for $i = 1, \ldots, k$, where $k \geq 0$ is a natural number and $c_1, \ldots, c_k \in \mathbf{C}$ are complex numbers. An *unsigned simple function* $f\colon \mathbf{R}^d \to [0, +\infty]$, is defined similarly, but with the c_i taking values in $[0, +\infty]$ rather than $\mathbf{C}$.

It is clear from construction that the space $\mathrm{Simp}(\mathbf{R}^d)$ of complex-valued simple functions forms a complex vector space; also, $\mathrm{Simp}(\mathbf{R}^d)$ also closed under pointwise product $f, g \mapsto fg$ and complex conjugation $f \mapsto \overline{f}$. In short, $\mathrm{Simp}(\mathbf{R}^d)$ is a *commutative $*$-algebra.* Meanwhile, the space $\mathrm{Simp}^+(\mathbf{R}^d)$ of unsigned simple functions is a $[0, +\infty]$-module; it is closed under addition, and under scalar multiplication by elements in $[0, +\infty]$.

In this definition, we did not require the $E_1, \ldots, E_k$ to be disjoint. However, it is easy enough to arrange this, basically by exploiting Venn diagrams (or, to use fancier language, finite boolean algebras). Indeed, any k subsets $E_1, \ldots, E_k$ of $\mathbf{R}^d$ partition $\mathbf{R}^d$ into 2^k disjoint sets, each of which is an intersection of E_i or the complement $\mathbf{R}^d \backslash E_i$ for $i = 1, \ldots, k$ (and in particular, is measurable). The (complex or unsigned) simple function is constant on each of these sets, and so can easily be decomposed as a linear combination of the indicator function of these sets. One easy consequence of this is that if f is a complex-valued simple function, then its absolute value $|f|: x \mapsto |f(x)|$ is an unsigned simple function.

It is geometrically intuitive that we should define the integral $\int_{\mathbf{R}^d} 1_E(x)\ dx$ of an indicator function of a measurable set E to equal $m(E)$:

$$\int_{\mathbf{R}^d} 1_E(x)\ dx = m(E).$$

Using this and applying the laws of integration formally, we are led to propose the following definition for the integral of an unsigned simple function:

Definition 1.3.3 (Integral of a unsigned simple function)**.** If $f = c_1 1_{E_1} + \ldots + c_k 1_{E_k}$ is an unsigned simple function, the integral $\operatorname{Simp} \int_{\mathbf{R}^d} f(x)\ dx$ is defined by the formula

$$\operatorname{Simp} \int_{\mathbf{R}^d} f(x)\ dx := c_1 m(E_1) + \ldots + c_k m(E_k),$$

thus $\operatorname{Simp} \int_{\mathbf{R}^d} f(x)\ dx$ will take values in $[0, +\infty]$.

However, one has to actually check that this definition is well-defined, in the sense that different representations

$$f = c_1 1_{E_1} + \ldots + c_k 1_{E_k} = c'_1 1_{E'_1} + \ldots + c'_{k'} 1_{E'_{k'}}$$

of a function as a finite unsigned combination of indicator functions of measurable sets will give the same value for the integral $\operatorname{Simp} \int_{\mathbf{R}^d} f(x)\ dx$. This is the purpose of the following lemma:

Lemma 1.3.4 (Well-definedness of simple integral)**.** *Let $k, k' \geq 0$ be natural numbers, $c_1, \ldots, c_k, c'_1, \ldots, c'_{k'} \in [0, +\infty]$, and let $E_1, \ldots, E_k, E'_1, \ldots, E'_{k'} \subset \mathbf{R}^d$ be Lebesgue measurable sets such that the identity*

$$c_1 1_{E_1} + \ldots + c_k 1_{E_k} = c'_1 1_{E'_1} + \ldots + c'_{k'} 1_{E'_{k'}} \tag{1.9}$$

holds identically on $\mathbf{R}^d$. Then one has

$$c_1 m(E_1) + \ldots + c_k m(E_k) = c'_1 m(E'_1) + \ldots + c'_{k'} m(E'_{k'}).$$

Proof. We again use a Venn diagram argument. The $k + k'$ sets $E_1, \ldots, E_k, E'_1, \ldots, E'_{k'}$ partition $\mathbf{R}^d$ into $2^{k+k'}$ disjoint sets, each of which is an intersection of some of the $E_1, \ldots, E_k, E'_1, \ldots, E'_{k'}$ and their complements. We throw away any sets that are empty, leaving us with a partition of $\mathbf{R}^d$ into m non-empty disjoint sets $A_1, \ldots, A_m$ for some $0 \leq m \leq 2^{k+k'}$. As the $E_1, \ldots, E_k, E'_1, \ldots, E'_k$ are Lebesgue measurable, the $A_1, \ldots, A_m$ are too. By construction, each of the $E_1, \ldots, E_k, E'_1, \ldots, E_{k'}$ arise as unions of some of the $A_1, \ldots, A_m$, thus we can write

$$E_i = \bigcup_{j \in J_i} A_j$$

and

$$E'_{i'} = \bigcup_{j' \in J'_{i'}} A_{j'}$$

for all $i = 1, \ldots, k$ and $i' = 1, \ldots, k'$, and some subsets $J_i, J'_{i'} \subset \{1, \ldots, m\}$. By finite additivity of Lebesgue measure, we thus have

$$m(E_i) = \sum_{j \in J_i} m(A_j)$$

and

$$m(E'_{i'}) = \sum_{j \in J'_{i'}} m(A_j)$$

Thus, our objective is now to show that

$$\sum_{i=1}^{k} c_i \sum_{j \in J_i} m(A_j) = \sum_{i'=1}^{k'} c'_{i'} \sum_{j \in J'_{i'}} m(A_j). \tag{1.10}$$

To obtain this, we fix $1 \leq j \leq m$ and evaluate (1.9) at a point x in the non-empty set A_j. At such a point, $1_{E_i}(x)$ is equal to $1_{J_i}(j)$, and similarly $1_{E'_{i'}}$ is equal to $1_{J'_{i'}}(j)$. From (1.9) we conclude that

$$\sum_{i=1}^{k} c_i 1_{J_i}(j) = \sum_{i'=1}^{k'} c'_{i'} 1_{J'_{i'}}(j).$$

Multiplying this by $m(A_j)$ and then summing over all $j = 1, \ldots, m$ we obtain (1.10). □

We now make some important definitions that we will use repeatedly in this text:

Definition 1.3.5 (Almost everywhere and support)**.** A property $P(x)$ of a point $x \in \mathbf{R}^d$ is said to hold *(Lebesgue) almost everywhere* in $\mathbf{R}^d$, or *for (Lebesgue) almost every point* $x \in \mathbf{R}^d$, if the set of $x \in \mathbf{R}^d$ for which $P(x)$ fails has Lebesgue measure zero (i.e. P is true outside of a null set). We usually omit the prefix Lebesgue, and often abbreviate "almost everywhere" or "almost every" as a.e.

Two functions $f, g \colon \mathbf{R}^d \to Z$ into an arbitrary range Z are said to *agree almost everywhere* if one has $f(x) = g(x)$ for almost every $x \in \mathbf{R}^d$.

The *support* of a function $f \colon \mathbf{R}^d \to \mathbf{C}$ or $f \colon \mathbf{R}^d \to [0, +\infty]$ is defined to be the set $\{x \in \mathbf{R}^d \colon f(x) \neq 0\}$ where f is non-zero.

Note that if $P(x)$ holds for almost every x, and $P(x)$ implies $Q(x)$, then $Q(x)$ holds for almost every x. Also, if $P_1(x), P_2(x), \ldots$ are an at most countable family of properties, each of which individually holds for almost every x, then they will *simultaneously* be true for almost every x, because the countable union of null sets is still a null set. Because of these properties, one can (as a rule of thumb) treat the almost universal quantifier "for almost every" as if it was the truly universal quantifier "for every", as long as one is only concatenating at most countably many properties together, and as long as one never specialises the free variable x to a null set. Observe also that the property of agreeing almost everywhere is an equivalence relation, which we will refer to as *almost everywhere equivalence*.

In *An epsilon of room, Vol. I* we will also see the notion of the *closed support* of a function $f \colon \mathbf{R}^d \to \mathbf{C}$, defined as the closure of the support.

The following properties of the simple unsigned integral are easily obtained from the definitions:

Exercise 1.3.1 (Basic properties of the simple unsigned integral)**.** Let $f, g \colon \mathbf{R}^d \to [0, +\infty]$ be simple unsigned functions.

(i) (Unsigned linearity) We have

$$\text{Simp}\int_{\mathbf{R}^d} f(x)+g(x)\ dx = \text{Simp}\int_{\mathbf{R}^d} f(x)\ dx + \text{Simp}\int_{\mathbf{R}^d} g(x)\ dx$$

and

$$\text{Simp}\int_{\mathbf{R}^d} cf(x)\ dx = c \times \text{Simp}\int_{\mathbf{R}^d} f(x)\ dx$$

for all $c \in [0,+\infty]$.

(ii) (Finiteness) We have $\text{Simp}\int_{\mathbf{R}^d} f(x)\ dx < \infty$ if and only if f is finite almost everywhere, and its support has finite measure.

(iii) (Vanishing) We have $\text{Simp}\int_{\mathbf{R}^d} f(x)\ dx = 0$ if and only if f is zero almost everywhere.

(iv) (Equivalence) If f and g agree almost everywhere, then $\text{Simp}\int_{\mathbf{R}^d} f(x)\ dx = \text{Simp}\int_{\mathbf{R}^d} g(x)\ dx$.

(v) (Monotonicity) If $f(x) \leq g(x)$ for almost every $x \in \mathbf{R}^d$, then $\text{Simp}\int_{\mathbf{R}^d} f(x)\ dx \leq \text{Simp}\int_{\mathbf{R}^d} g(x)\ dx$.

(vi) (Compatibility with Lebesgue measure) For any Lebesgue measurable E, one has $\text{Simp}\int_{\mathbf{R}^d} 1_E(x)\ dx = m(E)$.

Furthermore, show that the simple unsigned integral $f \mapsto \text{Simp}\int_{\mathbf{R}^d} f(x)\ dx$ is the only map from the space $\text{Simp}^+(\mathbf{R}^d)$ of unsigned simple functions to $[0,+\infty]$ that obeys all of the above properties.

We can now define an absolutely convergent counterpart to the simple unsigned integral. This integral will soon be superceded by the absolutely Lebesgue integral, but we give it here as motivation for that more general notion of integration.

Definition 1.3.6 (Absolutely convergent simple integral)**.** A complex-valued simple function $f\colon \mathbf{R}^d \to \mathbf{C}$ is said to be *absolutely integrable* of $\text{Simp}\int_{\mathbf{R}^d} |f(x)|\ dx < \infty$. If f is absolutely integrable, the integral $\text{Simp}\int_{\mathbf{R}^d} f(x)\ dx$ is defined for real signed f by the formula

$$\text{Simp}\int_{\mathbf{R}^d} f(x)\ dx := \text{Simp}\int_{\mathbf{R}^d} f_+(x)\ dx - \text{Simp}\int_{\mathbf{R}^d} f_-(x)\ dx$$

where $f_+(x) := \max(f(x), 0)$ and $f_-(x) := \max(-f(x), 0)$ (note that these are unsigned simple functions that are pointwise dominated by $|f|$ and thus have finite integral), and for complex-valued f by the formula[13]

$$\operatorname{Simp} \int_{\mathbf{R}^d} f(x)\ dx := \operatorname{Simp} \int_{\mathbf{R}^d} \operatorname{Re} f(x)\ dx + i \operatorname{Simp} \int_{\mathbf{R}^d} \operatorname{Im} f(x)\ dx.$$

Note from the preceding exercise that a complex-valued simple function f is absolutely integrable if and only if it has finite measure support (since finiteness almost everywhere is automatic). In particular, the space $\operatorname{Simp}^{abs}(\mathbf{R}^d)$ of absolutely integrable simple functions is closed under addition and scalar multiplication by complex numbers, and is thus a complex vector space.

The properties of the unsigned simple integral then can be used to deduce analogous properties for the complex-valued integral:

Exercise 1.3.2 (Basic properties of the complex-valued simple integral)**.** Let $f, g \colon \mathbf{R}^d \to \mathbf{C}$ be absolutely integrable simple functions.

(i) (*-linearity) We have

$$\operatorname{Simp} \int_{\mathbf{R}^d} f(x) + g(x)\ dx = \operatorname{Simp} \int_{\mathbf{R}^d} f(x)\ dx + \operatorname{Simp} \int_{\mathbf{R}^d} g(x)\ dx$$

and

$$\operatorname{Simp} \int_{\mathbf{R}^d} cf(x)\ dx = c \times \operatorname{Simp} \int_{\mathbf{R}^d} f(x)\ dx \tag{1.11}$$

for all $c \in \mathbf{C}$. Also we have

$$\operatorname{Simp} \int_{\mathbf{R}^d} \overline{f}(x)\ dx = \overline{\operatorname{Simp} \int_{\mathbf{R}^d} f(x)\ dx}.$$

(ii) (Equivalence) If f and g agree almost everywhere, then $\operatorname{Simp} \int_{\mathbf{R}^d} f(x)\ dx = \operatorname{Simp} \int_{\mathbf{R}^d} g(x)\ dx$.

[13] Strictly speaking, this is an abuse of notation as we have now defined the simple integral $\operatorname{Simp} \int_{\mathbf{R}^d}$ three different times, for unsigned, real signed, and complex-valued simple functions, but one easily verifies that these three definitions agree with each other on their common domains of definition, so it is safe to use a single notation for all three.

(iii) (Compatibility with Lebesgue measure) For any Lebesgue measurable E, one has $\mathrm{Simp}\int_{\mathbf{R}^d} 1_E(x)\ dx = m(E)$.

(*Hints:* Work out the real-valued counterpart of the linearity property first. To establish (1.11), treat the cases $c > 0, c = 0, c = -1$ separately. To deal with the additivity for real functions f, g, start with the identity

$$f + g = (f + g)_+ - (f + g)_- = (f_+ - f_-) + (g_+ - g_-)$$

and rearrange the second inequality so that no subtraction appears.) Furthermore, show that the complex-valued simple integral $f \mapsto \mathrm{Simp}\int_{\mathbf{R}^d} f(x)\ dx$ is the only map from the space $\mathrm{Simp}^{abs}(\mathbf{R}^d)$ of absolutely integrable simple functions to $\mathbf{C}$ that obeys all of the above properties.

We now comment further on the fact that (simple) functions that agree almost everywhere, have the same integral. We can view this as an assertion that integration is a *noise-tolerant* operation: one can have "noise" or "errors" in a function $f(x)$ on a null set, and this will not affect the final value of the integral. Indeed, once one has this noise tolerance, one can even integrate functions f that are not defined *everywhere* on $\mathbf{R}^d$, but merely defined *almost everywhere* on $\mathbf{R}^d$ (i.e. f is defined on some set $\mathbf{R}^d \backslash N$ where N is a null set), simply by extending f to all of $\mathbf{R}^d$ in some arbitrary fashion (e.g. by setting f equal to zero on N). This is extremely convenient for analysis, as there are many natural functions (e.g. $\frac{\sin x}{x}$ in one dimension, or $\frac{1}{|x|^\alpha}$ for various $\alpha > 0$ in higher dimensions) that are only defined almost everywhere instead of everywhere (often due to "division by zero" problems when a denominator vanishes). While such functions cannot be evaulated at certain singular points, they can still be integrated (provided they obey some integrability condition, of course, such as absolute integrability), and so one can still perform a large portion of analysis on such functions.

In fact, in the subfield of analysis known as *functional analysis*, it is convenient to abstract the notion of an almost everywhere defined function somewhat, by replacing any such function f with the *equivalence class* of almost everywhere defined functions that are equal to f almost everywhere. Such classes are then no longer functions in the standard set-theoretic sense (they do not map each point in the domain to a unique point in the range, since points in $\mathbf{R}^d$ have measure zero), but the properties of various function spaces improve when one does this (various semi-norms become norms, various topologies become Hausdorff, and so forth). See §1.3 of *An epsilon of room, Vol. I* for further discussion.

Remark 1.3.7. The "Lebesgue philosophy" that one is willing to lose control on sets of measure zero is a perspective that distinguishes Lebesgue-type analysis from other types of analysis, most notably that of *descriptive set theory*, which is also interested in studying subsets of $\mathbf{R}^d$, but can give completely different structural classifications to a pair of sets that agree almost everywhere. This loss of control on null sets is the price one has to pay for gaining access to the powerful tool of the Lebesgue integral; if one needs to control a function at absolutely *every* point, and not just almost every point, then one often needs to use other tools than integration theory (unless one has some regularity on the function, such as continuity, that lets one pass from almost everywhere true statements to everywhere true statements).

1.3.2. Measurable functions.

Much as the piecewise constant integral can be completed to the Riemann integral, the unsigned simple integral can be completed to the unsigned Lebesgue integral, by extending the class of unsigned simple functions to the larger class of unsigned Lebesgue measurable functions. One of the shortest ways to define this class is as follows:

Definition 1.3.8 (Unsigned measurable function)**.** An unsigned function $f\colon \mathbf{R}^d \to [0, +\infty]$ is *unsigned Lebesgue measurable*, or *measurable* for short, if it is the pointwise limit of unsigned simple functions, i.e. if there exists a sequence $f_1, f_2, f_3, \ldots : \mathbf{R}^d \to [0, +\infty]$ of unsigned simple functions such that $f_n(x) \to f(x)$ for every $x \in \mathbf{R}^d$.

This particular definition is not always the most tractable. Fortunately, it has many equivalent forms:

Lemma 1.3.9 (Equivalent notions of measurability)**.** *Let* $f\colon \mathbf{R}^d \to [0, +\infty]$ *be an unsigned function. Then the following are equivalent:*

(i) *f is unsigned Lebesgue measurable.*

(ii) *f is the pointwise limit of unsigned simple functions f_n (thus the limit $\lim_{n\to\infty} f_n(x)$ exists and is equal to $f(x)$ for all $x \in \mathbf{R}^d$).*

(iii) *f is the pointwise almost everywhere limit of unsigned simple functions f_n (thus the limit $\lim_{n\to\infty} f_n(x)$ exists and is equal to $f(x)$ for almost every $x \in \mathbf{R}^d$).*

(iv) *f is the supremum $f(x) = \sup_n f_n(x)$ of an increasing sequence $0 \leq f_1 \leq f_2 \leq \ldots$ of unsigned simple functions f_n, each of which are bounded with finite measure support.*

(v) *For every $\lambda \in [0, +\infty]$, the set $\{x \in \mathbf{R}^d : f(x) > \lambda\}$ is Lebesgue measurable.*

(vi) *For every $\lambda \in [0, +\infty]$, the set $\{x \in \mathbf{R}^d : f(x) \geq \lambda\}$ is Lebesgue measurable.*

(vii) *For every $\lambda \in [0, +\infty]$, the set $\{x \in \mathbf{R}^d : f(x) < \lambda\}$ is Lebesgue measurable.*

(ix) *For every $\lambda \in [0, +\infty]$, the set $\{x \in \mathbf{R}^d : f(x) \leq \lambda\}$ is Lebesgue measurable.*

(x) *For every interval $I \subset [0, +\infty)$, the set $f^{-1}(I) := \{x \in \mathbf{R}^d : f(x) \in I\}$ is Lebesgue measurable.*

(xi) *For every (relatively) open set $U \subset [0, +\infty)$, the set $f^{-1}(U) := \{x \in \mathbf{R}^d : f(x) \in U\}$ is Lebesgue measurable.*

(xii) *For every (relatively) closed set $K \subset [0, +\infty)$, the set $f^{-1}(K) := \{x \in \mathbf{R}^d : f(x) \in K\}$ is Lebesgue measurable.*

Proof. (i) and (ii) are equivalent by definition. (ii) clearly implies (iii). As every monotone sequence in $[0, +\infty]$ converges, (iv) implies (ii). Now we show that (iii) implies (v). If f is the pointwise almost everywhere limit of f_n, then for almost every $x \in \mathbf{R}^d$ one has

$$f(x) = \lim_{n\to\infty} f_n(x) = \limsup_{n\to\infty} f_n(x) = \inf_{N>0} \sup_{n \geq N} f_n(x).$$

This implies that, for any λ, the set $\{x \in \mathbf{R}^d : f(x) > \lambda\}$ is equal to

$$\bigcup_{M>0} \bigcap_{N>0} \{x \in \mathbf{R}^d : \sup_{n \geq N} f_n(x) > \lambda + \frac{1}{M}\}$$

outside of a set of measure zero; this set in turn is equal to

$$\bigcup_{M>0} \bigcap_{N>0} \bigcup_{n \geq N} \{x \in \mathbf{R}^d : f_n(x) > \lambda + \frac{1}{M}\}$$

outside of a set of measure zero. But as each f_n is an unsigned simple function, the sets $\{x \in \mathbf{R}^d : f_n(x) > \lambda + \frac{1}{M}\}$ are Lebesgue measurable. Since countable unions or countable intersections of Lebesgue measurable sets are Lebesgue measurable, and modifying a Lebesgue measurable set on a null set produces another Lebesgue measurable set, we obtain (v).

To obtain the equivalence of (v) and (vi), observe that

$$\{x \in \mathbf{R}^d : f(x) \geq \lambda\} = \bigcap_{\lambda' \in \mathbf{Q}^+ : \lambda' < \lambda} \{x \in \mathbf{R}^d : f(x) > \lambda'\}$$

for $\lambda \in (0, +\infty]$ and

$$\{x \in \mathbf{R}^d : f(x) > \lambda\} = \bigcup_{\lambda' \in \mathbf{Q}^+ : \lambda' > \lambda} \{x \in \mathbf{R}^d : f(x) \geq \lambda'\}$$

$\lambda \in [0, +\infty)$, where $\mathbf{Q}^+ := \mathbf{Q} \cap [0, +\infty]$ are the non-negative rationals. The claim then easily follows from the countable nature of $\mathbf{Q}^+$ (treating the extreme cases $\lambda = 0, +\infty$ separately if necessary). A similar argument lets one deduce (v) or (vi) from (ix).

The equivalence of (v), (vi) with (vii), (viii) comes from the observation that $\{x \in \mathbf{R}^d : f(x) \leq \lambda\}$ is the complement of $\{x \in \mathbf{R}^d : f(x) > \lambda\}$, and $\{x \in \mathbf{R}^d : f(x) < \lambda\}$ is the complement of $\{x \in \mathbf{R}^d : f(x) \geq \lambda\}$. A similar argument shows that (x) and (xi) are equivalent.

By expressing an interval as the intersection of two half-intervals, we see that (ix) follows from (v)-(viii), and so all of (v)-(ix) are now shown to be equivalent.

Clearly (x) implies (vii), and hence (v)-(ix). Conversely, because every open set in $[0, +\infty)$ is the union of countably many open intervals in $[0, +\infty)$, (ix) implies (x).

The only remaining task is to show that (v)-(xi) implies (iv). Let f obey (v)-(xi). For each positive integer n, we let $f_n(x)$ be defined to be the largest integer multiple of 2^{-n} that is less than or equal to $\min(f(x), n)$ when $|x| \leq n$, with $f_n(x) := 0$ for $|x| > n$.

From construction it is easy to see that the $f_n\colon \mathbf{R}^d \to [0,+\infty]$ are increasing and have f as their supremum. Furthermore, each f_n takes on only finitely many values, and for each non-zero value c it attains, the set $f_n^{-1}(c)$ takes the form $f^{-1}(I_c) \cap \{x \in \mathbf{R}^d : |x| \leq n\}$ for some interval or ray I_c, and is thus measurable. As a consequence, f_n is a simple function, and by construction it is bounded and has finite measure support. The claim follows. $\square$

With these equivalent formulations, we can now generate plenty of measurable functions:

Exercise 1.3.3.

(i) Show that every continuous function $f\colon \mathbf{R}^d \to [0,+\infty]$ is measurable.

(ii) Show that every unsigned simple function is measurable.

(iii) Show that the supremum, infimum, limit superior, or limit inferior of unsigned measurable functions is unsigned measurable.

(iv) Show that an unsigned function that is equal almost everywhere to an unsigned measurable function, is itself measurable.

(v) Show that if a sequence f_n of unsigned measurable functions converges pointwise almost everywhere to an unsigned limit f, then f is also measurable.

(vi) If $f\colon \mathbf{R}^d \to [0,+\infty]$ is measurable and $\phi\colon [0,+\infty] \to [0,+\infty]$ is continuous, show that $\phi \circ f\colon \mathbf{R}^d \to [0,+\infty]$ is measurable.

(vii) If f, g are unsigned measurable functions, show that $f+g$ and fg are measurable.

In view of Exercise 1.3.3(iv), one can define the concept of measurability for an unsigned function that is only defined almost everywhere on $\mathbf{R}^d$, rather than everywhere on $\mathbf{R}^d$, by extending that function arbitrarily to the null set where it is currently undefined.

Exercise 1.3.4. Let $f\colon \mathbf{R}^d \to [0,+\infty]$. Show that f is a *bounded* unsigned measurable function if and only if f is the *uniform* limit of *bounded* simple functions.

Exercise 1.3.5. Show that an unsigned function $f\colon \mathbf{R}^d \to [0,+\infty]$ is a simple function if and only if it is measurable and takes on at most finitely many values.

Exercise 1.3.6. Let $f\colon \mathbf{R}^d \to [0,+\infty]$ be an unsigned measurable function. Show that the region $\{(x,t) \in \mathbf{R}^d \times \mathbf{R} : 0 \leq t \leq f(x)\}$ is a measurable subset of $\mathbf{R}^{d+1}$. (There is a converse to this statement, but we will wait until Exercise 1.7.24 to prove it, once we have the Fubini-Tonelli theorem (Corollary 1.7.23) available to us.)

Remark 1.3.10. Lemma 1.3.9 tells us that if $f\colon \mathbf{R}^d \to [0,+\infty]$ is measurable, then $f^{-1}(E)$ is Lebesgue measurable for many classes of sets E. However, we caution that it is *not* necessarily the case that $f^{-1}(E)$ is Lebesgue measurable if E is Lebesgue measurable. To see this, we let C be the Cantor set

$$C := \{\sum_{j=1}^\infty a_j 3^{-j} : a_j \in \{0,2\} \text{ for all } j\}$$

and let $f\colon \mathbf{R} \to [0,+\infty]$ be the function defined by setting

$$f(x) := \sum_{j=1}^\infty 2b_j 3^{-j}$$

whenever $x \in [0,1]$ is not a terminating binary decimal, and so has a unique binary expansion $x = \sum_{j=1}^\infty b_j 2^{-j}$ for some $b_j \in \{0,1\}$, and $f(x) := 0$ otherwise. We thus see that f takes values in C, and is bijective on the set A of non-terminating decimals in $[0,1]$. Using Lemma 1.3.9, it is not difficult to show that f is measurable. On the other hand, by modifying the construction from the previous notes, we can find a subset F of A which is non-measurable. If we set $E := f(F)$, then E is a subset of the null set C and is thus itself a null set; but $f^{-1}(E) = F$ is non-measurable, and so the inverse image of a Lebesgue measurable set by a measurable function need not remain Lebesgue measurable.

However, we will later see that it is still true that $f^{-1}(E)$ is Lebesgue measurable if E has a slightly stronger measurability property than Lebesgue measurability, namely *Borel measurability*; see Exercise 1.4.29(iii).

Now we can define the concept of a complex-valued measurable function. As discussed earlier, it will be convenient to allow for such functions to only be defined *almost everywhere*, rather than *everywhere*, to allow for the possibility that the function becomes singular or otherwise undefined on a null set.

Definition 1.3.11 (Complex measurability)**.** An almost everywhere defined complex-valued function $f: \mathbf{R}^d \to \mathbf{C}$ is *Lebesgue measurable*, or *measurable* for short, if it is the pointwise almost everywhere limit of complex-valued simple functions.

As before, there are several equivalent definitions:

Exercise 1.3.7. Let $f: \mathbf{R}^d \to \mathbf{C}$ be an almost everywhere defined complex-valued function. Then the following are equivalent:

(i) f is measurable.

(ii) f is the pointwise almost everywhere limit of complex-valued simple functions.

(iii) The (magnitudes of the) positive and negative parts of $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are unsigned measurable functions.

(iv) $f^{-1}(U)$ is Lebesgue measurable for every open set $U \subset \mathbf{C}$.

(v) $f^{-1}(K)$ is Lebesgue measurable for every closed set $K \subset \mathbf{C}$.

From the above exercise, we see that the notion of complex-valued measurability and unsigned measurability are compatible when applied to a function that takes values in $[0, +\infty) = [0, +\infty] \cap \mathbf{C}$ everywhere (or almost everywhere).

Exercise 1.3.8.

(i) Show that every continuous function $f: \mathbf{R}^d \to \mathbf{C}$ is measurable.

(ii) Show that a function $f: \mathbf{R}^d \to \mathbf{C}$ is simple if and only if it is measurable and takes on at most finitely many values.

(iii) Show that a complex-valued function that is equal almost everywhere to an measurable function, is itself measurable.

(iv) Show that if a sequence f_n of complex-valued measurable functions converges pointwise almost everywhere to an complex-valued limit f, then f is also measurable.

(v) If $f\colon \mathbf{R}^d \to \mathbf{C}$ is measurable and $\phi\colon \mathbf{C} \to \mathbf{C}$ is continuous, show that $\phi \circ f\colon \mathbf{R}^d \to \mathbf{C}$ is measurable.

(vi) If f, g are measurable functions, show that $f + g$ and fg are measurable.

Exercise 1.3.9. Let $f\colon [a, b] \to \mathbf{R}$ be a Riemann integrable function. Show that if one extends f to all of $\mathbf{R}$ by defining $f(x) = 0$ for $x \notin [a, b]$, then f is measurable.

1.3.3. Unsigned Lebesgue integrals. We are now ready to integrate unsigned measurable functions. We begin with the notion of the lower unsigned Lebesgue integral, which can be defined for arbitrary unsigned functions (not necessarily measurable):

Definition 1.3.12 (Lower unsigned Lebesgue integral)**.** Let $f\colon \mathbf{R}^d \to [0, +\infty]$ be an unsigned function (not necessarily measurable). We define the *lower unsigned Lebesgue integral* $\underline{\int_{\mathbf{R}^d}} f(x)\ dx$ to be the quantity

$$\underline{\int_{\mathbf{R}^d}} f(x)\ dx := \sup_{0 \leq g \leq f;\, g \text{ simple}} \operatorname{Simp} \int_{\mathbf{R}^d} g(x)\ dx$$

where g ranges over all unsigned simple functions $g\colon \mathbf{R}^d \to [0, +\infty]$ that are pointwise bounded by f.

One can also define the *upper unsigned Lebesgue integral*

$$\overline{\int_{\mathbf{R}^d}} f(x)\ dx := \inf_{h \geq f;\, h \text{ simple}} \operatorname{Simp} \int_{\mathbf{R}^d} h(x)\ dx$$

but we will use this integral much more rarely. Note that both integrals take values in $[0, +\infty]$, and that the upper Lebesgue integral is always at least as large as the lower Lebesgue integral.

In the definition of the lower unsigned Lebesgue integral, g is required to be bounded by f pointwise everywhere, but it is easy to see that one could also require g to just be bounded by f pointwise almost everywhere without affecting the value of the integral, since the simple integral is not affected by modifications on sets of measure zero.

The following properties of the lower Lebesgue integral are easy to establish:

Exercise 1.3.10 (Basic properties of the lower Lebesgue integral)**.** Let $f, g \colon \mathbf{R}^d \to [0, +\infty]$ be unsigned functions (not necessarily measurable).

(i) (Compatibility with the simple integral) If f is simple, then $\underline{\int_{\mathbf{R}^d}} f(x)\ dx = \overline{\int_{\mathbf{R}^d}} f(x)\ dx = \operatorname{Simp} \int_{\mathbf{R}^d} f(x)\ dx$.

(ii) (Monotonicity) If $f \leq g$ pointwise almost everywhere, then $\underline{\int_{\mathbf{R}^d}} f(x)\ dx \leq \underline{\int_{\mathbf{R}^d}} g(x)\ dx$ and $\overline{\int_{\mathbf{R}^d}} f(x)\ dx \leq \overline{\int_{\mathbf{R}^d}} g(x)\ dx$.

(iii) (Homogeneity) If $c \in [0, +\infty)$, then $\underline{\int_{\mathbf{R}^d}} cf(x)\ dx = c \underline{\int_{\mathbf{R}^d}} f(x)\ dx$. (The claim unfortunately fails for $c = +\infty$, but this is somewhat tricky to show.)

(iv) (Equivalence) If f, g agree almost everywhere, then $\underline{\int_{\mathbf{R}^d}} f(x)\ dx = \underline{\int_{\mathbf{R}^d}} g(x)\ dx$ and $\overline{\int_{\mathbf{R}^d}} f(x)\ dx = \overline{\int_{\mathbf{R}^d}} g(x)\ dx$.

(v) (Superadditivity) $\underline{\int_{\mathbf{R}^d}} f(x) + g(x)\ dx \geq \underline{\int_{\mathbf{R}^d}} f(x)\ dx + \underline{\int_{\mathbf{R}^d}} g(x)\ dx$.

(vi) (Subadditivity of upper integral) $\overline{\int_{\mathbf{R}^d}} f(x) + g(x)\ dx \leq \overline{\int_{\mathbf{R}^d}} f(x)\ dx + \overline{\int_{\mathbf{R}^d}} g(x)\ dx$

(vii) (Divisibility) For any measurable set E, one has $\underline{\int_{\mathbf{R}^d}} f(x)\ dx = \underline{\int_{\mathbf{R}^d}} f(x) 1_E(x)\ dx + \underline{\int_{\mathbf{R}^d}} f(x) 1_{\mathbf{R}^d \backslash E}(x)\ dx$.

(viii) (Horizontal truncation) As $n \to \infty$, $\underline{\int_{\mathbf{R}^d}} \min(f(x), n)\ dx$ converges to $\underline{\int_{\mathbf{R}^d}} f(x)\ dx$.

(ix) (Vertical truncation) As $n \to \infty$, $\underline{\int_{\mathbf{R}^d}} f(x) 1_{|x| \leq n}\ dx$ converges to $\underline{\int_{\mathbf{R}^d}} f(x)\ dx$. *Hint:* From Exercise 1.2.11 one has $m(E \cap \{x : |x| \leq n\}) \to m(E)$ for any measurable set E.

(x) (Reflection) If $f + g$ is a simple function that is bounded with finite measure support (i.e. it is absolutely integrable), then $\operatorname{Simp} \int_{\mathbf{R}^d} f(x) + g(x)\ dx = \underline{\int_{\mathbf{R}^d}} f(x)\ dx + \overline{\int_{\mathbf{R}^d}} g(x)\ dx$.

Do the horizontal and vertical truncation properties hold if the lower Lebesgue integral is replaced with the upper Lebesgue integral?

Now we restrict attention to measurable functions.

Definition 1.3.13 (Unsigned Lebesgue integral)**.** If $f\colon \mathbf{R}^d \to [0,+\infty]$ is measurable, we define the unsigned Lebesgue integral $\int_{\mathbf{R}^d} f(x)\ dx$ of f to equal the lower unsigned Lebesgue integral $\underline{\int_{\mathbf{R}^d}} f(x)\ dx$. (For non-measurable functions, we leave the unsigned Lebesgue integral undefined.)

One nice feature of measurable functions is that the lower and upper Lebesgue integrals can match, if one also assumes some boundedness:

Exercise 1.3.11. Let $f\colon \mathbf{R}^d \to [0,+\infty]$ be measurable, bounded, and vanishing outside of a set of finite measure. Show that the lower and upper Lebesgue integrals of f agree. (*Hint:* use Exercise 1.3.4.) There is a converse to this statement, but we will defer it to later notes. What happens if f is allowed to be unbounded, or is not supported inside a set of finite measure?

This gives an important corollary:

Corollary 1.3.14 (Finite additivity of the Lebesgue integral)**.** *Let $f, g\colon \mathbf{R}^d \to [0,+\infty]$ be measurable. Then $\int_{\mathbf{R}^d} f(x) + g(x)\ dx = \int_{\mathbf{R}^d} f(x)\ dx + \int_{\mathbf{R}^d} g(x)\ dx$.*

Proof. From the horizontal truncation property and a limiting argument, we may assume that f, g are bounded. From the vertical truncation property and another limiting argument, we may assume that f, g are supported inside a bounded set. From Exercise 1.3.11, we now see that the lower and upper Lebesgue integrals of f, g, and $f + g$ agree. The claim now follows by combining the superadditivity of the lower Lebesgue integral with the subadditivity of the upper Lebesgue integral. □

In the next section we will improve this finite additivity property for the unsigned Lebesgue integral further, to countable additivity; this property is also known as the *monotone convergence theorem* (Theorem 1.4.44).

Exercise 1.3.12 (Upper Lebesgue integral and outer Lebesgue measure)**.** Show that for any set $E \subset \mathbf{R}^d$, $\overline{\int_{\mathbf{R}^d}} 1_E(x)\ dx = m^*(E)$. Conclude that the upper and lower Lebesgue integrals are not necessarily additive if no measurability hypotheses are assumed.

Exercise 1.3.13 (Area interpretation of integral)**.** If $f \colon \mathbf{R}^d \to [0, +\infty]$ is measurable, show that $\int_{\mathbf{R}^d} f(x)\ dx$ is equal to the $d+1$-dimensional Lebesgue measure of the region $\{(x, t) \in \mathbf{R}^d \times \mathbf{R} : 0 \le t \le f(x)\}$. (This can be used as an alternate, and more geometrically intuitive, definition of the unsigned Lebesgue integral; it is a more convenient formulation for establishing the basic convergence theorems, but not quite as convenient for establishing basic properties such as additivity.) (*Hint:* use Exercise 1.2.22.)

Exercise 1.3.14 (Uniqueness of the Lebesgue integral)**.** Show that the Lebesgue integral $f \mapsto \int_{\mathbf{R}^d} f(x)\ dx$ is the only map from measurable unsigned functions $f \colon \mathbf{R}^d \to [0, +\infty]$ to $[0, +\infty]$ that obeys the following properties for measurable $f, g \colon \mathbf{R}^d \to [0, +\infty]$:

(i) (Compatibility with the simple integral) If f is simple, then $\int_{\mathbf{R}^d} f(x)\ dx = \operatorname{Simp} \int_{\mathbf{R}^d} f(x)\ dx$.

(ii) (Finite additivity) $\int_{\mathbf{R}^d} f(x) + g(x)\ dx = \int_{\mathbf{R}^d} f(x)\ dx + \int_{\mathbf{R}^d} g(x)\ dx$.

(iii) (Horizontal truncation) As $n \to \infty$, $\int_{\mathbf{R}^d} \min(f(x), n)\ dx$ converges to $\int_{\mathbf{R}^d} f(x)\ dx$.

(iv) (Vertical truncation) As $n \to \infty$, $\int_{\mathbf{R}^d} f(x) 1_{|x| \le n}\ dx$ converges to $\int_{\mathbf{R}^d} f(x)\ dx$.

Exercise 1.3.15 (Translation invariance)**.** Let $f \colon \mathbf{R}^d \to [0, +\infty]$ be measurable. Show that $\int_{\mathbf{R}^d} f(x+y)\ dx = \int_{\mathbf{R}^d} f(x)\ dx$ for any $y \in \mathbf{R}^d$.

Exercise 1.3.16 (Linear change of variables)**.** Let $f \colon \mathbf{R}^d \to [0, +\infty]$ be measurable, and let $T \colon \mathbf{R}^d \to \mathbf{R}^d$ be an invertible linear transformation. Show that $\int_{\mathbf{R}^d} f(T^{-1}(x))\ dx = |\det T| \int_{\mathbf{R}^d} f(x)\ dx$, or equivalently $\int_{\mathbf{R}^d} f(Tx)\ dx = \frac{1}{|\det T|} \int_{\mathbf{R}^d} f(x)\ dx$.

Exercise 1.3.17 (Compatibility with the Riemann integral)**.** Let $f \colon [a, b] \to [0, +\infty]$ be Riemann integrable. If we extend f to $\mathbf{R}$ by declaring f to equal zero outside of $[a, b]$, show that $\int_{\mathbf{R}} f(x)\ dx = \int_a^b f(x)\ dx$.

We record a basic inequality, known as *Markov's inequality*, that asserts that the Lebesgue integral of an unsigned measurable function controls how often that function can be large:

Lemma 1.3.15 (Markov's inequality)**.** *Let $f \colon \mathbf{R}^d \to [0, +\infty]$ be measurable. Then for any $0 < \lambda < \infty$, one has*

$$m(\{x \in \mathbf{R}^d : f(x) \geq \lambda\}) \leq \frac{1}{\lambda} \int_{\mathbf{R}^d} f(x)\ dx.$$

Proof. We have the trivial pointwise inequality

$$\lambda 1_{\{x \in \mathbf{R}^d : f(x) \geq \lambda\}} \leq f(x).$$

From the definition of the lower Lebesgue integral, we conclude that

$$\lambda m(\{x \in \mathbf{R}^d : f(x) \geq \lambda\}) \leq \int_{\mathbf{R}^d} f(x)\ dx$$

and the claim follows. □

By sending λ to infinity or to zero, we obtain the following important corollary:

Exercise 1.3.18. Let $f \colon \mathbf{R}^d \to [0, +\infty]$ be measurable.

(i) Show that if $\int_{\mathbf{R}^d} f(x)\ dx < \infty$, then f is finite almost everywhere. Give a counterexample to show that the converse statement is false.

(ii) Show that $\int_{\mathbf{R}^d} f(x)\ dx = 0$ if and only if f is zero almost everywhere.

Remark 1.3.16. The use of the integral $\int_{\mathbf{R}^d} f(x)\ dx$ to control the distribution of f is known as the *first moment method*. One can also control this distribution using higher moments such as $\int_{\mathbf{R}^d} |f(x)|^p\ dx$ for various values of p, or exponential moments such as $\int_{\mathbf{R}^d} e^{tf(x)}\ dx$ or the Fourier moments $\int_{\mathbf{R}^d} e^{itf(x)}\ dx$ for various values of t; such moment methods are fundamental to probability theory.

1.3.4. Absolute integrability. Having set out the theory of the unsigned Lebesgue integral, we can now define the absolutely convergent Lebesgue integral.

Definition 1.3.17 (Absolute integrability)**.** An almost everywhere defined measurable function $f\colon \mathbf{R}^d \to \mathbf{C}$ is said to be *absolutely integrable* if the unsigned integral

$$\|f\|_{L^1(\mathbf{R}^d)} := \int_{\mathbf{R}^d} |f(x)|\ dx$$

is finite. We refer to this quantity $\|f\|_{L^1(\mathbf{R}^d)}$ as the $L^1(\mathbf{R}^d)$ *norm* of f, and use $L^1(\mathbf{R}^d)$ or $L^1(\mathbf{R}^d \to \mathbf{C})$ to denote the space of absolutely integrable functions. If f is real-valued and absolutely integrable, we define the Lebesgue integral $\int_{\mathbf{R}^d} f(x)\ dx$ by the formula

$$\int_{\mathbf{R}^d} f(x)\ dx := \int_{\mathbf{R}^d} f_+(x)\ dx - \int_{\mathbf{R}^d} f_-(x)\ dx \tag{1.12}$$

where $f_+ := \max(f, 0)$, $f_- := \max(-f, 0)$ are the magnitudes of the positive and negative components of f (note that the two unsigned integrals on the right-hand side are finite, as f_+, f_- are pointwise dominated by $|f|$). If f is complex-valued and absolutely integrable, we define the Lebesgue integral $\int_{\mathbf{R}^d} f(x)\ dx$ by the formula

$$\int_{\mathbf{R}^d} f(x)\ dx := \int_{\mathbf{R}^d} \operatorname{Re} f(x)\ dx + i \int_{\mathbf{R}^d} \operatorname{Im} f(x)\ dx$$

where the two integrals on the right are interpreted as real-valued absolutely integrable Lebesgue integrals. It is easy to see that the unsigned, real-valued, and complex-valued Lebesgue integrals defined in this manner are compatible on their common domains of definition.

Note from construction that the absolutely integrable Lebesgue integral extends the absolutely integrable simple integral, which is now redundant and will not be needed any further in the sequel.

Remark 1.3.18. One can attempt to define integrals for non-absolutely-integrable functions, analogous to the improper integrals $\int_0^\infty f(x)\ dx := \lim_{R\to\infty} \int_0^R f(x)\ dx$ or the principal value integrals $p.v. \int_{-\infty}^\infty f(x)\ dx := \lim_{R\to\infty} \int_{-R}^R f(x)\ dx$ one sees in the classical one-dimensional Riemannian theory. While one can certainly generate any number of such extensions of the Lebesgue integral concept, such extensions tend to be poorly behaved with respect to various important operations, such as change of variables or exchanging limits and integrals, so it is usually not worthwhile to try to set up a systematic theory for such non-absolutely-integrable integrals that is anywhere near as complete as the absolutely integrable theory, and instead deal with such exotic integrals on an *ad hoc* basis.

From the pointwise triangle inequality $|f(x)+g(x)| \leq |f(x)| + |g(x)|$, we conclude the L^1 triangle inequality

$$\|f+g\|_{L^1(\mathbf{R}^d)} \leq \|f\|_{L^1(\mathbf{R}^d)} + \|g\|_{L^1(\mathbf{R}^d)} \tag{1.13}$$

for any almost everywhere defined measurable $f, g\colon \mathbf{R}^d \to \mathbf{C}$. It is also easy to see that

$$\|cf\|_{L^1(\mathbf{R}^d)} = |c| \|f\|_{L^1(\mathbf{R}^d)}$$

for any complex number c. As such, we see that $L^1(\mathbf{R}^d \to \mathbf{C})$ is a complex vector space. (The L^1 norm is then a *seminorm* on this space; see §1.3 of *An epsilon of room, Vol. I.*) From Exercise 1.3.18 we make the important observation that a function $f \in L^1(\mathbf{R}^d \to \mathbf{C})$ has zero L^1 norm, $\|f\|_{L^1(\mathbf{R}^d)} = 0$, if and only if f is zero almost everywhere.

Given two functions $f, g \in L^1(\mathbf{R}^d \to \mathbf{C})$, we can define the L^1 *distance* $d_{L^1}(f,g)$ between them by the formula

$$d_{L^1}(f,g) := \|f-g\|_{L^1(\mathbf{R}^d)}.$$

Thanks to (1.13), this distance obeys almost all the axioms of a *metric* on $L^1(\mathbf{R}^d)$, with one exception: it is possible for two different functions $f, g \in L^1(\mathbf{R}^d \to \mathbf{C})$ to have a zero L^1 distance, if they agree almost everywhere. As such, d_{L^1} is only a semi-metric (also known as a *pseudo-metric*) rather than a metric. However, if one adopts the convention that any two functions that agree almost everywhere are considered equivalent (or more formally, one works in the quotient space of $L^1(\mathbf{R}^d)$ by the equivalence relation of almost everywhere agreement, which by abuse of notation is also denoted $L^1(\mathbf{R}^d)$), then one recovers a genuine metric. (Later on, we will establish the important fact that this metric makes the (quotient space) $L^1(\mathbf{R}^d)$ a complete metric space, a fact known as the L^1 *Riesz-Fischer theorem*; this completeness is one of the main reasons we spend so much effort setting up Lebesgue integration theory in the first place.)

The linearity properties of the unsigned integral induce analogous linearity properties of the absolutely convergent Lebesgue integral:

Exercise 1.3.19 (Integration is linear)**.** Show that integration $f \mapsto \int_{\mathbf{R}^d} f(x)\ dx$ is a (complex) linear operation from $L^1(\mathbf{R}^d)$ to $\mathbf{C}$. In other words, show that

$$\int_{\mathbf{R}^d} f(x) + g(x)\ dx = \int_{\mathbf{R}^d} f(x)\ dx + \int_{\mathbf{R}^d} g(x)\ dx$$

and

$$\int_{\mathbf{R}^d} cf(x)\ dx = c \int_{\mathbf{R}^d} f(x)\ dx$$

for all absolutely integrable $f, g \colon \mathbf{R}^d \to \mathbf{C}$ and complex numbers c. Also establish the identity

$$\int_{\mathbf{R}^d} \overline{f(x)}\ dx = \overline{\int_{\mathbf{R}^d} f(x)\ dx},$$

which makes integration not just a linear operation, but a **-linear* operation.

Exercise 1.3.20. Show that Exercises 1.3.15, 1.3.16, and 1.3.17 also hold for complex-valued, absolutely integrable functions rather than for unsigned measurable functions.

Exercise 1.3.21 (Absolute summability is a special case of absolute integrability)**.** Let $(c_n)_{n\in\mathbf{Z}}$ be a doubly infinite sequence of complex numbers, and let $f \colon \mathbf{R} \to \mathbf{C}$ be the function

$$f(x) := \sum_{n\in\mathbf{Z}} c_n 1_{[n,n+1)}(x) = c_{\lfloor x \rfloor}$$

where $\lfloor x \rfloor$ is the greatest integer less than x. Show that f is absolutely integrable if and only if the series $\sum_{n\in\mathbf{Z}} c_n$ is absolutely convergent, in which case one has $\int_{\mathbf{R}} f(x)\ dx = \sum_{n\in\mathbf{Z}} c_n$.

We can localise the absolutely convergent integral to any measurable subset E of $\mathbf{R}^d$. Indeed, if $f \colon E \to \mathbf{C}$ is a function, we say that f is measurable (resp. absolutely integrable) if its extension

$\tilde{f}\colon \mathbf{R}^d \to \mathbf{C}$ is measurable (resp. absolutely integrable), where $\tilde{f}(x)$ is defined to equal $f(x)$ when $x \in E$ and zero otherwise, and then we define $\int_E f(x)\ dx := \int_{\mathbf{R}^d} \tilde{f}(x)\ dx$. Thus, for instance, the absolutely integrable analogue of Exercise 1.3.17 tells us that

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

for any Riemann-integrable $f\colon [a,b] \to \mathbf{C}$.

Exercise 1.3.22. If E, F are disjoint measurable subsets of $\mathbf{R}^d$, and $f\colon E \cup F \to \mathbf{C}$ is absolutely integrable, show that

$$\int_E f(x)\ dx = \int_{E \cup F} f(x) 1_E(x)\ dx$$

and

$$\int_E f(x)\ dx + \int_F f(x)\ dx = \int_{E \cup F} f(x)\ dx.$$

We will study the properties of the absolutely convergent Lebesgue integral in more detail in later notes, as a special case of the more general Lebesgue integration theory on abstract measure spaces. For now, we record one very basic inequality:

Lemma 1.3.19 (Triangle inequality). *Let $f \in L^1(\mathbf{R}^d \to \mathbf{C})$. Then*

$$|\int_{\mathbf{R}^d} f(x)\ dx| \leq \int_{\mathbf{R}^d} |f(x)|\ dx.$$

Proof. If f is real-valued, then $|f| = f_+ + f_-$ and the claim is obvious from (1.12). When f is complex-valued, one cannot argue quite so simply; a naive mimicking of the real-valued argument would lose a factor of 2, giving the inferior bound

$$|\int_{\mathbf{R}^d} f(x)\ dx| \leq 2 \int_{\mathbf{R}^d} |f(x)|\ dx.$$

To do better, we exploit the phase rotation invariance properties of the absolute value operation and of the integral, as follows. Note that for any complex number z, one can write $|z|$ as $ze^{i\theta}$ for some real θ. In particular, we have

$$|\int_{\mathbf{R}^d} f(x)\ dx| = e^{i\theta} \int_{\mathbf{R}^d} f(x)\ dx = \int_{\mathbf{R}^d} e^{i\theta} f(x)\ dx$$

for some real θ. Taking real parts of both sides, we obtain

$$|\int_{\mathbf{R}^d} f(x)\ dx| = \int_{\mathbf{R}^d} \mathrm{Re}(e^{i\theta} f(x))\ dx.$$

Since $\mathrm{Re}(e^{i\theta} f(x)) \leq |e^{i\theta} f(x)| = |f(x)|$, we obtain the claim. □

1.3.5. Littlewood's three principles. *Littlewood's three principles* are informal heuristics that convey much of the basic intuition behind the measure theory of Lebesgue. Briefly, the three principles are as follows:

(i) Every (measurable) set is nearly a finite sum of intervals;

(ii) Every (absolutely integrable) function is nearly continuous; and

(iii) Every (pointwise) convergent sequence of functions is nearly uniformly convergent.

Various manifestations of the first principle were given in Exercise 1.2.7 and Exercise 1.2.16. Now we turn to the second principle. Define a *step function* to be a finite linear combination of indicator functions 1_B of boxes B.

Theorem 1.3.20 (Approximation of L^1 functions)**.** *Let $f \in L^1(\mathbf{R}^d)$ and $\varepsilon > 0$.*

(i) *There exists an absolutely integrable simple function g such that $\|f - g\|_{L^1(\mathbf{R}^d)} \leq \varepsilon$.*

(ii) *There exists a step function g such that $\|f - g\|_{L^1(\mathbf{R}^d)} \leq \varepsilon$.*

(iii) *There exists a continuous, compactly supported g such that $\|f - g\|_{L^1(\mathbf{R}^d)} \leq \varepsilon$.*

To put things another way, the absolutely integrable simple functions, the step functions, and the continuous, compactly supported functions are all dense subsets of $L^1(\mathbf{R}^d)$ with respect to the $L^1(\mathbf{R}^d)$ (semi-)metric. In §1.13 of *An epsilon of room, Vol. I* it is shown that a similar statement holds if one replaces continuous, compactly supported functions with *smooth*, compactly supported functions, also known as *test functions*; this is an important fact for the theory of *distributions*.

Proof. We begin with part (i). When f is unsigned, we see from the definition of the lower Lebesgue integral that there exists an unsigned simple function g such that $g \leq f$ (so, in particular, g is absolutely integrable) and

$$\int_{\mathbf{R}^d} g(x)\ dx \geq \int_{\mathbf{R}^d} f(x)\ dx - \varepsilon,$$

which by linearity implies that $\|f-g\|_{L^1(\mathbf{R}^d)} \leq \varepsilon$. This gives (i) when f is unsigned. The case when f is real-valued then follows by splitting f into positive and negative parts (and adjusting ε as necessary), and the case when f is complex-valued then follows by splitting f into real and imaginary parts (and adjusting ε yet again).

To establish part (ii), we see from (i) and the triangle inequality in L^1 that it suffices to show this when f is an absolutely integrable simple function. By linearity (and more applications of the triangle inequality), it then suffices to show this when $f = 1_E$ is the indicator function of a measurable set $E \subset \mathbf{R}^d$ of finite measure. But then, by Exercise (1.2.16), such a set can be approximated (up to an error of measure at most ε) by an elementary set, and the claim follows.

To establish part (iii), we see from (ii) and the argument from the preceding paragraph that it suffices to show this when $f = 1_E$ is the indicator function of a box. But one can then establish the claim by direct construction. Indeed, if one makes a slightly larger box F that contains the closure of E in its interior, but has a volume at most ε more than that of E, then one can directly construct a piecewise linear continuous function g supported on F that equals 1 on E (e.g. one can set $g(x) = \max(1 - R\operatorname{dist}(x, E), 0)$ for some sufficiently large R; one may also invoke *Urysohn's lemma*, see §1.10 of *An epsilon of room, Vol. I*). It is then clear from construction that $\|f - g\|_{L^1(\mathbf{R}^d)} \leq \varepsilon$ as required. □

This is not the only way to make Littlewood's second principle manifest; we return to this point shortly. For now, we turn to Littlewood's third principle. We recall three basic ways in which a sequence $f_n \colon \mathbf{R}^d \to \mathbf{C}$ of functions can converge to a limit $f \colon \mathbf{R}^d \to \mathbf{C}$:

(i) (Pointwise convergence) $f_n(x) \to f(x)$ for every $x \in \mathbf{R}^d$.

(ii) (Pointwise almost everywhere convergence) $f_n(x) \to f(x)$ for almost every $x \in \mathbf{R}^d$.

(iii) (Uniform convergence) For every $\varepsilon > 0$, there exists N such that $|f_n(x) - f(x)| \leq \varepsilon$ for all $n \geq N$ and all $x \in \mathbf{R}^d$.

Uniform convergence implies pointwise convergence, which in turn implies pointwise almost everywhere convergence.

We now add a fourth mode of convergence, that is weaker than uniform convergence but stronger than pointwise convergence:

Definition 1.3.21 (Locally uniform convergence)**.** A sequence of functions $f_n\colon \mathbf{R}^d \to \mathbf{C}$ converges *locally uniformly* to a limit $f\colon \mathbf{R}^d \to \mathbf{C}$ if, for every bounded subset E of $\mathbf{R}^d$, f_n converges uniformly to f on E. In other words, for every bounded $E \subset \mathbf{R}^d$ and every $\varepsilon > 0$, there exists $N > 0$ such that $|f_n(x) - f(x)| \leq \varepsilon$ for all $n \geq N$ and $x \in E$.

Remark 1.3.22. At least as far as $\mathbf{R}^d$ is concerned, an equivalent definition of local uniform convergence is: f_n converges locally uniformly to f if, for every point $x_0 \in \mathbf{R}^d$, there exists an open neighbourhood U of x_0 such that f_n converges uniformly to f on U. The equivalence of the two definitions is immediate from the *Heine-Borel theorem*. More generally, the adverb "locally" in mathematics is usually used in this fashion; a propery P is said to hold *locally* on some domain X if, for every point x_0 in that domain, there is an open neighbourhood of x_0 in X on which P holds.

One should caution, though, that on domains on which the Heine-Borel theorem does not hold, the bounded-set notion of local uniform convergence is not equivalent to the open-set notion of local uniform convergence (though, for *locally compact spaces*, one can recover equivalence of one replaces "bounded" by "compact").

Example 1.3.23. The functions $x \mapsto x/n$ on $\mathbf{R}$ for $n = 1, 2, \ldots$ converge locally uniformly (and hence pointwise) to zero on $\mathbf{R}$, but do not converge uniformly.

Example 1.3.24. The partial sums $\sum_{n=0}^{N} \frac{x^n}{n!}$ of the Taylor series $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ converges to e^x locally uniformly (and hence pointwise) on $\mathbf{R}$, but not uniformly.

Example 1.3.25. The functions $f_n(x) := \frac{1}{nx} 1_{x>0}$ for $n = 1, 2, \ldots$ (with the convention that $f_n(0) = 0$) converge pointwise everywhere to zero, but do not converge locally uniformly.

From the preceding example, we see that pointwise convergence (either everywhere or almost everywhere) is a weaker concept than local uniform convergence. Nevertheless, a remarkable theorem of Egorov, which demonstrates Littlewood's third principle, asserts that one can recover local uniform convergence as long as one is willing to delete a set of small measure:

Theorem 1.3.26 (Egorov's theorem)**.** *Let $f_n \colon \mathbf{R}^d \to \mathbf{C}$ be a sequence of measurable functions that converge pointwise almost everywhere to another function $f \colon \mathbf{R}^d \to \mathbf{C}$, and let $\varepsilon > 0$. Then there exists a Lebesgue measurable set A of measure at most ε, such that f_n converges locally uniformly to f outside of A.*

Note that Example 1.3.25 demonstrates that the exceptional set A in Egorov's theorem cannot be taken to have zero measure, at least if one uses the bounded-set definition of local uniform convergence from Definition 1.3.21. (If one instead takes the "open neighbourhood" definition, then the sequence in Example 1.3.25 does converge locally uniformly on $\mathbf{R} \backslash \{0\}$ in the open neighbourhood sense, even if it does not do so in the bounded-set sense. On a domain such as $\mathbf{R}^d \backslash A$, bounded-set locally uniform convergence implies open-neighbourhood locally uniform convergence, but not conversely, so for the purposes of applying Egorov's theorem, the distinction is not too important since one local uniform convergence in both senses.)

Proof. By modifying f_n and f on a set of measure zero (that can be absorbed into A at the end of the argument) we may assume that f_n converges pointwise everywhere to f, thus for every $x \in \mathbf{R}^d$ and $m > 0$ there exists $N \geq 0$ such that $|f_n(x) - f(x)| \leq 1/m$ for all $n \geq N$. We can rewrite this fact set-theoretically as

$$\bigcap_{N=0}^{\infty} E_{N,m} = \emptyset$$

for each m, where

$$E_{N,m} := \{x \in \mathbf{R}^d : |f_n(x) - f(x)| > 1/m \text{ for some } n \geq N\}.$$

It is clear that the $E_{N,m}$ are Lebesgue measurable, and are decreasing in N. Applying downward monotone convergence (Exercise 1.2.11(ii)) we conclude that, for any radius $R > 0$, one has

$$\lim_{N\to\infty} m(E_{N,m} \cap B(0,R)) = 0.$$

(The restriction to the ball $B(0,R)$ is necessary, because the downward monotone convergence property only works when the sets involved have finite measure.) In particular, for any $m \geq 1$, we can find N_m such that

$$m(E_{N,m} \cap B(0,m)) \leq \frac{\varepsilon}{2^m}$$

for all $N \geq N_m$.

Now let $A := \bigcup_{m=1}^\infty E_{N_m,m} \cap B(0,m)$. Then A is Lebesgue measurable, and by countable subadditivity, $m(A) \leq \varepsilon$. By construction, we have

$$|f_n(x) - f(x)| \leq 1/m$$

whenever $m \geq 1$, $x \in \mathbf{R}^d \backslash A$, $|x| \leq m$, and $n \geq N_m$. In particular, we see for any ball $B(0, m_0)$ with an integer radius, f_n converges uniformly to f on $B(0,m_0)\backslash A$. Since every bounded set is contained in such a ball, the claim follows. □

Remark 1.3.27. Unfortunately, one cannot in general upgrade local uniform convergence to uniform convergence in Egorov's theorem. A basic example here is the *moving bump* example $f_n := 1_{[n,n+1]}$ on $\mathbf{R}$, which "escapes to horizontal infinity". This sequence converges pointwise (and locally uniformly) to the zero function $f \equiv 0$. However, for any $0 < \varepsilon < 1$ and any n, we have $|f_n(x) - f(x)| > \varepsilon$ on a set of measure 1, namely on the interval $[n, n+1]$. Thus, if one wanted f_n to converge uniformly to f outside of a set A, then that set A has to contain a set of measure 1. In fact, it must contain the intervals $[n, n+1]$ for all sufficiently large n and must therefore have infinite measure.

However, if all the f_n and f were supported on a *fixed* set E of finite measure (e.g. on a ball $B(0,R)$), then the above "escape to horizontal infinity" cannot occur, it is easy to see from the above argument that one can recover uniform convergence (and not just locally uniform convergence) outside of a set of arbitrarily small measure.

We now use Theorem 1.3.20 to give another version of Littlewood's second principle, known as *Lusin's theorem*:

Theorem 1.3.28 (Lusin's theorem)**.** *Let $f\colon \mathbf{R}^d \to \mathbf{C}$ be absolutely integrable, and let $\varepsilon > 0$. Then there exists a Lebesgue measurable set $E \subset \mathbf{R}^d$ of measure at most ε such that the restriction of f to the complementary set $\mathbf{R}^d \backslash E$ is continuous on that set.*

A word of caution: this theorem does *not* imply that the unrestricted function f is continuous on $\mathbf{R}^d \backslash E$. For instance, the absolutely integrable function $1_{\mathbf{Q}}\colon \mathbf{R} \to \mathbf{C}$ is nowhere continuous, so is certainly not continuous on $\mathbf{R}\backslash E$ for any E of finite measure; but on the other hand, if one deletes the measure zero set $E := \mathbf{Q}$ from the reals, then the restriction of f to $\mathbf{R}\backslash E$ is identically zero and thus continuous.

Proof. By Theorem 1.3.20, for any $n \geq 1$ one can find a continuous, compactly supported function f_n such that $\|f - f_n\|_{L^1(\mathbf{R}^d)} \leq \varepsilon/4^n$ (say). By Markov's inequality (Lemma 1.3.15), that implies that $|f(x) - f_n(x)| \leq 1/2^{n-1}$ for all x outside of a Lebesgue measurable set E_n of measure at most $\varepsilon/2^{n+1}$. Letting $E := \bigcup_{n=1}^\infty E_n$, we conclude that E is Lebesgue measurable with measure at most $\varepsilon/2$, and f_n converges uniformly to f outside of E. But the uniform limit of continuous functions is continuous, and the same is true for local uniform limits (because continuity is itself a local property). We conclude that the restriction f to $\mathbf{R}^d \backslash E$ is continuous, as required. □

Exercise 1.3.23. Show that the hypothesis that f is absolutely integrable in Lusin's theorem can be relaxed to being locally absolutely integrable (i.e. absolutely integrable on every bounded set), and then relaxed further to that of being measurable (but still finite everywhere or almost everywhere). (To achieve the latter goal, one can replace f locally with a horizontal truncation $f 1_{|f| \leq n}$; alternatively, one can replace f with a bounded variant, such as $\frac{f}{(1+|f|^2)^{1/2}}$.)

Exercise 1.3.24. Show that a function $f\colon \mathbf{R}^d \to \mathbf{C}$ is measurable if and only if it is the pointwise almost everywhere limit of continuous functions $f_n\colon \mathbf{R}^d \to \mathbf{C}$. (*Hint:* if $f\colon \mathbf{R}^d \to \mathbf{C}$ is measurable and $n \geq 1$, show that there exists a continuous function $f_n\colon \mathbf{R}^d \to \mathbf{C}$ for which the set $\{x \in B(0,n) : |f(x) - f_n(x)| \geq 1/n\}$ has measure at most $\frac{1}{2^n}$. You may find Exercise 1.3.25 below to be useful for this.) Use this (and Egorov's theorem, Theorem 1.3.26) to give an alternate proof of Lusin's theorem for arbitrary measurable functions.

Remark 1.3.29. This is a trivial but important remark: when dealing with unsigned measurable functions such as $f \colon \mathbf{R}^d \to [0,+\infty]$, then Lusin's theorem does not apply directly because f could be infinite on a set of positive measure, which is clearly in contradiction with the conclusion of Lusin's theorem (unless one allows the continuous function to also take values in the extended non-negative reals $[0,+\infty]$ with the extended topology). However, if one knows already that f is almost everywhere finite (which is for instance the case when f is absolutely integrable), then Lusin's theorem applies (since one can simply zero out f on the null set where it is infinite, and add that null set to the exceptional set of Lusin's theorem).

Remark 1.3.30. By combining Lusin's theorem with inner regularity (Exercise 1.2.15) and the *Tietze extension theorem* (see §1.10 of *An epsilon of room, Vol. I*), one can conclude that every measurable function $f \colon \mathbf{R}^d \to \mathbf{C}$ agrees (outside of a set of arbitrarily small measure) with a continuous function $g \colon \mathbf{R}^d \to \mathbf{C}$.

Exercise 1.3.25 (Littlewood-like principles)**.** The following facts are not, strictly speaking, instances of any of Littlewood's three principles, but are in a similar spirit.

(i) (Absolutely integrable functions almost have bounded support) Let $f \colon \mathbf{R}^d \to \mathbf{C}$ be an absolutely integrable function, and let $\varepsilon > 0$. Show that there exists a ball $B(0,R)$ outside of which f has an L^1 norm of at most ε, or in other words that $\int_{\mathbf{R}^d \backslash B(0,R)} |f(x)|\, dx \leq \varepsilon$.

(ii) (Measurable functions are almost locally bounded) Let $f \colon \mathbf{R}^d \to \mathbf{C}$ be a measurable function supported on a set of finite measure, and let $\varepsilon > 0$. Show that there exists a measurable set $E \subset \mathbf{R}^d$ of measure at most ε outside of which f is locally bounded, or in other words that for every $R > 0$ there exists $M < \infty$ such that $|f(x)| \leq M$ for all $x \in B(0,R) \backslash E$.

As with Remark 1.3.29, it is important in the second part of the exercise that f is known to be finite everywhere (or at least almost everywhere); the result would of course fail if f was, say, unsigned but took the value $+\infty$ on a set of positive measure.

1.4. Abstract measure spaces

Thus far, we have only focused on measure and integration theory in the context of Euclidean spaces $\mathbf{R}^d$. Now, we will work in a more abstract and general setting, in which the Euclidean space $\mathbf{R}^d$ is replaced by a more general space X.

It turns out that in order to properly define measure and integration on a general space X, it is not enough to just specify the set X. One also needs to specify two additional pieces of data:

(i) A collection $\mathcal{B}$ of subsets of X that one is allowed to measure; and

(ii) The measure $\mu(E) \in [0, +\infty]$ one assigns to each measurable set $E \in \mathcal{B}$.

For instance, Lebesgue measure theory covers the case when X is a Euclidean space $\mathbf{R}^d$, $\mathcal{B}$ is the collection $\mathcal{B} = \mathcal{L}[\mathbf{R}^d]$ of all Lebesgue measurable subsets of $\mathbf{R}^d$, and $\mu(E)$ is the Lebesgue measure $\mu(E) = m(E)$ of E.

The collection $\mathcal{B}$ has to obey a number of axioms (e.g. being closed with respect to countable unions) that make it a *σ-algebra*, which is a stronger variant of the more well-known concept of a *boolean algebra*. Similarly, the measure μ has to obey a number of axioms (most notably, a countable additivity axiom) in order to obtain a measure and integration theory comparable to the Lebesgue theory on Euclidean spaces. When all these axioms are satisfied, the triple $(X, \mathcal{B}, \mu)$ is known as a *measure space*. These play much the same role in abstract measure theory that metric spaces or topological spaces play in abstract point-set topology, or that vector spaces play in abstract linear algebra.

On any measure space, one can set up the unsigned and absolutely convergent integrals in almost exactly the same way as was done in the previous notes for the Lebesgue integral on Euclidean spaces, although the approximation theorems are largely unavailable at this level of generality due to the lack of such concepts as "elementary set" or "continuous function" for an abstract measure space. On the other hand, one *does* have the fundamental convergence theorems for the subject, namely Fatou's lemma, the monotone convergence theorem and the dominated convergence theorem, and we present these results here.

One question that will not be addressed much in this section is how one actually *constructs* interesting examples of measures. We will return to this issue in Section 1.7 (although one of the most powerful tools for such constructions, namely the *Riesz representation theorem*, will not be covered here, but instead in §1.10 of *An epsilon of room, Vol. I*).

1.4.1. Boolean algebras.

We begin by recalling the concept of a *Boolean algebra*.

Definition 1.4.1 (Boolean algebras)**.** Let X be a set. A (concrete) *Boolean algebra* on X is a collection $\mathcal{B}$ of X which obeys the following properties:

(i) (Empty set) $\emptyset \in \mathcal{B}$.

(ii) (Complement) If $E \in \mathcal{B}$, then the complement $E^c := X \backslash E$ also lies in $\mathcal{B}$.

(iii) (Finite unions) If $E, F \in \mathcal{B}$, then $E \cup F \in \mathcal{B}$.

We sometimes say that E is *$\mathcal{B}$-measurable*, or *measurable with respect to $\mathcal{B}$*, if $E \in \mathcal{B}$.

Given two Boolean algebras $\mathcal{B}, \mathcal{B}'$ on X, we say that $\mathcal{B}'$ is *finer than*, *a sub-algebra of*, or *a refinement of* $\mathcal{B}$, or that $\mathcal{B}$ is *coarser than* or *a coarsening of* $\mathcal{B}'$, if $\mathcal{B} \subset \mathcal{B}'$.

We have chosen a "minimalist" definition of a Boolean algebra, in which one is only assumed to be closed under two of the basic Boolean operations, namely complement and finite union. However, by using the laws of Boolean algebra (such as de Morgan's laws), it is easy to see that a Boolean algebra is also closed under other

Boolean algebra operations such as intersection $E \cap F$, set differerence $E \backslash F$, and symmetric difference $E \Delta F$. So one could have placed these additional closure properties inside the definition of a Boolean algebra without any loss of generality. However, when we are *verifying* that a given collection $\mathcal{B}$ of sets is indeed a Boolean algebra, it is convenient to have as minimal a set of axioms as possible.

Remark 1.4.2. One can also consider *abstract Boolean algebras* $\mathcal{B}$, which do not necessarily live in an ambient domain X, but for which one has a collection of abstract Boolean operations such as meet $\wedge$ and join $\vee$ instead of the concrete operations of intersection $\cap$ and union $\cup$. We will not take this abstract perspective here, but see §2.3 of *An epsilon of room, Vol. I* for some further discussion of the relationship between concrete and abstract Boolean algebras, which is codified by *Stone's theorem*.

Example 1.4.3 (Trivial and discrete algebra)**.** Given any set X, the coarsest Boolean algebra is the *trivial algebra* $\{\emptyset, X\}$, in which the only measurable sets are the empty set and the whole set. The finest Boolean algebra is the *discrete algebra* $2^X := \{E : E \subset X\}$, in which *every* set is measurable. All other Boolean algebras are intermediate between these two extremes: finer than the trivial algebra, but coarser than the discrete one.

Exercise 1.4.1 (Elementary algebra)**.** Let $\overline{\mathcal{E}[\mathbf{R}^d]}$ be the collection of those sets $E \subset \mathbf{R}^d$ that are either elementary sets, or co-elementary sets (i.e. the complement of an elementary set). Show that $\overline{\mathcal{E}[\mathbf{R}^d]}$ is a Boolean algebra. We will call this algebra the *elementary Boolean algebra* of $\mathbf{R}^d$.

Example 1.4.4 (Jordan algebra)**.** Let $\overline{\mathcal{J}[\mathbf{R}^d]}$ be the collection of subsets of $\mathbf{R}^d$ that are either Jordan measurable or co-Jordan measurable (i.e. the complement of a Jordan measurable set). Then $\overline{\mathcal{J}[\mathbf{R}^d]}$ is a Boolean algebra that is finer than the elementary algebra. We refer to this algebra as the *Jordan algebra* on $\mathbf{R}^d$ (but caution that there is a completely different concept of a Jordan algebra in abstract algebra.)

Example 1.4.5 (Lebesgue algebra)**.** Let $\mathcal{L}[\mathbf{R}^d]$ be the collection of Lebesgue measurable subsets of $\mathbf{R}^d$. Then $\mathcal{L}[\mathbf{R}^d]$ is a Boolean algebra that is finer than the Jordan algebra; we refer to this as the *Lebesgue algebra* on $\mathbf{R}^d$.

Example 1.4.6 (Null algebra)**.** Let $\mathcal{N}(\mathbf{R}^d)$ be the collection of subsets of $\mathbf{R}^d$ that are either Lebesgue null sets or Lebesgue co-null sets (the complement of null sets). Then $\mathcal{N}(\mathbf{R}^d)$ is a Boolean algebra that is coarser than the Lebesgue algebra; we refer to it as the *null algebra* on $\mathbf{R}^d$.

Exercise 1.4.2 (Restriction)**.** Let $\mathcal{B}$ be a Boolean algebra on a set X, and let Y be a subset of X (not necessarily $\mathcal{B}$-measurable). Show that the *restriction* $\mathcal{B}\downharpoonright_Y := \{E \cap Y : E \in \mathcal{B}\}$ of $\mathcal{B}$ to Y is a Boolean algebra on Y. If Y is $\mathcal{B}$-measurable, show that

$$\mathcal{B}\downharpoonright_Y = \mathcal{B} \cap 2^Y = \{E \subset Y : E \in \mathcal{B}\}.$$

Example 1.4.7 (Atomic algebra)**.** Let X be partitioned into a union $X = \bigcup_{\alpha \in I} A_\alpha$ of disjoint sets A_α, which we refer to as *atoms*. Then this partition generates a Boolean algebra $\mathcal{A}((A_\alpha)_{\alpha \in I})$, defined as the collection of all the sets E of the form $E = \bigcup_{\alpha \in J} A_\alpha$ for some $J \subset I$, i.e. $\mathcal{A}((A_\alpha)_{\alpha \in I})$ is the collection of all sets that can be represented as the union of one or more atoms. This is easily verified to be a Boolean algebra, and we refer to it as the *atomic algebra* with atoms $(A_\alpha)_{\alpha \in I}$. The trivial algebra corresponds to the trivial partition $X = X$ into a single atom; at the other extreme, the discrete algebra corresponds to the discrete partition $X = \bigcup_{x \in X} \{x\}$ into singleton atoms. More generally, note that finer (resp. coarser) partitions lead to finer (resp. coarser) atomic algebra. In this definition, we permit some of the atoms in the partition to be empty; but it is clear that empty atoms have no impact on the final atomic algebra, and so without loss of generality one can delete all empty atoms and assume that all atoms are non-empty if one wishes.

Example 1.4.8 (Dyadic algebras)**.** Let n be an integer. The *dyadic algebra* $\mathcal{D}_n(\mathbf{R}^d)$ at scale 2^{-n} in $\mathbf{R}^d$ is defined to be the atomic algebra generated by the half-open dyadic cubes

$$\left[\frac{i_1}{2^n}, \frac{i_1+1}{2^n}\right) \times \ldots \times \left[\frac{i_d}{2^n}, \frac{i_d+1}{2^n}\right)$$

of length 2^{-n} (see Exercise 1.1.14). These are Boolean algebras which are increasing in n: $\mathcal{D}_{n+1} \supset \mathcal{D}_n$. Draw a diagram to indicate how these algebras sit in relation to the elementary, Jordan, and Lebesgue, null, discrete, and trivial algebras.

Remark 1.4.9. The dyadic algebras are analogous to the finite resolution one has on modern computer monitors, which subdivide space into square pixels. A low resolution monitor (in which each pixel has a large size) can only resolve a very small set of "blocky" images, as opposed to the larger class of images that can be resolved by a finer resolution monitor.

Exercise 1.4.3. Show that the non-empty atoms of an atomic algebra are determined up to relabeling. More precisely, show that if $X = \bigcup_{\alpha \in I} A_\alpha = \bigcup_{\alpha' \in I'} A'_{\alpha'}$ are two partitions of X into non-empty atoms A_α, $A'_{\alpha'}$, then $\mathcal{A}((A_\alpha)_{\alpha \in I}) = \mathcal{A}((A'_{\alpha'})_{\alpha' \in I'})$ if and only if exists a bijection $\phi \colon I \to I'$ such that $A'_{\phi(\alpha)} = A_\alpha$ for all $\alpha \in I$.

While many Boolean algebras are atomic, many are not, as the following two exercises indicate.

Exercise 1.4.4. Show that every finite Boolean algebra is an atomic algebra. (A Boolean algebra $\mathcal{B}$ is *finite* if its cardinality is finite, i.e. there are only finitely many measurable sets.) Conclude that every finite Boolean algebra has a cardinality of the form 2^n for some natural number n. From this exercise and Exercise 1.4.3 we see that there is a one-to-one correspondence between finite Boolean algebras on X and finite partitions of X into non-empty sets (up to relabeling).

Exercise 1.4.5. Show that the elementary, Jordan, Lebesgue, and null algebras are *not* atomic algebras. (*Hint:* argue by contradiction. If these algebras were atomic, what must the atoms be?)

Now we describe some further ways to generate Boolean algebras.

Exercise 1.4.6 (Intersection of algebras)**.** Let $(\mathcal{B}_\alpha)_{\alpha \in I}$ be a family of Boolean algebras on a set X, indexed by a (possibly infinite or uncountable) label set I. Show that the intersection $\bigwedge_{\alpha \in I} \mathcal{B}_\alpha := \bigcap_{\alpha \in I} \mathcal{B}_\alpha$ of these algebras is still a Boolean algebra, and is the finest

Boolean algebra that is coarser than all of the $\mathcal{B}_\alpha$. (If I is empty, we adopt the convention that $\bigwedge_{\alpha \in I} \mathcal{B}_\alpha$ is the discrete algebra.)

Definition 1.4.10 (Generation of algebras)**.** Let $\mathcal{F}$ be any family of sets in X. We define $\langle \mathcal{F} \rangle_{\rm bool}$ to be the intersection of all the Boolean algebras that contain $\mathcal{F}$, which is again a Boolean algebra by Exercise 1.4.6. Equivalently, $\langle \mathcal{F} \rangle_{\rm bool}$ is the coarsest Boolean algebra that contains $\mathcal{F}$. We say that $\langle \mathcal{F} \rangle_{\rm bool}$ is the Boolean algebra *generated* by $\mathcal{F}$.

Example 1.4.11. $\mathcal{F}$ is a Boolean algebra if and only if $\langle \mathcal{F} \rangle_{\rm bool} = \mathcal{F}$; thus each Boolean algebra is generated by itself.

Exercise 1.4.7. Show that the elementary algebra $\mathcal{E}(\mathbf{R}^d)$ is generated by the collection of boxes in $\mathbf{R}^d$.

Exercise 1.4.8. Let n be a natural number. Show that if $\mathcal{F}$ is a finite collection of n sets, then $\langle \mathcal{F} \rangle_{\rm bool}$ is a finite Boolean algebra of cardinality at most 2^{2^n} (in particular, finite sets generate finite algebras). Give an example to show that this bound is best possible. (*Hint:* for the latter, it may be convenient to use a discrete ambient space such as the discrete cube $X = \{0,1\}^n$.)

The Boolean algebra $\langle \mathcal{F} \rangle_{\rm bool}$ can be described explicitly in terms of $\mathcal{F}$ as follows:

Exercise 1.4.9 (Recursive description of a generated Boolean algebra)**.** Let $\mathcal{F}$ be a collection of sets in a set X. Define the sets $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \ldots$ recursively as follows:

(i) $\mathcal{F}_0 := \mathcal{F}$.

(ii) For each $n \geq 1$, we define $\mathcal{F}_n$ to be the collection of all sets that either the union of a finite number of sets in $\mathcal{F}_{n-1}$ (including the empty union $\emptyset$), or the complement of such a union.

Show that $\langle \mathcal{F} \rangle_{\rm bool} = \bigcup_{n=0}^\infty \mathcal{F}_n$.

1.4.2. σ-algebras and measurable spaces. In order to obtain a measure and integration theory that can cope well with limits, the finite union axiom of a Boolean algebra is insufficient, and must be improved to a countable union axiom:

Definition 1.4.12 (Sigma algebras)**.** Let X be a set. A *σ-algebra* on X is a collection $\mathcal{B}$ of X which obeys the following properties:

(i) (Empty set) $\emptyset \in \mathcal{B}$.

(ii) (Complement) If $E \in \mathcal{B}$, then the complement $E^c := X \backslash E$ also lies in $\mathcal{B}$.

(iii) (Countable unions) If $E_1, E_2, \ldots \in \mathcal{B}$, then $\bigcup_{n=1}^\infty E_n \in \mathcal{B}$.

We refer to the pair $(X, \mathcal{B})$ of a set X together with a σ-algebra on that set as a *measurable space*.

Remark 1.4.13. The prefix σ usually denotes "countable union". Other instances of this prefix include a *σ-compact topological space* (a countable union of compact sets), a *σ-finite measure space* (a countable union of sets of finite measure), or *F_σ set* (a countable union of closed sets) for other instances of this prefix.

From de Morgan's law (which is just as valid for infinite unions and intersections as it is for finite ones), we see that σ-algebras are closed under countable intersections as well as countable unions.

By padding a finite union into a countable union by using the empty set, we see that every σ-algebra is automatically a Boolean algebra. Thus, we automatically inherit the notion of being measurable with respect to a σ-algebra, or of one σ-algebra being coarser or finer than another.

Exercise 1.4.10. Show that all atomic algebras are σ-algebras. In particular, the discrete algebra and trivial algebra are σ-algebras, as are the finite algebras and the dyadic algebras on Euclidean spaces.

Exercise 1.4.11. Show that the Lebesgue and null algebras are σ-algebras, but the elementary and Jordan algebras are not.

Exercise 1.4.12. Show that any restriction $\mathcal{B}\downharpoonright_Y$ of a σ-algebra $\mathcal{B}$ to a subspace Y of X (as defined in Exercise 1.4.2) is again a σ-algebra on the subspace Y.

There is an exact analogue of Exercise 1.4.6:

Exercise 1.4.13 (Intersection of σ-algebras)**.** Show that the intersection $\bigwedge_{\alpha \in I} \mathcal{B}_\alpha := \bigcap_{\alpha \in I} \mathcal{B}_\alpha$ of an arbitrary (and possibly infinite or uncountable) number of σ-algebras $\mathcal{B}_\alpha$ is again a σ-algebra, and is the finest σ-algebra that is coarser than all of the $\mathcal{B}_\alpha$.

Similarly, we have a notion of generation:

Definition 1.4.14 (Generation of σ-algebras)**.** Let $\mathcal{F}$ be any family of sets in X. We define $\langle \mathcal{F} \rangle$ to be the intersection of all the σ-algebras that contain $\mathcal{F}$, which is again a σ-algebra by Exercise 1.4.13. Equivalently, $\langle \mathcal{F} \rangle$ is the coarsest σ-algebra that contains $\mathcal{F}$. We say that $\langle \mathcal{F} \rangle$ is the σ-algebra *generated* by $\mathcal{F}$.

Since every σ-algebra is a Boolean algebra, we have the trivial inclusion

$$\langle \mathcal{F} \rangle_{\mathrm{bool}} \subset \langle \mathcal{F} \rangle.$$

However, equality need not hold; it only holds if and only if $\langle \mathcal{F} \rangle_{\mathrm{bool}}$ is a σ-algebra. For instance, if $\mathcal{F}$ is the collection of all boxes in $\mathbf{R}^d$, then $\langle \mathcal{F} \rangle_{\mathrm{bool}}$ is the elementary algebra (Exercise 1.4.7), but $\langle \mathcal{F} \rangle$ cannot equal this algebra, as it is not a σ-algebra.

Remark 1.4.15. From the definitions, it is clear that we have the following principle, somewhat analogous to the principle of mathematical induction: if $\mathcal{F}$ is a family of sets in X, and $P(E)$ is a property of sets $E \subset X$ which obeys the following axioms:

(i) $P(\emptyset)$ is true.
(ii) $P(E)$ is true for all $E \in \mathcal{F}$.
(iii) If $P(E)$ is true for some $E \subset X$, then $P(X \backslash E)$ is true also.
(iv) If $E_1, E_2, \ldots \subset X$ are such that $P(E_n)$ is true for all n, then $P(\bigcup_{n=1}^\infty E_n)$ is true also.

Then one can conclude that $P(E)$ is true for all $E \in \langle \mathcal{F} \rangle$. Indeed, the set of all E for which $P(E)$ holds is a σ-algebra that contains $\mathcal{F}$, whence the claim. This principle is particularly useful for establishing properties of Borel measurable sets (see below).

We now turn to an important example of a σ-algebra:

Definition 1.4.16 (Borel σ-algebra)**.** Let X be a metric space, or more generally a topological space. The *Borel σ-algebra* $\mathcal{B}[X]$ of X is defined to be the σ-algebra generated by the open subsets of X. Elements of $\mathcal{B}[X]$ will be called *Borel measurable*.

Thus, for instance, the Borel σ-algebra contains the open sets, the closed sets (which are complements of open sets), the countable unions of closed sets (known as F_σ *sets*), the countable intersections of open sets (known as G_δ *sets*), the countable intersections of F_σ sets, and so forth.

In $\mathbf{R}^d$, every open set is Lebesgue measurable, and so we see that the Borel σ-algebra is coarser than the Lebesgue σ-algebra. We will shortly see, though, that the two σ-algebras are not equal.

We defined the Borel σ-algebra to be generated by the open sets. However, they are also generated by several other sets:

Exercise 1.4.14. Show that the Borel σ-algebra $\mathcal{B}[\mathbf{R}^d]$ of a Euclidean set is generated by any of the following collections of sets:

(i) The open subsets of $\mathbf{R}^d$.
(ii) The closed subsets of $\mathbf{R}^d$.
(iii) The compact subsets of $\mathbf{R}^d$.
(iv) The open balls of $\mathbf{R}^d$.
(v) The boxes in $\mathbf{R}^d$.
(vi) The elementary sets in $\mathbf{R}^d$.

(*Hint:* To show that two families $\mathcal{F}, \mathcal{F}'$ of sets generate the same σ-algebra, it suffices to show that every σ-algebra that contains $\mathcal{F}$, contains $\mathcal{F}'$ also, and conversely.)

There is an analogue of Exercise 1.4.9, which illustrates the extent to which a generated σ-algebra is "larger" than the analogous generated Boolean algebra:

Exercise 1.4.15 (Recursive description of a generated σ-algebra)**.** (This exercise requires familiarity with the theory of *ordinals*, which is reviewed in §2.4 of *An epsilon of room, Vol. I.* Recall that we are assuming the axiom of choice throughout this text.) Let $\mathcal{F}$ be a collection of sets in a set X, and let ω_1 be the first uncountable ordinal. Define the sets $\mathcal{F}_\alpha$ for every countable ordinal $\alpha \in \omega_1$ via transfinite induction as follows:

(i) $\mathcal{F}_\alpha := \mathcal{F}$.

(ii) For each countable successor ordinal $\alpha = \beta + 1$, we define $\mathcal{F}_\alpha$ to be the collection of all sets that either the union of an at most countable number of sets in $\mathcal{F}_{n-1}$ (including the empty union $\emptyset$), or the complement of such a union.

(iii) For each countable limit ordinal $\alpha = \sup_{\beta<\alpha} \beta$, we define $\mathcal{F}_\alpha := \bigcup_{\beta<\alpha} \mathcal{F}_\beta$.

Show that $\langle \mathcal{F} \rangle = \bigcup_{\alpha \in \omega_1} \mathcal{F}_\alpha$.

Remark 1.4.17. The first uncountable ordinal ω_1 will make several further cameo appearances here and in *An epsilon of room, Vol. I*, for instance by generating counterexamples to various plausible statements in point-set topology. In the case when $\mathcal{F}$ is the collection of open sets in a topological space, so that $\langle \mathcal{F} \rangle$, then the sets $\mathcal{F}_\alpha$ are essentially the *Borel hierarchy* (which starts at the open and closed sets, then moves on to the F_σ and G_δ sets, and so forth); these play an important role in *descriptive set theory.*

Exercise 1.4.16. (This exercise requires familiarity with the theory of *cardinals.*) Let $\mathcal{F}$ be an infinite family of subsets of X of cardinality κ (thus κ is an infinite cardinal). Show that $\langle \mathcal{F} \rangle$ has cardinality at most $\kappa^{\aleph_0}$. (*Hint:* use Exercise 1.4.15.) In particular, show that the Borel σ-algebra $\mathcal{B}[\mathbf{R}^d]$ has cardinality at most $c := 2^{\aleph_0}$.

Conclude that there exist Jordan measurable (and hence Lebesgue measurable) subsets of $\mathbf{R}^d$ which are not Borel measurable. (*Hint:* How many subsets of the Cantor set are there?) Use this to place the Borel σ-algebra on the diagram that you drew for Exercise 1.4.8.

Remark 1.4.18. Despite this demonstration that not all Lebesgue measurable subsets are Borel measurable, it is remarkably difficult (though not impossible) to exhibit a *specific* set that is not Borel measurable. Indeed, a large majority of the explicitly constructible sets that one actually encounters in practice tend to be Borel measurable, and one can view the property of Borel measurability intuitively as a kind of "constructibility" property. (Indeed, as a *very* crude first approximation, one can view the Borel measurable sets as those sets of "countable descriptive complexity"; in contrast, sets of finite descriptive complexity tend to be Jordan measurable (assuming they are bounded, of course).

Exercise 1.4.17. Let E, F be Borel measurable subsets of $\mathbf{R}^{d_1}, \mathbf{R}^{d_2}$ respectively. Show that $E \times F$ is a Borel measurable subset of $\mathbf{R}^{d_1+d_2}$. (*Hint:* first establish this in the case when F is a box, by using Remark 1.4.15. To obtain the general case, apply Remark 1.4.15 yet again.)

The above exercise has a partial converse:

Exercise 1.4.18. Let E be a Borel measurable subset of $\mathbf{R}^{d_1+d_2}$.

(i) Show that for any $x_1 \in \mathbf{R}^{d_1}$, the slice $\{x_2 \in \mathbf{R}^{d_2} : (x_1, x_2) \in E\}$ is a Borel measurable subset of $\mathbf{R}^{d_2}$. Similarly, show that for every $x_2 \in \mathbf{R}^{d_2}$, the slice $\{x_1 \in \mathbf{R}^{d_1} : (x_1, x_2) \in E\}$ is a Borel measurable subset of $\mathbf{R}^{d_1}$.

(ii) Give a counterexample to show that this claim is *not* true if "Borel" is replaced with "Lebesgue" throughout. (*Hint:* the Cartesian product of any set with a point is a null set, even if the first set was not measurable.)

Exercise 1.4.19. Show that the Lebesgue σ-algebra on $\mathbf{R}^d$ is generated by the union of the Borel σ-algebra and the null σ-algebra.

1.4.3. Countably additive measures and measure spaces. Having set out the concept of a σ-algebra a measurable space, we now endow these structures with a measure.

We begin with the finitely additive theory, although this theory is too weak for our purposes and will soon be supplanted by the countably additive theory.

Definition 1.4.19 (Finitely additive measure)**.** Let $\mathcal{B}$ be a Boolean algebra on a space X. An (unsigned) *finitely additive measure* μ on $\mathcal{B}$ is a map $\mu \colon \mathcal{B} \to [0, +\infty]$ that obeys the following axioms:

(i) (Empty set) $\mu(\emptyset) = 0$.

(ii) (Finite additivity) Whenever $E, F \in \mathcal{B}$ are disjoint, then $\mu(E \cup F) = \mu(E) + \mu(F)$.

Remark 1.4.20. The empty set axiom is needed in order to rule out the degenerate situation in which every set (including the empty set) has infinite measure.

Example 1.4.21. Lebesgue measure m is a finitely additive measure on the Lebesgue σ-algebra, and hence on all sub-algebras (such as the null algebra, the Jordan algebra, or the elementary algebra). In particular, Jordan measure and elementary measure are finitely additive (adopting the convention that co-Jordan measurable sets have infinite Jordan measure, and co-elementary sets have infinite elementary measure).

On the other hand, as we saw in previous notes, Lebesgue outer measure is *not* finitely additive on the discrete algebra, and Jordan outer measure is *not* finitely additive on the Lebesgue algebra.

Example 1.4.22 (Dirac measure)**.** Let $x \in X$ and $\mathcal{B}$ be an arbitrary Boolean algebra on X. Then the *Dirac measure* δ_x at x, defined by setting $\delta_x(E) := 1_E(x)$, is finitely additive.

Example 1.4.23 (Zero measure)**.** The zero measure $0\colon E \mapsto 0$ is a finitely additive measure on any Boolean algebra.

Example 1.4.24 (Linear combinations of measures)**.** If $\mathcal{B}$ is a Boolean algebra on X, and $\mu, \nu\colon \mathcal{B} \to [0, +\infty]$ are finitely additive measures on $\mathcal{B}$, then $\mu+\nu\colon E \mapsto \mu(E)+\nu(E)$ is also a finitely additive measure, as is $c\mu\colon E \mapsto c \times \mu(E)$ for any $c \in [0, +\infty]$. Thus, for instance, the sum of Lebesgue measure and a Dirac measure is also a finitely additive measure on the Lebesgue algebra (or on any of its sub-algebras).

Example 1.4.25 (Restriction of a measure)**.** If $\mathcal{B}$ is a Boolean algebra on X, $\mu\colon \mathcal{B} \to [0, +\infty]$ is a finitely additive measure, and Y is a $\mathcal{B}$-*measurable* subset of X, then the restriction $\mu\downharpoonright_Y\colon \mathcal{B}\downharpoonright_Y \to [0, +\infty]$ of $\mathcal{B}$ to Y, defined by setting $\mu\downharpoonright_Y(E) := \mu(E)$ whenever $E \in \mathcal{B}\downharpoonright_Y$ (i.e. if $E \in \mathcal{B}$ and $E \subset Y$), is also a finitely additive measure.

Example 1.4.26 (Counting measure)**.** If $\mathcal{B}$ is a Boolean algebra on X, then the function $\#\colon \mathcal{B} \to [0, +\infty]$ defined by setting $\#(E)$ to be the cardinality of E if E is finite, and $\#(E) := +\infty$ if E is infinite, is a finitely additive measure, known as *counting measure*.

As with our definition of Boolean algebras and σ-algebras, we adopted a "minimalist" definition so that the axioms are easy to verify. But they imply several further useful properties:

Exercise 1.4.20. Let $\mu\colon \mathcal{B} \to [0, +\infty]$ be a finitely additive measure on a Boolean σ-algebra $\mathcal{B}$. Establish the following facts:

(i) (Monotonicity) If E, F are $\mathcal{B}$-measurable and $E \subset F$, then $\mu(E) \leq \mu(F)$.

(ii) (Finite additivity) If k is a natural number, and $E_1, \ldots, E_k$ are $\mathcal{B}$-measurable and disjoint, then $\mu(E_1 \cup \ldots \cup E_k) = \mu(E_1) + \ldots + \mu(E_k)$.

(iii) (Finite subadditivity) If k is a natural number, and $E_1, \ldots, E_k$ are $\mathcal{B}$-measurable, then $\mu(E_1 \cup \ldots \cup E_k) \leq \mu(E_1) + \ldots + \mu(E_k)$.

(iv) (Inclusion-exclusion for two sets) If E, F are $\mathcal{B}$-measurable, then $\mu(E \cup F) + \mu(E \cap F) = \mu(E) + \mu(F)$.

(*Caution:* remember that the cancellation law $a+c = b+c \implies a = b$ does not hold in $[0, +\infty]$ if c is infinite, and so the use of cancellation (or subtraction) should be avoided if possible.)

One can characterise measures completely for any finite algebra:

Exercise 1.4.21. Let $\mathcal{B}$ be a finite Boolean algebra, generated by a finite family $A_1, \ldots, A_k$ of non-empty atoms. Show that for every finitely additive measure μ on $\mathcal{B}$ there exists $c_1, \ldots, c_k \in [0, +\infty]$ such that

$$\mu(E) = \sum_{1 \leq j \leq k: A_j \subset E} c_j.$$

Equivalently, if x_j is a point in A_j for each $1 \leq j \leq k$, then

$$\mu = \sum_{j=1}^{k} c_j \delta_{x_j}.$$

Furthermore, show that the $c_1, \ldots, c_k$ are uniquely determined by μ.

This is about the limit of what one can say about finitely additive measures at this level of generality. We now specialise to the *countably additive measures* on σ-algebras.

Definition 1.4.27 (Countably additive measure)**.** Let $(X, \mathcal{B})$ be a measurable space. An (unsigned) *countably additive measure* μ on $\mathcal{B}$, or *measure* for short, is a map $\mu\colon \mathcal{B} \to [0, +\infty]$ that obeys the following axioms:

(i) (Empty set) $\mu(\emptyset) = 0$.

(ii) (Countable additivity) Whenever $E_1, E_2, \ldots \in \mathcal{B}$ are a countable sequence of disjoint measurable sets, then $\mu(\bigcup_{n=1}^\infty E_n) = \sum_{n=1}^\infty \mu(E_n)$.

A triplet $(X, \mathcal{B}, \mu)$, where $(X, \mathcal{B})$ is a measurable space and $\mu\colon \mathcal{B} \to [0, +\infty]$ is a countably additive measure, is known as a *measure space*.

Note the distinction between a measure space and a measurable space. The latter has the *capability* to be equipped with a measure, but the former is *actually* equipped with a measure.

Example 1.4.28. Lebesgue measure is a countably additive measure on the Lebesgue σ-algebra, and hence on every sub-σ-algebra (such as the Borel σ-algebra).

Example 1.4.29. The Dirac measures from Exercise 1.4.22 are countably additive, as is counting measure.

Example 1.4.30. Any restriction of a countably additive measure to a measurable subspace is again countably additive.

Exercise 1.4.22 (Countable combinations of measures)**.** Let $(X, \mathcal{B})$ be a measurable space.

(i) If μ is a countably additive measure on $\mathcal{B}$, and $c \in [0, +\infty]$, then $c\mu$ is also countably additive.

(ii) If $\mu_1, \mu_2, \ldots$ are a sequence of countably additive measures on $\mathcal{B}$, then the sum $\sum_{n=1}^\infty \mu_n : E \mapsto \sum_{n=1}^\infty \mu_n(E)$ is also a countably additive measure.

Note that countable additivity measures are necessarily finitely additive (by padding out a finite union into a countable union using the empty set), and so countably additive measures inherit all the properties of finitely additive properties, such as monotonicity and finite subadditivity. But one also has additional properties:

Exercise 1.4.23. Let $(X, \mathcal{B}, \mu)$ be a measure space.

(i) (Countable subadditivity) If $E_1, E_2, \dots$ are $\mathcal{B}$-measurable, then $\mu(\bigcup_{n=1}^\infty E_n) \leq \sum_{n=1}^\infty \mu(E_n)$.

(ii) (Upwards monotone convergence) If $E_1 \subset E_2 \subset \dots$ are $\mathcal{B}$-measurable, then

$$\mu(\bigcup_{n=1}^\infty E_n) = \lim_{n\to\infty} \mu(E_n) = \sup_n \mu(E_n).$$

(iii) (Downwards monotone convergence) If $E_1 \supset E_2 \supset \dots$ are $\mathcal{B}$-measurable, and $\mu(E_n) < \infty$ for at least one n, then

$$\mu(\bigcap_{n=1}^\infty E_n) = \lim_{n\to\infty} \mu(E_n) = \inf_n \mu(E_n).$$

Show that the downward monotone convergence claim can fail if the hypothesis that $\mu(E_n) < \infty$ for at least one n is dropped. (*Hint:* mimic the solution to Exercise 1.2.11.)

Exercise 1.4.24 (Dominated convergence for sets)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space. Let $E_1, E_2, \dots$ be a sequence of $\mathcal{B}$-measurable sets that converge to another set E, in the sense that 1_{E_n} converges pointwise to 1_E.

(i) Show that E is also $\mathcal{B}$-measurable.

(ii) If there exists a $\mathcal{B}$-measurable set F of finite measure (i.e. $\mu(F) < \infty$) that contains all of the E_n, show that $\lim_{n\to\infty} \mu(E_n) = \mu(E)$. (*Hint:* Apply downward monotonicity to the sets $\bigcup_{n>N} (E_n \Delta E)$.)

(iii) Show that the previous part of this exercise can fail if the hypothesis that all the E_n are contained in a set of finite measure is omitted.

Exercise 1.4.25. Let X be an at most countable set with the discrete σ-algebra. Show that every measure μ on this measurable space can be uniquely represented in the form

$$\mu = \sum_{x \in X} c_x \delta_x$$

for some $c_x \in [0, +\infty]$, thus

$$\mu(E) = \sum_{x \in E} c_x$$

for all $E \subset X$. (This claim fails in the uncountable case, although showing this is slightly tricky.)

A useful technical property, enjoyed by some measure spaces, is that of *completeness*:

Definition 1.4.31 (Completeness)**.** A *null set* of a measure space $(X, \mathcal{B}, \mu)$ is defined to be a $\mathcal{B}$-measurable set of measure zero. A *sub-null set* is any subset of a null set. A measure space is said to be *complete* if every sub-null set is a null set.

Thus, for instance, the Lebesgue measure space $(\mathbf{R}^d, \mathcal{L}[\mathbf{R}^d], m)$ is complete, but the Borel measure space $(\mathbf{R}^d, \mathcal{B}[\mathbf{R}^d], m)$ is not (as can be seen from the solution to Exercise 1.4.16).

Completion is a convenient property to have in some cases, particularly when dealing with properties that hold almost everywhere. Fortunately, it is fairly easy to modify any measure space to be complete:

Exercise 1.4.26 (Completion)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space. Show that there exists a unique refinement $(X, \overline{\mathcal{B}}, \overline{\mu})$, known as the *completion* of $(X, \mathcal{B}, \mu)$, which is the coarsest refinement of $(X, \mathcal{B}, \mu)$ that is complete. Furthermore, show that $\overline{\mathcal{B}}$ consists precisely of those sets that differ from a $\mathcal{B}$-measurable set by a $\mathcal{B}$-subnull set.

Exercise 1.4.27. Show that the Lebesgue measure space $(\mathbf{R}^d, \mathcal{L}[\mathbf{R}^d], m)$ is the completion of the Borel measure space $(\mathbf{R}^d, \mathcal{B}[\mathbf{R}^d], m)$.

Exercise 1.4.28 (Approximation by an algebra)**.** Let $\mathcal{A}$ be a Boolean algebra on X, and let μ be a measure on $\langle \mathcal{A} \rangle$.

(i) If $\mu(X) < \infty$, show that for every $E \in \langle \mathcal{A} \rangle$ and $\varepsilon > 0$ there exists $F \in \mathcal{A}$ such that $\mu(E \Delta F) < \varepsilon$.

(ii) More generally, if $X = \bigcup_{n=1}^\infty A_n$ for some $A_1, A_2, \ldots \in \mathcal{A}$ with $\mu(A_n) < \infty$ for all n, $E \in \langle \mathcal{A} \rangle$ has finite measure, and $\varepsilon > 0$, show that there exists $F \in \mathcal{A}$ such that $\mu(E \Delta F) < \varepsilon$.

1.4.4. Measurable functions, and integration on a measure space. Now we are ready to define integration on measure spaces. We first need the notion of a *measurable function*, which is analogous to that of a continuous function in topology. Recall that a function $f\colon X \to Y$ between two topological spaces X, Y is continuous if the inverse image $f^{-1}(U)$ of any open set is open. In a similar spirit, we have

Definition 1.4.32. Let $(X, \mathcal{B})$ be a measurable space, and let $f\colon X \to [0, +\infty]$ or $f\colon X \to \mathbf{C}$ be an unsigned or complex-valued function. We say that f is *measurable* if $f^{-1}(U)$ is $\mathcal{B}$-measurable for every open subset U of $[0, +\infty]$ or $\mathbf{C}$.

From Lemma 1.3.9, we see that this generalises the notion of a Lebesgue measurable function.

Exercise 1.4.29. Let $(X, \mathcal{B})$ be a measurable space.

(i) Show that a function $f\colon X \to [0, +\infty]$ is measurable if and only if the level sets $\{x \in X : f(x) > \lambda\}$ are $\mathcal{B}$-measurable.

(ii) Show that an indicator function 1_E of a set $E \subset X$ is measurable if and only if E itself is $\mathcal{B}$-measurable.

(iii) Show that a function $f\colon X \to [0, +\infty]$ or $f\colon X \to \mathbf{C}$ is measurable if and only if $f^{-1}(E)$ is $\mathcal{B}$-measurable for every Borel-measurable subset E of $[0, +\infty]$ or $\mathbf{C}$.

(iv) Show that a function $f\colon X \to \mathbf{C}$ is measurable if and only if its real and imaginary parts are measurable.

(v) Show that a function $f\colon X \to \mathbf{R}$ is measurable if and only if the magnitudes $f_+ := \max(f, 0)$, $f_- := \max(-f, 0)$ of its positive and negative parts are measurable.

(vi) If $f_n\colon X \to [0, +\infty]$ are a sequence of measurable functions that converge pointwise to a limit $f\colon X \to [0, +\infty]$, then show that f is also measurable. Obtain the same claim if $[0, +\infty]$ is replaced by $\mathbf{C}$.

(vii) If $f: X \to [0, +\infty]$ is measurable and $\phi: [0, +\infty] \to [0, +\infty]$ is continuous, show that $\phi \circ f$ is measurable. Obtain the same claim if $[0, +\infty]$ is replaced by $\mathbf{C}$.

(viii) Show that the sum or product of two measurable functions in $[0, +\infty]$ or $\mathbf{C}$ is still measurable.

Remark 1.4.33. One can also view measurable functions in a more *category theoretic* fashion. Define *measurable morphism* or *measurable map* f from one measurable space $(X, \mathcal{B})$ to another $(Y, \mathcal{C})$ to be a function $f: X \to Y$ with the property that $f^{-1}(E)$ is $\mathcal{B}$-measurable for every $\mathcal{C}$-measurable set E. Then a measurable function $f: X \to [0, +\infty]$ or $f: X \to \mathbf{C}$ is the same thing as a measurable morphism from X to $[0, +\infty]$ or $\mathbf{C}$, where the latter is equipped with the Borel σ-algebra. Also, one σ-algebra $\mathcal{B}$ on a space X is coarser than another $\mathcal{B}'$ precisely when the identity map $\mathrm{id}_X: X \to X$ is a measurable morphism from $(X, \mathcal{B}')$ to $(X, \mathcal{B})$. The main purpose of adopting this viewpoint is that it is obvious that the composition of measurable morphisms is again a measurable morphism. This is important in those fields of mathematics, such as *ergodic theory* (discussed in [**Ta2009**]), in which one frequently wishes to compose measurable transformations (and in particular, to compose a transformation $T: (X, \mathcal{B}) \to (X, \mathcal{B})$ with itself repeatedly); but it will not play a major role in this text.

Measurable functions are particularly easy to describe on atomic spaces:

Exercise 1.4.30. Let $(X, \mathcal{B})$ be a measurable space that is atomic, thus $\mathcal{B} = \mathcal{A}((A_\alpha)_{\alpha \in I})$ for some partition $X = \bigcup_{\alpha \in I} A_\alpha$ of X into disjoint non-empty atoms. Show that a function $f: X \to [0, +\infty]$ or $f: X \to \mathbf{C}$ is measurable if and only if it is constant on each atom, or equivalently if one has a representation of the form

$$f = \sum_{\alpha \in I} c_\alpha 1_{A_\alpha}$$

for some constants c_α in $[0, +\infty]$ or in $\mathbf{C}$ as appropriate. Furthermore, the c_α are uniquely determined by f.

Exercise 1.4.31 (Egorov's theorem)**.** Let $(X, \mathcal{B}, \mu)$ be a finite measure space (so $\mu(X) < \infty$), and let $f_n: X \to \mathbf{C}$ be a sequence of measurable functions that converge pointwise almost everywhere to a limit $f\colon X \to \mathbf{C}$, and let $\varepsilon > 0$. Show that there exists a measurable set E of measure at most ε such that f_n converges uniformly to f outside of E. Give an example to show that the claim can fail when the measure μ is not finite.

In Section 1.3 we defined first an simple integral, then an unsigned integral, and then finally an absolutely convergent integral. We perform the same three stages here. We begin with the simple integral, which in the abstract setting becomes integration in the case when the σ-algebra is finite:

Definition 1.4.34 (Simple integral)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space with $\mathcal{B}$ finite. By Exercise 1.4.4, X is partitioned into a finite number of atoms $A_1, \ldots, A_n$. If $f\colon X \to [0, +\infty]$ is measurable, then by Exercise 1.4.30 it has a unique representation of the form

$$f = \sum_{i=1}^{n} c_i 1_{A_i}$$

for some $c_1, \ldots, c_n \in [0, +\infty]$. We then define the *simple integral* $\operatorname{Simp} \int_X f\ d\mu$ of f by the formula

$$\operatorname{Simp} \int_X f\ d\mu := \sum_{i=1}^{n} c_i \mu(A_i).$$

Note that, thanks to Exercise 1.4.3, the precise decomposition into atoms does not affect the definition of the simple integral.

Exercise 1.4.32. Propose a definition for the simple integral for absolutely convergent complex-valued functions on a measurable space with a finite σ-algebra.

With this definition, it is clear that one has the monotonicity property

$$\operatorname{Simp} \int_X f\ d\mu \leq \operatorname{Simp} \int_X g\ d\mu$$

whenever $f \leq g$ are unsigned measurable, as well as the linearity properties

$$\operatorname{Simp} \int_X f + g\ d\mu = \operatorname{Simp} \int_X f\ d\mu + \operatorname{Simp} \int_X g\ d\mu$$

and

$$\mathrm{Simp}\int_X cf\ d\mu = c \times \mathrm{Simp}\int_X f\ d\mu$$

for unsigned measurable f, g and $c \in [0, +\infty]$. We also make the following important technical observation:

Exercise 1.4.33 (Simple integral unaffected by refinements)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $(X, \mathcal{B}', \mu')$ be a *refinement* of $(X, \mathcal{B}, \mu)$, which means that $\mathcal{B}'$ contains $\mathcal{B}$ and $\mu' \colon \mathcal{B}' \to [0, +\infty]$ agrees with $\mu \colon \mathcal{B} \to [0, +\infty]$ on $\mathcal{B}$. Suppose that both $\mathcal{B}, \mathcal{B}'$ are finite, and let $f \colon \mathcal{B} \to [0, +\infty]$ be measurable. Show that

$$\mathrm{Simp}\int_X f\ d\mu = \mathrm{Simp}\int_X f\ d\mu'.$$

This allows one to extend the simple integral to simple functions:

Definition 1.4.35 (Integral of simple functions)**.** An (unsigned) *simple function* $f \colon X \to [0, +\infty]$ on a measurable space $(X, \mathcal{B})$ is a measurable function that takes on finitely many values $a_1, \ldots, a_k$. Note that such a function is then automatically measurable with respect to at least one finite sub-σ-algebra $\mathcal{B}'$ of $\mathcal{B}$, namely the σ-algebra $\mathcal{B}'$ generated by the preimages $f^{-1}(\{a_1\}), \ldots, f^{-1}(\{a_k\})$ of $a_1, \ldots, a_k$. We then define the simple integral $\mathrm{Simp}\int_X f\ d\mu$ by the formula

$$\mathrm{Simp}\int_X f\ d\mu := \mathrm{Simp}\int_X f\ d\mu\downharpoonright_{\mathcal{B}'},$$

where $\mu\downharpoonright_{\mathcal{B}'} \colon \mathcal{B}' \to [0, +\infty]$ is the restriction of $\mu \colon \mathcal{B} \to [0, +\infty]$ to $\mathcal{B}'$.

Note that there could be multiple finite σ-algebras with respect to which f is measurable, but Exercise 1.4.33 guarantees that all such extensions will give the same simple integral. Indeed, if f were measurable with respect to two separate finite sub-σ-algebras $\mathcal{B}'$ and $\mathcal{B}''$ of $\mathcal{B}$, then it would also be measurable with respect to their common refinement $\mathcal{B}' \vee \mathcal{B}'' := \langle \mathcal{B}' \cup \mathcal{B}'' \rangle$, which is also finite (by Exercise 1.4.8), and then by Exercise 1.4.33, $\int_X f\ d\mu\downharpoonright_{\mathcal{B}'}$ and $\int_X f\ d\mu\downharpoonright_{\mathcal{B}''}$ are both equal to $\int_X f\ d\mu\downharpoonright_{\mathcal{B}' \vee \mathcal{B}''}$, and hence equal to each other.

From this we can deduce the following properties of the simple integral. As with the Lebesgue theory, we say that a property $P(x)$ of an element $x \in X$ of a measure space $(X, \mathcal{B}, \mu)$ holds μ-*almost everywhere* if it holds outside of a sub-null set.

Exercise 1.4.34 (Basic properties of the simple integral)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f, g \colon X \to [0, +\infty]$ be simple functions.

(i) (Monotonicity) If $f \leq g$ pointwise, then $\mathrm{Simp} \int_X f \, d\mu \leq \mathrm{Simp} \int_X g \, d\mu$.

(ii) (Compatibility with measure) For every $\mathcal{B}$-measurable set E, we have $\mathrm{Simp} \int_X 1_E \, d\mu = \mu(E)$.

(iii) (Homogeneity) For every $c \in [0, +\infty]$, one has $\mathrm{Simp} \int_X cf \, d\mu = c \times \mathrm{Simp} \int_X f \, d\mu$.

(iv) (Finite additivity) $\mathrm{Simp} \int_X (f + g) \, d\mu = \mathrm{Simp} \int_X f \, d\mu + \mathrm{Simp} \int_X g \, d\mu$.

(v) (Insensitivity to refinement) If $(X, \mathcal{B}', \mu')$ is a refinement of $(X, \mathcal{B}, \mu)$ (as defined in Exercise 1.4.33), then $\mathrm{Simp} \int_X f \, d\mu = \mathrm{Simp} \int_X f \, d\mu'$.

(vi) (Almost everywhere equivalence) If $f(x) = g(x)$ for μ-almost every $x \in X$, then $\mathrm{Simp} \int_X f \, d\mu = \mathrm{Simp} \int_X g \, d\mu$.

(vii) (Finiteness) $\mathrm{Simp} \int_X f \, d\mu < \infty$ if and only if f is finite almost everywhere, and is supported on a set of finite measure.

(viii) (Vanishing) $\mathrm{Simp} \int_X f \, d\mu = 0$ if and only if f is zero almost everywhere.

Exercise 1.4.35 (Inclusion-exclusion principle)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $A_1, \ldots, A_n$ be $\mathcal{B}$-measurable sets of finite measure. Show that

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{J \subset \{1,\ldots,n\} : J \neq \emptyset} (-1)^{|J|-1} \mu\left(\bigcap_{i \in J} A_i\right).$$

(*Hint:* Compute $\mathrm{Simp} \int_X (1 - \prod_{i=1}^n (1 - 1_{A_i})) \, d\mu$ in two different ways.)

Remark 1.4.36. The simple integral could also be defined on finitely additive measure spaces, rather than countably additive ones, and all the above properties would still apply. However, on a finitely additive measure space one would have difficulty extending the integral beyond simple functions, as we will now do.

From the simple integral, we can now define the unsigned integral, in analogy to how the unsigned Lebesgue integral was constructed in Section 1.3.3.

Definition 1.4.37. Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f \colon X \to [0, +\infty]$ be measurable. Then we define the unsigned integral $\int_X f \, d\mu$ of f by the formula

$$\int_X f \, d\mu := \sup_{0 \leq g \leq f; g \text{ simple}} \operatorname{Simp} \int_X g \, d\mu. \tag{1.14}$$

Clearly, this definition generalises Definition 1.3.13. Indeed, if $f \colon \mathbf{R}^d \to [0, +\infty]$ is Lebesgue measurable, then $\int_{\mathbf{R}^d} f(x) \, dx = \int_{\mathbf{R}^d} f \, dm$.

We record some easy properties of this integral:

Exercise 1.4.36 (Easy properties of the unsigned integral)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f, g \colon X \to [0, +\infty]$ be measurable.

(i) (Almost everywhere equivalence) If $f = g$ μ-almost everywhere, then $\int_X f \, d\mu = \int_X g \, d\mu$

(ii) (Monotonicity) If $f \leq g$ μ-almost everywhere, then $\int_X f \, d\mu \leq \int_X g \, d\mu$.

(iii) (Homogeneity) We have $\int_X cf \, d\mu = c \int_X f \, d\mu$ for every $c \in [0, +\infty]$.

(iv) (Superadditivity) We have $\int_X (f+g) \, d\mu \geq \int_X f \, d\mu + \int_X g \, d\mu$.

(v) (Compatibility with the simple integral) If f is simple, then $\int_X f \, d\mu = \operatorname{Simp} \int_X f \, d\mu$.

(vi) (Markov's inequality) For any $0 < \lambda < \infty$, one has

$$\mu(\{x \in X : f(x) \geq \lambda\}) \leq \frac{1}{\lambda} \int_X f \, d\mu.$$

In particular, if $\int_X f \, d\mu < \infty$, then the sets $\{x \in X : f(x) \geq \lambda\}$ have finite measure for each $\lambda > 0$.

(vii) (Finiteness) If $\int_X f \, d\mu < \infty$, then $f(x)$ is finite for μ-almost every x.

(viii) (Vanishing) If $\int_X f \, d\mu = 0$, then $f(x)$ is zero for μ-almost every x.

(ix) (Vertical truncation) We have $\lim_{n\to\infty} \int_X \min(f,n)\ d\mu = \int_X f\ d\mu$.

(x) (Horizontal truncation) If $E_1 \subset E_2 \subset \dots$ is an increasing sequence of $\mathcal{B}$-measurable sets, then

$$\lim_{n\to\infty} \int_X f 1_{E_n}\ d\mu = \int_X f 1_{\bigcup_{n=1}^\infty E_n}\ d\mu.$$

(xi) (Restriction) If Y is a measurable subset of X, then $\int_X f 1_Y\ d\mu = \int_Y f\ \downharpoonright_Y\ d\mu\ \downharpoonright_Y$, where $f\ \downharpoonright_Y : Y \to [0,+\infty]$ is the restriction of $f\colon X \to [0,+\infty]$ to Y, and the restriction $\mu\ \downharpoonright_Y$ was defined in Example 1.4.25. We will often abbreviate $\int_Y f\ \downharpoonright_Y\ d\mu\ \downharpoonright_Y$ (by slight abuse of notation) as $\int_Y f\ d\mu$.

As before, one of the key properties of this integral is its additivity:

Theorem 1.4.38. *Let $(X,\mathcal{B},\mu)$ be a measure space, and let $f,g\colon X \to [0,+\infty]$ be measurable. Then*

$$\int_X (f+g)\ d\mu = \int_X f\ d\mu + \int_X g\ d\mu.$$

Proof. In view of superadditivity, it suffices to establish the subadditivity property

$$\int_X (f+g)\ d\mu \leq \int_X f\ d\mu + \int_X g\ d\mu$$

We establish this in stages. We first deal with the case when μ is a *finite* measure (which means that $\mu(X) < \infty$) and f,g are bounded. Pick an $\varepsilon > 0$, and let f_ε be f rounded down to the nearest integer multiple of ε, and f^ε be f rounded up to the nearest integer multiple. Clearly, we have the pointwise bounds

$$f_\varepsilon(x) \leq f(x) \leq f^\varepsilon(x)$$

and

$$f^\varepsilon(x) - f_\varepsilon(x) \leq \varepsilon.$$

Since f is bounded, f_ε and f^ε are simple. Similarly define $g_\varepsilon, g^\varepsilon$. We then have the pointwise bound

$$f+g \leq f^\varepsilon + g^\varepsilon \leq f_\varepsilon + g_\varepsilon + 2\varepsilon,$$

hence by Exercise 1.4.36 and the properties of the simple integral,

$$\begin{aligned}\int_X f+g\ d\mu &\leq \int_X f_\varepsilon + g_\varepsilon + 2\varepsilon\ d\mu \\ &= \mathrm{Simp}\int_X f_\varepsilon + g_\varepsilon + 2\varepsilon\ d\mu \\ &= \mathrm{Simp}\int_X f_\varepsilon\ d\mu + \mathrm{Simp}\int_X g_\varepsilon\ d\mu + 2\varepsilon\mu(X).\end{aligned}$$

From (1.14) we conclude that

$$\int_X f+g\ d\mu \leq \int_X f\ d\mu + \int_X g\ d\mu + 2\varepsilon\mu(X).$$

Letting $\varepsilon \to 0$ and using the assumption that $\mu(X)$ is finite, we obtain the claim.

Now we continue to assume that μ is a finite measure, but now do not assume that f, g are bounded. Then for any natural number n, we can use the previous case to deduce that

$$\int_X \min(f,n)+\min(g,n)\ d\mu \leq \int_X \min(f,n)\ d\mu + \int_X \min(g,n)\ d\mu.$$

Since $\min(f+g,n) \leq \min(f,n)+\min(g,n)$, we conclude that

$$\int_X \min(f+g,n) \leq \int_X \min(f,n)\ d\mu + \int_X \min(g,n)\ d\mu.$$

Taking limits as $n \to \infty$ using vertical truncation, we obtain the claim.

Finally, we no longer assume that μ is of finite measure, and also do not require f, g to be bounded. If either $\int_X f\ d\mu$ or $\int_X g\ d\mu$ is infinite, then by monotonicity, $\int_X f+g\ d\mu$ is infinite as well, and the claim follows; so we may assume that $\int_X f\ d\mu$ and $\int_X g\ d\mu$ are both finite. By Markov's inequality (Exercise 1.4.36(vi)), we conclude that for each natural number n, the set $E_n := \{x \in X : f(x) > \frac{1}{n}\} \cup \{x \in X : g(x) > \frac{1}{n}\}$ has finite measure. These sets are increasing in n, and $f, g, f+g$ are supported on $\bigcup_{n=1}^\infty E_n$, and so by horizontal truncation

$$\int_X (f+g)\ d\mu = \lim_{n\to\infty}\int_X (f+g)1_{E_n}\ d\mu.$$

From the previous case, we have

$$\int_X (f+g)1_{E_n}\ d\mu \leq \int_X f1_{E_n}\ d\mu + \int_X g1_{E_n}\ d\mu.$$

Letting $n \to \infty$ and using horizontal truncation we obtain the claim. $\square$

Exercise 1.4.37 (Linearity in μ)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f\colon X \to [0, +\infty]$ be measurable.

(i) Show that $\int_X f\ d(c\mu) = c \times \int_X f\ d\mu$ for every $c \in [0, +\infty]$.

(ii) If $\mu_1, \mu_2, \ldots$ are a sequence of measures on $\mathcal{B}$, show that

$$\int_X f\ d\sum_{n=1}^\infty \mu_n = \sum_{n=1}^\infty \int_X f\ d\mu_n.$$

Exercise 1.4.38 (Change of variables formula)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $\phi\colon X \to Y$ be a measurable morphism (as defined in Remark 1.4.33) from $(X, \mathcal{B})$ to another measurable space $(Y, \mathcal{C})$. Define the *pushforward* $\phi_*\mu\colon \mathcal{C} \to [0, +\infty]$ of μ by ϕ by the formula $\phi_*\mu(E) := \mu(\phi^{-1}(E))$.

(i) Show that $\phi_*\mu$ is a measure on $\mathcal{C}$, so that $(Y, \mathcal{C}, \phi_*\mu)$ is a measure space.

(ii) If $f\colon Y \to [0, +\infty]$ is measurable, show that $\int_Y f\ d\phi_*\mu = \int_X (f \circ \phi)\ d\mu$.

(*Hint:* the quickest proof here is via the monotone convergence theorem (Theorem 1.4.44) below, but it is also possible to prove the exercise without this theorem.)

Exercise 1.4.39. Let $T\colon \mathbf{R}^d \to \mathbf{R}^d$ be an invertible linear transformation, and let m be Lebesgue measure on $\mathbf{R}^d$. Show that $T_*m = \frac{1}{|\det T|}m$, where the pushforward T_*m of m was defined in Exercise 1.4.38.

Exercise 1.4.40 (Sums as integrals)**.** Let X be an arbitrary set (with the discrete σ-algebra), let $\#$ be counting measure (see Exercise 1.4.26), and let $f\colon X \to [0, +\infty]$ be an arbitrary unsigned function.

Show that f is measurable with

$$\int_X f \, d\# = \sum_{x \in X} f(x).$$

Once one has the unsigned integral, one can define the absolutely convergent integral exactly as in the Lebesgue case:

Definition 1.4.39 (Absolutely convergent integral)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space. A measurable function $f \colon X \to \mathbf{C}$ is said to be *absolutely integrable* if the unsigned integral

$$\|f\|_{L^1(X,\mathcal{B},\mu)} := \int_X |f| \, d\mu$$

is finite, and use $L^1(X, \mathcal{B}, \mu)$, $L^1(X)$, or $L^1(\mu)$ to denote the space of absolutely integrable functions. If f is real-valued and absolutely integrable, we define the integral $\int_X f \, d\mu$ by the formula

$$\int_X f \, d\mu := \int_X f_+ \, d\mu - \int_X f_- \, d\mu$$

where $f_+ := \max(f, 0)$, $f_- := \max(-f, 0)$ are the magnitudes of the positive and negative components of f. If f is complex-valued and absolutely integrable, we define the integral $\int_X f \, d\mu$ by the formula

$$\int_X f \, d\mu := \int_X \operatorname{Re} f \, d\mu + i \int_X \operatorname{Im} f \, d\mu$$

where the two integrals on the right are interpreted as real-valued integrals. It is easy to see that the unsigned, real-valued, and complex-valued integrals defined in this manner are compatible on their common domains of definition.

Clearly, this definition generalises the Definition 1.3.17.

We record some of the key facts about the absolutely convergent integral:

Exercise 1.4.41. Let $(X, \mathcal{B}, \mu)$ be a measure space.

(i) Show that $L^1(X, \mathcal{B}, \mu)$ is a complex vector space.

(ii) Show that the integration map $f \mapsto \int_X f \, d\mu$ is a complex-linear map from $L^1(X, \mathcal{B}, \mu)$ to $\mathbf{C}$.

(iii) Establish the triangle inequality $\|f+g\|_{L^1(\mu)} \leq \|f\|_{L^1(\mu)} + \|g\|_{L^1(\mu)}$ and the homogeneity property $\|cf\|_{L^1(\mu)} = |c|\|f\|_{L^1(\mu)}$ for all $f, g \in L^1(X, \mathcal{B}, \mu)$ and $c \in \mathbf{C}$.

(iv) Show that if $f, g \in L^1(X, \mathcal{B}, \mu)$ are such that $f(x) = g(x)$ for μ-almost every $x \in X$, then $\int_X f \ d\mu = \int_X g \ d\mu$.

(v) If $f \in L^1(X, \mathcal{B}, \mu)$, and $(X, \mathcal{B}', \mu')$ is a refinement of $(X, \mathcal{B}, \mu)$, then $f \in L^1(X, \mathcal{B}', \mu')$, and $\int_X f \ d\mu' = \int_X f \ d\mu$. (*Hint:* it is easy to get one inequality. To get the other inequality, first work in the case when f is both bounded and has finite measure support (i.e. is both vertically and horizontally truncated).)

(vi) Show that if $f \in L^1(X, \mathcal{B}, \mu)$, then $\|f\|_{L^1(\mu)} = 0$ if and only if f is zero μ-almost everywhere.

(vii) If $Y \subset X$ is $\mathcal{B}$-measurable and $f \in L^1(X, \mathcal{B}, \mu)$, then $f \downharpoonright_Y \in L^1(Y, \mathcal{B} \downharpoonright_Y, \mu \downharpoonright_Y)$ and $\int_Y f \downharpoonright_Y \ d\mu \downharpoonright_Y = \int_X f 1_Y \ d\mu$. As before, by abuse of notation we write $\int_Y f \ d\mu$ for $\int_Y f \downharpoonright_Y \ d\mu \downharpoonright_Y$.

1.4.5. The convergence theorems. Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f_1, f_2, \ldots : X \to [0, +\infty]$ be a sequence of measurable functions. Suppose that as $n \to \infty$, $f_n(x)$ converges pointwise either everywhere, or μ-almost everywhere, to a measurable limit f. A basic question in the subject is to determine the conditions under which such pointwise convergence would imply convergence of the integral:

$$\int_X f_n \ d\mu \stackrel{?}{\to} \int_X f \ d\mu.$$

To put it another way: when can we ensure that one can interchange integrals and limits,

$$\lim_{n\to\infty} \int_X f_n \ d\mu \stackrel{?}{=} \int_X \lim_{n\to\infty} f_n \ d\mu?$$

There are certainly some cases in which one can safely do this:

Exercise 1.4.42 (Uniform convergence on a finite measure space)**.** Suppose that $(X, \mathcal{B}, \mu)$ is a *finite* measure space (so $\mu(X) < \infty$), and $f_n \colon X \to [0, +\infty]$ (resp. $f_n \colon X \to \mathbf{C}$) are a sequence of unsigned measurable functions (resp. absolutely integrable functions) that converge *uniformly* to a limit f. Show that $\int_X f_n \, d\mu$ converges to $\int_X f \, d\mu$.

However, there are also cases in which one cannot interchange limits and integrals, even when the f_n are unsigned. We give the three classic examples, all of "moving bump" type, though the way in which the bump moves varies from example to example:

Example 1.4.40 (Escape to horizontal infinity)**.** Let X be the real line with Lebesgue measure, and let $f_n := 1_{[n,n+1]}$. Then f_n converges pointwise to $f := 0$, but $\int_{\mathbf{R}} f_n(x) \, dx = 1$ does not converge to $\int_{\mathbf{R}} f(x) \, dx = 0$. Somehow, all the mass in the f_n has escaped by moving off to infinity in a horizontal direction, leaving none behind for the pointwise limit f.

Example 1.4.41 (Escape to width infinity)**.** Let X be the real line with Lebesgue measure, and let $f_n := \frac{1}{n} 1_{[0,n]}$. Then f_n now converges *uniformly* $f := 0$, but $\int_{\mathbf{R}} f_n(x) \, dx = 1$ still does not converge to $\int_{\mathbf{R}} f(x) \, dx = 0$. Exercise 1.4.42 would prevent this from happening if all the f_n were supported in a single set of finite measure, but the increasingly wide nature of the support of the f_n prevents this from happening.

Example 1.4.42 (Escape to vertical infinity)**.** Let X be the unit interval $[0,1]$ with Lebesgue measure (restricted from $\mathbf{R}$), and let $f_n := n 1_{[\frac{1}{n}, \frac{2}{n}]}$. Now, we have finite measure, and f_n converges pointwise to f, but no uniform convergence. And again, $\int_{[0,1]} f_n(x) \, dx = 1$ is not converging to $\int_{[0,1]} f(x) \, dx = 0$. This time, the mass has escaped vertically, through the increasingly large values of f_n.

Remark 1.4.43. From the perspective of *time-frequency analysis* (or perhaps more accurately, space-frequency analysis), these three escapes are analogous (though not quite identical) to escape to spatial infinity, escape to zero frequency, and escape to infinite frequency respectively, thus describing the three different ways in which *phase space* fails to be compact (if one excises the zero frequency as being singular).

However, once one shuts down these avenues of escape to infinity, it turns out that one can recover convergence of the integral. There are two major ways to accomplish this. One is to enforce *monotonicity*, which prevents each f_n from abandoning the location where the mass of the preceding $f_1, \ldots, f_{n-1}$ was concentrated and which thus shuts down the above three escape scenarios. More precisely, we have the *monotone convergence theorem*:

Theorem 1.4.44 (Monotone convergence theorem)**.** *Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $0 \leq f_1 \leq f_2 \leq \ldots$ be a monotone non-decreasing sequence of unsigned measurable functions on X. Then we have*

$$\lim_{n\to\infty} \int_X f_n \ d\mu = \int_X \lim_{n\to\infty} f_n \ d\mu.$$

Note that in the special case when each f_n is an indicator function $f_n = 1_{E_n}$, this theorem collapses to the upwards monotone convergence property (Exercise 1.4.23(ii)). Conversely, the upwards monotone convergence property will play a key role in the proof of this theorem.

Proof. Write $f := \lim_{n\to\infty} f_n = \sup_n f_n$, then $f\colon X \to [0, +\infty]$ is measurable. Since the f_n are non-decreasing to f, we see from monotonicity that $\int_X f_n \ d\mu$ are non-decreasing and bounded above by $\int_X f \ d\mu$, which gives the bound

$$\lim_{n\to\infty} \int_X f_n \ d\mu \leq \int_X f \ d\mu.$$

It remains to establish the reverse inequality

$$\int_X f \ d\mu \leq \lim_{n\to\infty} \int_X f_n \ d\mu.$$

By definition, it suffices to show that

$$\int_X g \ d\mu \leq \lim_{n\to\infty} \int_X f_n \ d\mu.$$

whenever g is a simple function that is bounded pointwise by f. By vertical truncation we may assume without loss of generality that g

also is finite everywhere, then we can write

$$g = \sum_{i=1}^{k} c_i 1_{A_i}$$

for some $0 \leq c_i < \infty$ and some disjoint $\mathcal{B}$-measurable sets $A_1, \ldots, A_k$, thus

$$\int_X g \, d\mu = \sum_{i=1}^{k} c_i \mu(A_i).$$

Let $0 < \varepsilon < 1$ be arbitrary. Then we have

$$f(x) = \sup_n f_n(x) > (1 - \varepsilon) c_i$$

for all $x \in A_i$. Thus, if we define the sets

$$A_{i,n} := \{x \in A_i : f_n(x) > (1 - \varepsilon) c_i\}$$

then the $A_{i,n}$ increase to A_i and are measurable. By upwards monotonicity of measure, we conclude that

$$\lim_{n \to \infty} \mu(A_{i,n}) = \mu(A_i).$$

On the other hand, observe the pointwise bound

$$f_n \geq \sum_{i=1}^{k} (1 - \varepsilon) c_i 1_{A_{i,n}}$$

for any n; integrating this, we obtain

$$\int_X f_n \, d\mu \geq (1 - \varepsilon) \sum_{i=1}^{k} c_i \mu(A_{i,n}).$$

Taking limits as $n \to \infty$, we obtain

$$\lim_{n \to \infty} \int_X f_n \, d\mu \geq (1 - \varepsilon) \sum_{i=1}^{k} c_i \mu(A_i);$$

sending $\varepsilon \to 0$ we then obtain the claim. □

Remark 1.4.45. It is easy to see that the result still holds if the monotonicity $f_n \leq f_{n+1}$ only holds almost everywhere rather than everywhere.

This has a number of important corollaries. Firstly, we can generalise (part of) Tonelli's theorem for exchanging sums (see Theorem 0.0.2):

Corollary 1.4.46 (Tonelli's theorem for sums and integrals)**.** *Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f_1, f_2, \ldots : X \to [0, +\infty]$ be a sequence of unsigned measurable functions. Then one has*

$$\int_X \sum_{n=1}^\infty f_n \, d\mu = \sum_{n=1}^\infty \int_X f_n \, d\mu.$$

Proof. Apply the monotone convergence theorem (Theorem 1.4.44) to the partial sums $F_N := \sum_{n=1}^N f_n$. □

Exercise 1.4.43. Give an example to show that this corollary can fail if the f_n are assumed to be absolutely integrable rather than unsigned measurable, even if the sum $\sum_{n=1}^\infty f_n(x)$ is absolutely convergent for each x. (*Hint:* think about the three escapes to infinity.)

Exercise 1.4.44 (Borel-Cantelli lemma)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $E_1, E_2, E_3, \ldots$ be a sequence of $\mathcal{B}$-measurable sets such that $\sum_{n=1}^\infty \mu(E_n) < \infty$. Show that almost every $x \in X$ is contained in at most finitely many of the E_n (i.e. $\{n \in \mathbf{N} : x \in E_n\}$ is finite for almost every $x \in X$). (*Hint:* Apply Tonelli's theorem to the indicator functions 1_{E_n}.)

Exercise 1.4.45.

(i) Give an alternate proof of the Borel-Cantelli lemma (Exercise 1.4.44) that does not go through any of the convergence theorems, but instead exploits the more basic properties of measure from Exercise 1.4.23.

(ii) Give a counterexample that shows that the Borel-Cantelli lemma can fail if the condition $\sum_{n=1}^\infty \mu(E_n) < \infty$ is relaxed to $\lim_{n\to\infty} \mu(E_n) = 0$.

Secondly, when one does not have monotonicity, one can at least obtain an important inequality, known as *Fatou's lemma*:

Corollary 1.4.47 (Fatou's lemma)**.** *Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f_1, f_2, \ldots : X \to [0, +\infty]$ be a sequence of unsigned measurable functions. Then*

$$\int_X \liminf_{n\to\infty} f_n \ d\mu \leq \liminf_{n\to\infty} \int_X f_n \ d\mu.$$

Proof. Write $F_N := \inf_{n \geq N} f_n$ for each N. Then the F_N are measurable and non-decreasing, and hence by the monotone convergence theorem (Theorem 1.4.44)

$$\int_X \sup_{N>0} F_N \ d\mu = \sup_{N>0} \int_X F_N \ d\mu.$$

By definition of lim inf, we have $\sup_{N>0} F_N = \liminf_{n\to\infty} f_n$. By monotonicity, we have $\int_X F_N \ d\mu \leq \int_X f_n \ d\mu$ for all $n \geq N$, and thus

$$\int_X F_N \ d\mu \leq \inf_{n \geq N} \int_X f_n \ d\mu.$$

Hence we have

$$\int_X \liminf_{n\to\infty} f_n \ d\mu \leq \sup_{N>0} \inf_{n\geq N} \int_X f_n \ d\mu.$$

The claim then follows by another appeal to the definition of the lim inferior. □

Remark 1.4.48. Informally, Fatou's lemma tells us that when taking the pointwise limit of unsigned functions f_n, that mass $\int_X f_n \ d\mu$ can be destroyed in the limit (as was the case in the three key moving bump examples), but it cannot be created in the limit. Of course the unsigned hypothesis is necessary here (consider for instance multiplying any of the moving bump examples by -1). While this lemma was stated only for pointwise limits, the same general principle (that mass can be destroyed, but not created, by the process of taking limits) tends to hold for other "weak" notions of convergence. See §1.9 of *An epsilon of room, Vol. I* for some examples of this.

Finally, we give the other major way to shut down loss of mass via escape to infinity, which is to *dominate* all of the functions involved by an absolutely convergent one. This result is known as the *dominated convergence theorem*:

Theorem 1.4.49 (Dominated convergence theorem). *Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f_1, f_2, \ldots : X \to \mathbf{C}$ be a sequence of measurable functions that converge pointwise μ-almost everywhere to a measurable limit $f: X \to \mathbf{C}$. Suppose that there is an unsigned absolutely integrable function $G: X \to [0, +\infty]$ such that $|f_n|$ are pointwise μ-almost everywhere bounded by G for each n. Then we have*

$$\lim_{n\to\infty} \int_X f_n \; d\mu = \int_X f \; d\mu.$$

From the moving bump examples we see that this statement fails if there is no absolutely integrable dominating function G. The reader is encouraged to see why, in each of the moving bump examples, no such dominating function exists, without appealing to the above theorem. Note also that when each of the f_n is an indicator function $f_n = 1_{E_n}$, the dominated convergence theorem collapses to Exercise 1.4.24.

Proof. By modifying f_n, f on a null set, we may assume without loss of generality that the f_n converge to f pointwise everywhere rather than μ-almost everywhere, and similarly we can assume that $|f_n$ are bounded by G pointwise everywhere rather than μ-almost everywhere.

By taking real and imaginary parts we may assume without loss of generality that f_n, f are real, thus $-G \leq f_n \leq G$ pointwise. Of course, this implies that $-G \leq f \leq G$ pointwise also.

If we apply Fatou's lemma (Corollary1.4.47) to the unsigned functions $f_n + G$, we see that

$$\int_X f + G \; d\mu \leq \liminf_{n\to\infty} \int_X f_n + G \; d\mu,$$

which on subtracting the *finite* quantity $\int_X G \; d\mu$ gives

$$\int_X f \; d\mu \leq \liminf_{n\to\infty} \int_X f_n \; d\mu.$$

Similarly, if we apply that lemma to the unsigned functions $G - f_n$, we obtain

$$\int_X G - f \; d\mu \leq \liminf_{n\to\infty} \int_X G - f_n \; d\mu;$$

negating this inequality and then cancelling $\int_X G \, d\mu$ again we conclude that

$$\limsup_{n\to\infty} \int_X f_n \, d\mu \leq \int_X f \, d\mu.$$

The claim then follows by combining these inequalities. $\square$

Remark 1.4.50. We deduced the dominated convergence theorem from Fatou's lemma, and Fatou's lemma from the monotone convergence theorem. However, one can obtain these theorems in a different order, depending on one's taste, as they are so closely related. For instance, in [**StSk2005**], the logic is somewhat different; one first obtains the slightly simpler *bounded convergence theorem*, which is the dominated convergence theorem under the assumption that the functions are uniformly bounded and all supported on a single set of finite measure, and then uses that to deduce Fatou's lemma, which in turn is used to deduce the monotone convergence theorem; and then the horizontal and vertical truncation properties are used to extend the bounded convergence theorem to the dominated convergence theorem. It is instructive to view a couple different derivations of these key theorems to get more of an intuitive understanding as to how they work.

Exercise 1.4.46. Under the hypotheses of the dominated convergence theorem (Theorem 1.4.49), establish also that $\|f_n - f\|_{L^1} \to 0$ as $n \to \infty$.

Exercise 1.4.47 (Almost dominated convergence)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f_1, f_2, \ldots : X \to \mathbf{C}$ be a sequence of measurable functions that converge pointwise μ-almost everywhere to a measurable limit $f \colon X \to \mathbf{C}$. Suppose that there is an unsigned absolutely integrable functions $G, g_1, g_2, \ldots : X \to [0, +\infty]$ such that the $|f_n|$ are pointwise μ-almost everywhere bounded by $G + g_n$, and that $\int_X g_n \, d\mu \to 0$ as $n \to \infty$. Show that

$$\lim_{n\to\infty} \int_X f_n \, d\mu = \int_X f \, d\mu.$$

Exercise 1.4.48 (Defect version of Fatou's lemma)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f_1, f_2, \ldots : X \to [0, +\infty]$ be a sequence of unsigned absolutely integrable functions that converges pointwise to an absolutely integrable limit f. Show that

$$\int_X f_n\ d\mu - \int_X f\ d\mu - \|f - f_n\|_{L^1(\mu)} \to 0$$

as $n \to \infty$. (*Hint:* Apply the dominated convergence theorem (Theorem 1.4.49) to $\min(f_n, f)$.) Informally, this result (first established in [**BrLi1983**]) tells us that the gap between the left and right hand sides of Fatou's lemma can be measured by the quantity $\|f - f_n\|_{L^1(\mu)}$.

Exercise 1.4.49. Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $g\colon X \to [0, +\infty]$ be measurable. Show that the function $\mu_g\colon \mathcal{B} \to [0, +\infty]$ defined by the formula

$$\mu_g(E) := \int_X 1_E g\ d\mu = \int_E g\ d\mu$$

is a measure. (Such measures are studied in greater detail in §1.2 of *An epsilon of room, Vol. I.*)

The monotone convergence theorem is, in some sense, a defining property of the unsigned integral, as the following exercise illustrates.

Exercise 1.4.50 (Characterisation of the unsigned integral)**.** Let $(X, \mathcal{B})$ be a measurable space. $I\colon f \mapsto I(f)$ be a map from the space $\mathcal{U}(X, \mathcal{B})$ of unsigned measurable functions $f\colon X \to [0, +\infty]$ to $[0, +\infty]$ that obeys the following axioms:

(i) (Homogeneity) For every $f \in \mathcal{U}(X, \mathcal{B})$ and $c \in [0, +\infty]$, one has $I(cf) = cI(f)$.

(ii) (Finite additivity) For every $f, g \in \mathcal{U}(X, \mathcal{B})$, one has $I(f + g) = I(f) + I(g)$.

(iii) (Monotone convergence) If $0 \leq f_1 \leq f_2 \leq \ldots$ are a non-decreasing sequence of unsigned measurable functions, then $I(\lim_{n\to\infty} f_n) = \lim_{n\to\infty} I(f_n)$.

Then there exists a unique measure μ on $(X, \mathcal{B})$ such that $I(f) = \int_X f\ d\mu$ for all $f \in \mathcal{U}(X, \mathcal{B})$. Furthermore, μ is given by the formula $\mu(E) := I(1_E)$ for all $\mathcal{B}$-measurable sets E.

Exercise 1.4.51. Let $(X, \mathcal{B}, \mu)$ be a finite measure space (i.e. $\mu(X) < \infty$), and let $f\colon X \to \mathbf{R}$ be a bounded function. Suppose that μ is *complete* (see Definition 1.4.31). Suppose that the upper integral

$$\overline{\int}_X f\, d\mu := \inf_{g \geq f; g \text{ simple}} \int_X g\, d\mu$$

and lower integral

$$\underline{\int}_X f\, d\mu := \sup_{h \leq f; h \text{ simple}} \int_X h\, d\mu$$

agree. Show that f is measurable. (This is a converse to Exercise 1.3.11.)

We will continue to see the monotone convergence theorem, Fatou's lemma, and the dominated convergence theorem make an appearance throughout the rest of this text (and in *An epsilon of room, Vol. I*).

1.5. Modes of convergence

If one has a sequence $x_1, x_2, x_3, \ldots \in \mathbf{R}$ of real numbers x_n, it is unambiguous what it means for that sequence to converge to a limit $x \in \mathbf{R}$: it means that for every $\varepsilon > 0$, there exists an N such that $|x_n - x| \leq \varepsilon$ for all $n > N$. Similarly for a sequence $z_1, z_2, z_3, \ldots \in \mathbf{C}$ of complex numbers z_n converging to a limit $z \in \mathbf{C}$.

More generally, if one has a sequence $v_1, v_2, v_3, \ldots$ of d-dimensional vectors v_n in a real vector space $\mathbf{R}^d$ or complex vector space $\mathbf{C}^d$, it is also unambiguous what it means for that sequence to converge to a limit $v \in \mathbf{R}^d$ or $v \in \mathbf{C}^d$; it means that for every $\varepsilon > 0$, there exists an N such that $\|v_n - v\| \leq \varepsilon$ for all $n \geq N$. Here, the norm $\|v\|$ of a vector $v = (v^{(1)}, \ldots, v^{(d)})$ can be chosen to be the Euclidean norm $\|v\|_2 := (\sum_{j=1}^d (v^{(j)})^2)^{1/2}$, the supremum norm $\|v\|_\infty := \sup_{1 \leq j \leq d} |v^{(j)}|$, or any other number of norms, but for the purposes of convergence, these norms are all *equivalent*; a sequence of vectors converges in the Euclidean norm if and only if it converges in the supremum norm, and similarly for any other two norms on the finite-dimensional space $\mathbf{R}^d$ or $\mathbf{C}^d$.

If however one has a sequence $f_1, f_2, f_3, \ldots$ of functions $f_n \colon X \to \mathbf{R}$ or $f_n \colon X \to \mathbf{C}$ on a common domain X, and a putative limit $f \colon X \to \mathbf{R}$ or $f \colon X \to \mathbf{C}$, there can now be many different ways in which the sequence f_n may or may not converge to the limit f. (One could also consider convergence of functions $f_n \colon X_n \to \mathbf{C}$ on different domains X_n, but we will not discuss this issue at all here.) This is contrast with the situation with scalars x_n or z_n (which corresponds to the case when X is a single point) or vectors v_n (which corresponds to the case when X is a finite set such as $\{1, \ldots, d\}$). Once X becomes infinite, the functions f_n acquire an infinite number of degrees of freedom, and this allows them to approach f in any number of inequivalent ways.

What different types of convergence are there? As an undergraduate, one learns of the following two basic *modes of convergence*:

(i) We say that f_n converges to f *pointwise* if, for every $x \in X$, $f_n(x)$ converges to $f(x)$. In other words, for every $\varepsilon > 0$ and $x \in X$, there exists N (that depends on *both* ε and x) such that $|f_n(x) - f(x)| \leq \varepsilon$ whenever $n \geq N$.

(ii) We say that f_n converges to f *uniformly* if, for every $\varepsilon > 0$, there exists N such that for every $n \geq N$, $|f_n(x) - f(x)| \leq \varepsilon$ for every $x \in X$. The difference between uniform convergence and pointwise convergence is that with the former, the time N at which $f_n(x)$ must be permanently ε-close to $f(x)$ is not permitted to depend on x, but must instead be chosen uniformly in x.

Uniform convergence implies pointwise convergence, but not conversely. A typical example: the functions $f_n \colon \mathbf{R} \to \mathbf{R}$ defined by $f_n(x) := x/n$ converge pointwise to the zero function $f(x) := 0$, but not uniformly.

However, pointwise and uniform convergence are only two of dozens of many other modes of convergence that are of importance in analysis. We will not attempt to exhaustively enumerate these modes here (but see §1.9 of *An epsilon of room, Vol. I*). We will, however, discuss some of the modes of convergence that arise from measure theory, when the domain X is equipped with the structure of a measure space $(X, \mathcal{B}, \mu)$, and the functions f_n (and their limit f) are measurable with respect to this space. In this context, we have some additional modes of convergence:

(i) We say that f_n converges to f *pointwise almost everywhere* if, for (μ-)almost everywhere $x \in X$, $f_n(x)$ converges to $f(x)$.

(ii) We say that f_n converges to f *uniformly almost everywhere*, *essentially uniformly*, or *in L^∞ norm* if, for every $\varepsilon > 0$, there exists N such that for every $n \geq N$, $|f_n(x) - f(x)| \leq \varepsilon$ for μ-almost every $x \in X$.

(iii) We say that f_n converges to f *almost uniformly* if, for every $\varepsilon > 0$, there exists an exceptional set $E \in \mathcal{B}$ of measure $\mu(E) \leq \varepsilon$ such that f_n converges uniformly to f on the complement of E.

(iv) We say that f_n converges to f *in L^1 norm* if the quantity $\|f_n - f\|_{L^1(\mu)} = \int_X |f_n(x) - f(x)|\, d\mu$ converges to 0 as $n \to \infty$.

(v) We say that f_n converges to f *in measure* if, for every $\varepsilon > 0$, the measures $\mu(\{x \in X : |f_n(x) - f(x)| \geq \varepsilon\})$ converge to zero as $n \to \infty$.

Observe that each of these five modes of convergence is unaffected if one modifies f_n or f on a set of measure zero. In contrast, the pointwise and uniform modes of convergence can be affected if one modifies f_n or f even on a single point. The L^1 and L^∞ modes of converges are special cases of the L^p mode of convergence, which is discussed in §1.3 of *An epsilon of room, Vol. I.*

Remark 1.5.1. In the context of probability theory (see Section 2.3), in which f_n and f are interpreted as random variables, convergence in L^1 norm is often referred to as *convergence in mean*, pointwise convergence almost everywhere is often referred to as *almost sure convergence*, and convergence in measure is often referred to as *convergence in probability*.

Exercise 1.5.1 (Linearity of convergence)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space, let $f_n, g_n \colon X \to \mathbf{C}$ be sequences of measurable functions, and let $f, g \colon X \to \mathbf{C}$ be measurable functions.

(i) Show that f_n converges to f along one of the above seven modes of convergence if and only if $|f_n - f|$ converges to 0 along the same mode.

(ii) If f_n converges to f along one of the above seven modes of convergence, and g_n converges to g along the same mode, show that $f_n + g_n$ converges to $f + g$ along the same mode, and that cf_n converges to cf along the same mode for any $c \in \mathbf{C}$.

(iii) (Squeeze test) If f_n converges to 0 along one of the above seven modes, and $|g_n| \leq f_n$ pointwise for each n, show that g_n converges to 0 along the same mode.

We have some easy implications between modes:

Exercise 1.5.2 (Easy implications)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f_n \colon X \to \mathbf{C}$ and $f \colon X \to \mathbf{C}$ be measurable functions.

(i) If f_n converges to f uniformly, then f_n converges to f pointwise.

(ii) If f_n converges to f uniformly, then f_n converges to f in L^∞ norm. Conversely, if f_n converges to f in L^∞ norm, then f_n converges to f uniformly outside of a null set (i.e. there exists a null set E such that the restriction $f_n \downharpoonright_{X \backslash E}$ of f_n to the complement of E converges to the restriction $f \downharpoonright_{X \backslash E}$ of f).

(iii) If f_n converges to f in L^∞ norm, then f_n converges to f almost uniformly.

(iv) If f_n converges to f almost uniformly, then f_n converges to f pointwise almost everywhere.

(v) If f_n converges to f pointwise, then f_n converges to f pointwise almost everywhere.

(vi) If f_n converges to f in L^1 norm, then f_n converges to f in measure.

(vii) If f_n converges to f almost uniformly, then f_n converges to f in measure.

The reader is encouraged to draw a diagram that summarises the logical implications between the seven modes of convergence that the above exercise describes.

We give four key examples that distinguish between these modes, in the case when X is the real line $\mathbf{R}$ with Lebesgue measure. The first three of these examples already were introduced in Section 1.4, but the fourth is new, and also important.

Example 1.5.2 (Escape to horizontal infinity)**.** Let $f_n := 1_{[n,n+1]}$. Then f_n converges to zero pointwise (and thus, pointwise almost everywhere), but not uniformly, in L^∞ norm, almost uniformly, in L^1 norm, or in measure.

Example 1.5.3 (Escape to width infinity)**.** Let $f_n := \frac{1}{n}1_{[0,n]}$. Then f_n converges to zero uniformly (and thus, pointwise, pointwise almost everywhere, in L^∞ norm, almost uniformly, and in measure), but not in L^1 norm.

Example 1.5.4 (Escape to vertical infinity)**.** Let $f_n := n1_{[\frac{1}{n},\frac{2}{n}]}$. Then f_n converges to zero pointwise (and thus, pointwise almost everywhere) and almost uniformly (and hence in measure), but not uniformly, in L^∞ norm, or in L^1 norm.

Example 1.5.5 (Typewriter sequence)**.** Let f_n be defined by the formula

$$f_n := 1_{[\frac{n-2^k}{2^k},\frac{n-2^k+1}{2^k}]}$$

whenever $k \geq 0$ and $2^k \leq n < 2^{k+1}$. This is a sequence of indicator functions of intervals of decreasing length, marching across the unit interval $[0,1]$ over and over again. Then f_n converges to zero in measure and in L^1 norm, but not pointwise almost everywhere (and hence also not pointwise, not almost uniformly, nor in L^∞ norm, nor uniformly).

Remark 1.5.6. The L^∞ *norm* $\|f\|_{L^\infty(\mu)}$ of a measurable function $f: X \to \mathbf{C}$ is defined to the infimum of all the quantities $M \in [0, +\infty]$ that are *essential upper bounds* for f in the sense that $|f(x)| \leq M$ for almost every x. Then f_n converges to f in L^∞ norm if and only if $\|f_n - f\|_{L^\infty(\mu)} \to 0$ as $n \to \infty$. The L^∞ and L^1 norms are part of the larger family of L^p *norms*, studied in §1.3 of *An epsilon of room, Vol. I*.

One particular advantage of L^1 convergence is that, in the case when the f_n are absolutely integrable, it implies convergence of the integrals,

$$\int_X f_n \, d\mu \to \int_X f \, d\mu,$$

as one sees from the triangle inequality. Unfortunately, none of the other modes of convergence automatically imply this convergence of the integral, as the above examples show.

The purpose of these notes is to compare these modes of convergence with each other. Unfortunately, the relationship between these modes is not particularly simple; unlike the situation with pointwise and uniform convergence, one cannot simply rank these modes in a linear order from strongest to weakest. This is ultimately because the different modes react in different ways to the three "escape to infinity" scenarios described above, as well as to the "typewriter" behaviour when a single set is "overwritten" many times. On the other hand, if one imposes some additional assumptions to shut down one or more of these escape to infinity scenarios, such as a finite measure hypothesis $\mu(X) < \infty$ or a *uniform integrability hypothesis*, then one can obtain some additional implications between the different modes.

1.5.1. Uniqueness. Throughout these notes, $(X, \mathcal{B}, \mu)$ denotes a measure space. We abbreviate "μ-almost everywhere" as "almost everywhere" throughout.

Even though the modes of convergence all differ from each other, they are all *compatible* in the sense that they never disagree about *which* function f a sequence of functions f_n converges to, outside of a set of measure zero. More precisely:

Proposition 1.5.7. *Let $f_n \colon X \to \mathbf{C}$ be a sequence of measurable functions, and let $f, g \colon X \to \mathbf{C}$ be two additional measurable functions. Suppose that f_n converges to f along one of the seven modes of convergence defined above, and f_n converges to g along another of the seven modes of convergence (or perhaps the same mode of convergence as for f). Then f and g agree almost everywhere.*

Note that the conclusion is the best one can hope for in the case of the last five modes of convergence, since as remarked earlier, these modes of convergence are unaffected if one modifies f or g on a set of measure zero.

Proof. In view of Exercise 1.5.2, we may assume that f_n converges to f either pointwise almost everywhere, or in measure, and similarly that f_n converges to g either pointwise almost everywhere, or in measure.

Suppose first that f_n converges to both f and g pointwise almost everywhere. Then by Exercise 1.5.1, 0 converges to $f - g$ pointwise almost everywhere, which clearly implies that $f - g$ is zero almost everywhere, and the claim follows. A similar argument applies if f_n converges to both f and g in measure.

By symmetry, the only remaining case that needs to be considered is when f_n converges to f pointwise almost everywhere, and f_n converges to g in measure. We need to show that $f = g$ almost everywhere. It suffices to show that for every $\varepsilon > 0$, that $|f(x) - g(x)| \leq \varepsilon$ for almost every x, as the claim then follows by setting $\varepsilon = 1/m$ for $m = 1, 2, 3, \ldots$ and using the fact that the countable union of null sets is again a null set.

Fix $\varepsilon > 0$, and let $A := \{x \in X : |f(x) - g(x)| > \varepsilon\}$. This is a measurable set; our task is to show that it has measure zero. Suppose for contradiction that $\mu(A) > 0$. We consider the sets

$$A_N := \{x \in A : |f_n(x) - f(x)| \leq \varepsilon/2 \text{ for all } n \geq N\}.$$

These are measurable sets that are increasing in N. As f_n converges to f almost everywhere, we see that almost every $x \in A$ belongs to at least one of the A_N, thus $\bigcup_{N=1}^{\infty} A_N$ is equal to A outside of a null set. In particular,

$$\mu(\bigcup_{N=1}^{\infty} A_N) > 0.$$

Applying monotone convergence for sets, we conclude that

$$\mu(A_N) > 0$$

for some finite N. But by the triangle inequality, we have $|f_n(x) - g(x)| > \varepsilon/2$ for all $x \in A_N$ and all $n \geq N$. As a consequence, f_n cannot converge in measure to g, which gives the desired contradiction. $\square$

1.5.2. The case of a step function. One way to appreciate the distinctions between the above modes of convergence is to focus on the case when $f = 0$, and when each of the f_n is a *step function*, by which we mean a constant multiple $f_n = A_n 1_{E_n}$ of a measurable set E_n. For simplicity we will assume that the $A_n > 0$ are positive reals, and that the E_n have a positive measure $\mu(E_n) > 0$. We also assume the A_n exhibit one of two modes of behaviour: either the A_n converge to zero, or else they are bounded away from zero (i.e. there exists $c > 0$ such that $A_n \geq c$ for every n. It is easy to see that if a sequence A_n does not converge to zero, then it has a subsequence that is bounded away from zero, so it does not cause too much loss of generality to restrict to one of these two cases.

Given such a sequence $f_n = A_n 1_{E_n}$ of step functions, we now ask, for each of the seven modes of convergence, what it means for this sequence to converge to zero along that mode. It turns out that the answer to question is controlled more or less completely by the following three quantities:

(i) The *height* A_n of the n^{th} function f_n;
(ii) The *width* $\mu(E_n)$ of the n^{th} function f_n; and
(iii) The N^{th} *tail support* $E^*_N := \bigcup_{n \geq N} E_n$ of the sequence $f_1, f_2, f_3, \ldots$.

Indeed, we have:

Exercise 1.5.3 (Convergence for step functions)**.** Let the notation and assumptions be as above. Establish the following claims:

(i) f_n converges uniformly to zero if and only if $A_n \to 0$ as $n \to \infty$.

(ii) f_n converges in L^∞ norm to zero if and only if $A_n \to 0$ as $n \to \infty$.

(iii) f_n converges almost uniformly to zero if and only if $A_n \to 0$ as $n \to \infty$, or $\mu(E_N^*) \to 0$ as $N \to \infty$.

(iv) f_n converges pointwise to zero if and only if $A_n \to 0$ as $n \to \infty$, or $\bigcap_{N=1}^\infty E_N^* = \emptyset$.

(v) f_n converges pointwise almost everywhere to zero if and only if $A_n \to 0$ as $n \to \infty$, or $\bigcap_{N=1}^\infty E_N^*$ is a null set.

(vi) f_n converges in measure to zero if and only if $A_n \to 0$ as $n \to \infty$, or $\mu(E_n) \to 0$ as $n \to \infty$.

(vii) f_n converges in L^1 norm if and only if $A_n \mu(E_n) \to 0$ as $n \to \infty$.

To put it more informally: when the height goes to zero, then one has convergence to zero in all modes except possibly for L^1 convergence, which requires that the product of the height and the width goes to zero. If instead the height is bounded away from zero and the width is positive, then we never have uniform or L^∞ convergence, but we have convergence in measure if the width goes to zero, we have almost uniform convergence if the tail support (which has larger measure than the width) has measure that goes to zero, we have pointwise almost everywhere convergence if the tail support shrinks to a null set, and pointwise convergence if the tail support shrinks to the empty set.

It is instructive to compare this exercise with Exercise 1.5.2, or with the four examples given in the introduction. In particular:

(i) In the escape to horizontal infinity scenario, the height and width do not shrink to zero, but the tail set shrinks to the empty set (while remaining of infinite measure throughout).

(ii) In the escape to width infinity scenario, the height goes to zero, but the width (and tail support) go to infinity, causing the L^1 norm to stay bounded away from zero.

(iii) In the escape to vertical infinity, the height goes to infinity, but the width (and tail support) go to zero (or the empty set), causing the L^1 norm to stay bounded away from zero.

(iv) In the typewriter example, the width goes to zero, but the height and the tail support stay fixed (and thus bounded away from zero).

Remark 1.5.8. The monotone convergence theorem (Theorem 1.4.44) can also be specialised to this case. Observe that the $f_n = A_n 1_{E_n}$ are monotone increasing if and only if $A_n \leq A_{n+1}$ and $E_n \subset E_{n+1}$ for each n. In such cases, observe that the f_n converge pointwise to $f := A1_E$, where $A := \lim_{n\to\infty} A_n$ and $E := \bigcup_{n=1}^\infty E_n$. The monotone convergence theorem then asserts that $A_n\mu(E_n) \to A\mu(E)$ as $n \to \infty$, which is a consequence of the monotone convergence theorem $\mu(E_n) \to \mu(E)$ for sets.

1.5.3. Finite measure spaces. The situation simplifies somewhat if the space X has finite measure (and in particular, in the case when $(X, \mathcal{B}, \mu)$ is a *probability space*, see Section 2.3). This shuts down two of the four examples (namely, escape to horizontal infinity or width infinity) and creates a few more equivalences. Indeed, from Egorov's theorem (Exercise 1.4.31), we now have

Theorem 1.5.9 (Egorov's theorem, again)**.** *Let X have finite measure, and let $f_n : X \to \mathbf{C}$ and $f : X \to \mathbf{C}$ be measurable functions. Then f_n converges to f pointwise almost everywhere if and only if f_n converges to f almost uniformly.*

Note that when one specialises to step functions using Exercise 1.5.3, then Egorov's theorem collapses to the downward monotone convergence property for sets (Exercise 1.4.23(iii)).

Another nice feature of the finite measure case is that L^∞ convergence implies L^1 convergence:

Exercise 1.5.4. Let X have finite measure, and let $f_n : X \to \mathbf{C}$ and $f : X \to \mathbf{C}$ be measurable functions. Show that if f_n converges to f in L^∞ norm, then f_n also converges to f in L^1 norm.

1.5.4. Fast convergence. The typewriter example shows that L^1 convergence is not strong enough to force almost uniform or pointwise almost everywhere convergence. However, this can be rectified if one assumes that the L^1 convergence is sufficiently fast:

Exercise 1.5.5 (Fast L^1 convergence)**.** Suppose that $f_n, f: X \to \mathbf{C}$ are measurable functions such that $\sum_{n=1}^\infty \|f_n - f\|_{L^1(\mu)} < \infty$; thus, not only do the quantities $\|f_n - f\|_{L^1(\mu)}$ go to zero (which would mean L^1 convergence), but they converge in an absolutely summable fashion.

(i) Show that f_n converges pointwise almost everywhere to f.

(ii) Show that f_n converges almost uniformly to f.

(*Hint:* If you have trouble getting started, try working first in the special case in which $f_n = A_n 1_{E_n}$ are step functions and $f = 0$ and use Exercise 1.5.3 in order to gain some intuition. The second part of the exercise implies the first, but the first is a little easier to prove and may thus serve as a useful warmup. The $\varepsilon/2^n$ trick may come in handy for the second part.)

As a corollary, we see that L^1 convergence implies almost uniform or pointwise almost everywhere convergence if we are allowed to pass to a subsequence:

Corollary 1.5.10. *Suppose that $f_n: X \to \mathbf{C}$ are a sequence of measurable functions that converge in L^1 norm to a limit f. Then there exists a subsequence f_{n_j} that converges almost uniformly (and hence, pointwise almost everywhere) to f (while remaining convergent in L^1 norm, of course).*

Proof. Since $\|f_n - f\|_{L^1(\mu)} \to 0$ as $n \to \infty$, we can select $n_1 < n_2 < n_3 < \ldots$ such that $\|f_{n_j} - f\|_{L^1(\mu)} \leq 2^{-j}$ (say). This is enough for the previous exercise to apply. □

Actually, one can strengthen this corollary a bit by relaxing L^1 convergence to convergence in measure:

Exercise 1.5.6. Suppose that $f_n: X \to \mathbf{C}$ are a sequence of measurable functions that converge in measure to a limit f. Then there exists a subsequence f_{n_j} that converges almost uniformly (and hence, pointwise almost everywhere) to f. (*Hint:* Choose the n_j so that the sets $\{x \in X : |f_{n_j}(x) - f(x)| > 1/j\}$ have a suitably small measure.)

It is instructive to see how this subsequence is extracted in the case of the typewriter sequence. In general, one can view the operation of passing to a subsequence as being able to eliminate "typewriter" situations in which the tail support is much larger than the width.

Exercise 1.5.7. Let $(X, \mathcal{B}, \mu)$ be a measure space, let $f_n \colon X \to \mathbf{C}$ be a sequence of measurable functions converging pointwise almost everywhere as $n \to \infty$ to a measurable limit $f \colon X \to \mathbf{C}$, and for each n, let $f_{n,m} \colon X \to \mathbf{C}$ be a sequence of measurable functions converging pointwise almost everywhere as $m \to \infty$ (keeping n fixed) to f_n.

(i) If $\mu(X)$ is finite, show that there exists a sequence $m_1, m_2, \ldots$ such that f_{n,m_n} converges pointwise almost everywhere to f.

(ii) Show the same claim is true if, instead of assuming that $\mu(X)$ is finite, we merely assume that X is *σ-finite*, i.e. it is the countable union of sets of finite measure.

(The claim can fail if X is not σ-finite. A counterexample is if $X = \mathbf{N}^{\mathbf{N}}$ with counting measure, f_n and f are identically zero for all $n \in \mathbf{N}$, and $f_{n,m}$ is the indicator function the space of all sequences $(a_i)_{i \in \mathbf{N}} \in \mathbf{N}^{\mathbf{N}}$ with $a_n \geq m$.)

Exercise 1.5.8. Let $f_n \colon X \to \mathbf{C}$ be a sequence of measurable functions, and let $f \colon X \to \mathbf{C}$ be another measurable function. Show that the following are equivalent:

(i) f_n converges in measure to f.

(ii) Every subsequence f_{n_j} of the f_n has a further subsequence $f_{n_{j_i}}$ that converges almost uniformly to f.

1.5.5. Domination and uniform integrability. Now we turn to the reverse question, of whether almost uniform convergence, pointwise almost everywhere convergence, or convergence in measure can imply L^1 convergence. The escape to vertical and width infinity examples shows that without any further hypotheses, the answer to this question is no. However, one can do better if one places some *domination* hypotheses on the f_n that shut down both of these escape routes.

We say that a sequence $f_n\colon X \to \mathbf{C}$ is *dominated* if there exists an absolutely integrable function $g\colon X \to \mathbf{C}$ such that $|f_n(x)| \leq g(x)$ for all n and almost every x. For instance, if X has finite measure and the f_n are uniformly bounded, then they are dominated. Observe that the sequences in the vertical and width escape to infinity examples are *not* dominated (why?).

The dominated convergence theorem (Theorem 1.4.49) then asserts that if f_n converges to f pointwise almost everywhere, then it necessarily converges to f in L^1 norm (and hence also in measure). Here is a variant:

Exercise 1.5.9. Suppose that $f_n\colon X \to \mathbf{C}$ are a dominated sequence of measurable functions, and let $f\colon X \to \mathbf{C}$ be another measurable function. Show that f_n converges in L^1 norm to f if and only if f_n converges in measure to f. (*Hint:* one way to establish the "if" direction is first show that every subsequence of the f_n has a further subsequence that converges in L^1 to f, using Exercise 1.5.6 and the dominated convergence theorem (Theorem 1.4.49). Alternatively, use monotone convergence to find a set E of finite measure such that $\int_{X\backslash E} g\ d\mu$, and hence $\int_{X\backslash E} f_n\ d\mu$ and $\int_{X\backslash E} f\ d\mu$, are small.)

There is a more general notion than domination, known as *uniform integrability*, which serves as a substitute for domination in many (but not all) contexts.

Definition 1.5.11 (Uniform integrability)**.** A sequence $f_n\colon X \to \mathbf{C}$ of absolutely integrable functions is said to be *uniformly integrable* if the following three statements hold:

(i) (Uniform bound on L^1 norm) One has $\sup_n \|f_n\|_{L^1(\mu)} = \sup_n \int_X |f_n|\ d\mu < +\infty$.

(ii) (No escape to vertical infinity) One has $\sup_n \int_{|f_n|\geq M} |f_n|\ d\mu \to 0$ as $M \to +\infty$.

(iii) (No escape to width infinity) One has $\sup_n \int_{|f_n|\leq \delta} |f_n|\ d\mu \to 0$ as $\delta \to 0$.

Remark 1.5.12. It is instructive to understand uniform integrability in the step function case $f_n = A_n 1_{E_n}$. The uniform bound on the

L^1 norm then asserts that $A_n\mu(E_n)$ stays bounded. The lack of escape to vertical infinity means that along any subsequence for which $A_n \to \infty$, $A_n\mu(E_n)$ must go to zero. Similarly, the lack of escape to width infinity means that along any subsequence for which $A_n \to 0$, $A_n\mu(E_n)$ must go to zero.

Exercise 1.5.10. (i) Show that if f is an absolutely integrable function, then the constant sequence $f_n = f$ is uniformly integrable. (*Hint:* use the monotone convergence theorem.)

(ii) Show that every dominated sequence of measurable functions is uniformly integrable.

(iii) Give an example of a sequence that is uniformly integrable but not dominated.

In the case of a finite measure space, there is no escape to width infinity, and the criterion for uniform integrability simplifies to just that of excluding vertical infinity:

Exercise 1.5.11. Suppose that X has finite measure, and let $f_n \colon X \to \mathbf{C}$ be a sequence of measurable functions. Show that f_n is uniformly integrable if and only if $\sup_n \int_{|f_n| \geq M} |f_n|\ d\mu \to 0$ as $M \to +\infty$.

Exercise 1.5.12 (Uniform L^p bound on finite measure implies uniform integrability)**.** Suppose that X have finite measure, let $1 < p < \infty$, an d suppose that $f_n \colon X \to \mathbf{C}$ is a sequence of measurable functions such that $\sup_n \int_X |f_n|^p\ d\mu < \infty$. Show that the sequence f_n is uniformly integrable.

Exercise 1.5.13. Let $f_n \colon X \to \mathbf{C}$ be a uniformly integrable sequence of functions. Show that for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\int_E |f_n|\ d\mu \leq \varepsilon$$

whenever $n \geq 1$ and E is a measurable set with $\mu(E) \leq \delta$.

Exercise 1.5.14. This exercise is a partial converse to Exercise 1.5.13. Let X be a probability space, and let $f_n \colon X \to \mathbf{C}$ be a sequence of absolutely integrable functions with $\sup_n \|f_n\|_{L^1} < \infty$. Suppose that for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\int_E |f_n|\ d\mu \leq \varepsilon$$

whenever $n \geq 1$ and E is a measurable set with $\mu(E) \leq \delta$. Show that the sequence f_n is uniformly integrable.

The dominated convergence theorem (Theorem 1.4.49) does not have an analogue in the uniformly integrable setting:

Exercise 1.5.15. Give an example of a sequence f_n of uniformly integrable functions that converge pointwise almost everywhere to zero, but do *not* converge almost uniformly, in measure, or in L^1 norm.

However, one *does* have an analogue of Exercise 1.5.9:

Theorem 1.5.13 (Uniformly integrable convergence in measure)**.** *Let $f_n \colon X \to \mathbf{C}$ be a uniformly integrable sequence of functions, and let $f \colon X \to \mathbf{C}$ be another function. Then f_n converges in L^1 norm to f if and only if f_n converges to f in measure.*

Proof. The "only if" part follows from Exercise 1.5.2, so we establish the "if" part.

By uniform integrability, there exists a finite $A > 0$ such that

$$\int_X |f_n| \, d\mu \leq A$$

for all n. By Exercise 1.5.6, there is a subsequence of the f_n that converges pointwise almost everywhere to f. Applying Fatou's lemma (Corollary1.4.47), we conclude that

$$\int_X |f| \, d\mu \leq A,$$

thus f is absolutely integrable.

Now let $\varepsilon > 0$ be arbitrary. By uniform integrability, one can find $\delta > 0$ such that

$$\int_{|f_n| \leq \delta} |f_n| \, d\mu \leq \varepsilon \tag{1.15}$$

for all n. By monotone convergence, and decreasing δ if necessary, we may say the same for f, thus

$$\int_{|f| \leq \delta} |f| \, d\mu \leq \varepsilon. \tag{1.16}$$

Let $0 < \kappa < \delta/2$ be another small quantity (that can depend on A, ε, δ) that we will choose a bit later. From (1.15), (1.16) and the hypothesis $\kappa < \delta/2$ we have

$$\int_{|f_n - f| < \kappa; |f| \leq \delta/2} |f_n| \, d\mu \leq \varepsilon$$

and

$$\int_{|f_n - f| < \kappa; |f| \leq \delta/2} |f| \, d\mu \leq \varepsilon$$

and hence by the triangle inequality

$$\int_{|f_n - f| < \kappa; |f| \leq \delta/2} |f - f_n| \, d\mu \leq 2\varepsilon. \tag{1.17}$$

Finally, from Markov's inequality (Exercise 1.4.36(vi)) we have

$$\mu(\{x : |f(x)| > \delta/2\}) \leq \frac{A}{\delta/2}$$

and thus

$$\int_{|f_n - f| < \kappa; |f| > \delta/2} |f - f_n| \, d\mu \leq \varepsilon \leq \frac{A}{\delta/2}\kappa.$$

In particular, by shrinking κ further if necessary we see that

$$\int_{|f_n - f| < \kappa; |f| > \delta/2} |f - f_n| \, d\mu \leq \varepsilon$$

and hence by (1.17)

$$\int_{|f_n - f| < \kappa} |f - f_n| \, d\mu \leq 3\varepsilon \tag{1.18}$$

for all n.

Meanwhile, since f_n converges in measure to f, we know that there exists an N (depending on κ) such that

$$\mu(|f_n(x) - f(x)| \geq \kappa) \leq \kappa$$

for all $n \geq N$. Applying Exercise 1.5.13, we conclude (making κ smaller if necessary) that

$$\int_{|f_n - f| \geq \kappa} |f_n| \, d\mu \leq \varepsilon$$

and

$$\int_{|f_n - f| \geq \kappa} |f| \, d\mu \leq \varepsilon$$

and hence by the triangle inequality

$$\int_{|f_n - f| \geq \kappa} |f - f_n| \; d\mu \leq 2\varepsilon$$

for all $n \geq N$. Combining this with (1.18) we conclude that

$$\|f_n - f\|_{L^1(\mu)} = \int_X |f - f_n| \; d\mu \leq 5\varepsilon$$

for all $n \geq N$, and so f_n converges to f in L^1 norm as desired. □

Finally, we recall two results from the previous notes for unsigned functions.

Exercise 1.5.16 (Monotone convergence theorem)**.** Suppose that $f_n \colon X \to [0, +\infty)$ are measurable, monotone non-decreasing in n and are such that $\sup_n \int_X f_n \; d\mu < \infty$. Show that f_n converges in L^1 norm to $\sup_n f_n$. (Note that $\sup_n f_n$ can be infinite on a null set, but the definition of L^1 convergence can be easily modified to accomodate this.)

Exercise 1.5.17 (Defect version of Fatou's lemma)**.** Suppose that $f_n \colon X \to [0, +\infty)$ are measurable, are such that $\sup_n \int_X f_n \; d\mu < \infty$, and converge pointwise almost everywhere to some measurable limit $f \colon X \to [0, +\infty)$. Show that f_n converges in L^1 norm to f if and only if $\int_X f_n \; d\mu$ converges to $\int_X f \; d\mu$. Informally, we see that in the unsigned, bounded mass case, pointwise convergence implies L^1 norm convergence if and only if there is no loss of mass.

Exercise 1.5.18. Suppose that $f_n \colon X \to \mathbf{C}$ are a dominated sequence of measurable functions, and let $f \colon X \to \mathbf{C}$ be another measurable function. Show that f_n converges pointwise almost everywhere to f if and only if f_n converges in almost uniformly to f.

Exercise 1.5.19. Let X be a probability space (see Section 2.3). Given any real-valued measurable function $f \colon X \to \mathbf{R}$, we define the *cumulative distribution function* $F \colon \mathbf{R} \to [0, 1]$ of f to be the function $F(\lambda) := \mu(\{x \in X : f(x) \leq \lambda\})$. Given another sequence $f_n \colon X \to \mathbf{R}$ of real-valued measurable functions, we say that f_n *converges in distribution* to f if the cumulative distribution function $F_n(\lambda)$ of f_n

converges pointwise to the cumulative distribution function $F(\lambda)$ of f at all $\lambda \in \mathbf{R}$ for which F is continuous.

(i) Show that if f_n converges to f in any of the seven senses discussed above (uniformly, essentially uniformly, almost uniformly pointwise, pointwise almost everywhere, in L^1, or in measure), then it converges in distribution to f.

(ii) Give an example in which f_n converges to f in distribution, but not in any of the above seven senses.

(iii) Show that convergence in distribution is not linear, in the sense that if f_n converges to f in distribution, and g_n converges to g, then $f_n + g_n$ need not converge to $f + g$.

(iv) Show that a sequence f_n can converge in distribution to two different limits f, g, which are not equal almost everywhere.

Convergence in distribution (not to be confused with convergence in the sense of *distributions*, which is studied in S 1.13 of *An epsilon of room, Vol. I* is commonly used in probability; but, as the above exercise demonstrates, it is quite a weak notion of convergence, lacking many of the properties of the modes of convergence discussed here.

1.6. Differentiation theorems

Let $[a, b]$ be a compact interval of positive length (thus $-\infty < a < b < +\infty$). Recall that a function $F\colon [a, b] \to \mathbf{R}$ is said to be *differentiable* at a point $x \in [a, b]$ if the limit

$$F'(x) := \lim_{y \to x; y \in [a,b] \backslash \{x\}} \frac{F(y) - F(x)}{y - x} \tag{1.19}$$

exists. In that case, we call $F'(x)$ the *strong derivative*, *classical derivative*, or just *derivative* for short, of F at x. We say that F is *everywhere differentiable*, or *differentiable* for short, if it is differentiable at all points $x \in [a, b]$, and *differentiable almost everywhere* if it is differentiable at almost every point $x \in [a, b]$. If F is differentiable everywhere and its derivative F' is continuous, then we say that F is *continuously differentiable*.

Remark 1.6.1. In §1.13 of *An epsilon of room, Vol. I*, the notion of a *weak derivative* or *distributional derivative* is introduced. This type of derivative can be applied to a much rougher class of functions and is in many ways more suitable than the classical derivative for doing "Lebesgue" type analysis (i.e. analysis centred around the Lebesgue integral, and in particular allowing functions to be uncontrolled, infinite, or even undefined on sets of measure zero). However, for now we will stick with the classical approach to differentiation.

Exercise 1.6.1. If $F\colon [a,b] \to \mathbf{R}$ is everywhere differentiable, show that F is continuous and F' is measurable. If F is almost everywhere differentiable, show that the (almost everywhere defined) function F' is measurable (i.e. it is equal to an everywhere defined measurable function on $[a,b]$ outside of a null set), but give an example to demonstrate that F need not be continuous.

Exercise 1.6.2. Give an example of a function $F\colon [a,b] \to \mathbf{R}$ which is everywhere differentiable, but not continuously differentiable. (*Hint:* choose an F that vanishes quickly at some point, say at the origin 0, but which also oscillates rapidly near that point.)

In single-variable calculus, the operations of integration and differentiation are connected by a number of basic theorems, starting with *Rolle's theorem*.

Theorem 1.6.2 (Rolle's theorem)**.** *Let $[a,b]$ be a compact interval of positive length, and let $F\colon [a,b] \to \mathbf{R}$ be a differentiable function such that $F(a) = F(b)$. Then there exists $x \in (a,b)$ such that $F'(x) = 0$.*

Proof. By subtracting a constant from F (which does not affect differentiability or the derivative) we may assume that $F(a) = F(b) = 0$. If F is identically zero then the claim is trivial, so assume that F is non-zero somewhere. By replacing F with $-F$ if necessary, we may assume that F is positive somewhere, thus $\sup_{x\in[a,b]} F(x) > 0$. On the other hand, as F is continuous and $[a,b]$ is compact, F must attain its maximum somewhere, thus there exists $x \in [a,b]$ such that $F(x) \geq F(y)$ for all $y \in [a,b]$. Then $F(x)$ must be positive and so x cannot equal either a or b, and thus must lie in the interior. From the right limit of (1.19) we see that $F'(x) \leq 0$, while from the left limit we have $F'(x) \geq 0$. Thus $F'(x) = 0$ and the claim follows. □

Remark 1.6.3. Observe that the same proof also works if F is only differentiable in the interior (a, b) of the interval $[a, b]$, so long as it is continuous all the way up to the boundary of $[a, b]$.

Exercise 1.6.3. Give an example to show that Rolle's theorem can fail if f is merely assumed to be almost everywhere differentiable, even if one adds the additional hypothesis that f is continuous. This example illustrates that everywhere differentiability is a significantly stronger property than almost everywhere differentiability. We will see further evidence of this fact later in these notes; there are many theorems that assert in their conclusion that a function is almost everywhere differentiable, but few that manage to conclude *everywhere* differentiability.

Remark 1.6.4. It is important to note that Rolle's theorem only works in the real scalar case when F is real-valued, as it relies heavily on the least upper bound property for the domain $\mathbf{R}$. If, for instance, we consider complex-valued scalar functions $F\colon [a, b] \to \mathbf{C}$, then the theorem can fail; for instance, the function $F\colon [0, 1] \to \mathbf{C}$ defined by $F(x) := e^{2\pi ix} - 1$ vanishes at both endpoints and is differentiable, but its derivative $F'(x) = 2\pi i e^{2\pi ix}$ is never zero. (Rolle's theorem does imply that the real and imaginary parts of the derivative F' both vanish somewhere, but the problem is that they don't *simultaneously* vanish at the same point.) Similar remarks to functions taking values in a finite-dimensional vector space, such as $\mathbf{R}^n$.

One can easily amplify Rolle's theorem to the *mean value theorem*:

Corollary 1.6.5 (Mean value theorem)**.** *Let $[a, b]$ be a compact interval of positive length, and let $F\colon [a, b] \to \mathbf{R}$ be a differentiable function. Then there exists $x \in (a, b)$ such that $F'(x) = \frac{F(b)-F(a)}{b-a}$.*

Proof. Apply Rolle's theorem to the function $x \mapsto F(x) - \frac{F(b)-F(a)}{b-a}(x - a)$. □

Remark 1.6.6. As Rolle's theorem is only applicable to real scalar-valued functions, the more general mean value theorem is also only applicable to such functions.

Exercise 1.6.4 (Uniqueness of antiderivatives up to constants)**.** Let $[a, b]$ be a compact interval of positive length, and let $F\colon [a, b] \to \mathbf{R}$ and $G\colon [a, b] \to \mathbf{R}$ be differentiable functions. Show that $F'(x) = G'(x)$ for every $x \in [a, b]$ if and only if $F(x) = G(x) + C$ for some constant $C \in \mathbf{R}$ and all $x \in [a, b]$.

We can use the mean value theorem to deduce one of the fundamental theorems of calculus:

Theorem 1.6.7 (Second fundamental theorem of calculus)**.** *Let $F\colon [a, b] \to \mathbf{R}$ be a differentiable function, such that F' is Riemann integrable. Then the Riemann integral $\int_a^b F'(x)\ dx$ of F' is equal to $F(b) - F(a)$. In particular, we have $\int_a^b F'(x)\ dx = F(b) - F(a)$ whenever F is continuously differentiable.*

Proof. Let $\varepsilon > 0$. By the definition of Riemann integrability, there exists a finite partition $a = t_0 < t_1 < \ldots < t_k = b$ such that

$$|\sum_{j=1}^{k} F'(t_j^*)(t_j - t_{j-1}) - \int_a^b F'(x)| \leq \varepsilon$$

for every choice of $t_j^* \in [t_{j-1}, t_j]$.

Fix this partition. From the mean value theorem, for each $1 \leq j \leq k$ one can find $t_j^* \in [t_{j-1}, t_j]$ such that

$$F'(t_j^*)(t_j - t_{j-1}) = F(t_j) - F(t_{j-1})$$

and thus by telescoping series

$$|(F(b) - F(a)) - \int_a^b F'(x)| \leq \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, the claim follows. □

Remark 1.6.8. Even though the mean value theorem only holds for real scalar functions, the fundamental theorem of calculus holds for complex or vector-valued functions, as one can simply apply that theorem to each component of that function separately.

Of course, we also have the other half of the fundamental theorem of calculus:

Theorem 1.6.9 (First fundamental theorem of calculus)**.** *Let $[a, b]$ be a compact interval of positive length. Let $f\colon [a, b] \to \mathbf{C}$ be a continuous function, and let $F\colon [a, b] \to \mathbf{C}$ be the indefinite integral $F(x) := \int_a^x f(t)\ dt$. Then F is differentiable on $[a, b]$, with derivative $F'(x) = f(x)$ for all $x \in [a, b]$. In particular, F is continuously differentiable.*

Proof. It suffices to show that

$$\lim_{h\to 0^+} \frac{F(x+h) - F(x)}{h} = f(x)$$

for all $x \in [a, b)$, and

$$\lim_{h\to 0^-} \frac{F(x+h) - F(x)}{h} = f(x)$$

for all $x \in (a, b]$. After a change of variables, we can write

$$\frac{F(x+h) - F(x)}{h} = \int_0^1 f(x+ht)\ dt$$

for any $x \in [a, b)$ and any sufficiently small $h > 0$, or any $x \in (a, b]$ and any sufficiently small $h < 0$. As f is continuous, the function $t \mapsto f(x+ht)$ converges uniformly to $f(x)$ on $[0, 1]$ as $h \to 0$ (keeping x fixed). As the interval $[0, 1]$ is bounded, $\int_0^1 f(x+ht)\ dt$ thus converges to $\int_0^1 f(x)\ dt = f(x)$, and the claim follows. □

Corollary 1.6.10 (Differentiation theorem for continuous functions)**.** *Let $f\colon [a, b] \to \mathbf{C}$ be a continuous function on a compact interval. Then we have*

$$\lim_{h\to 0^+} \frac{1}{h} \int_{[x,x+h]} f(t)\ dt = f(x)$$

for all $x \in [a, b)$,

$$\lim_{h\to 0^+} \frac{1}{h} \int_{[x-h,x]} f(t)\ dt = f(x)$$

for all $x \in (a, b]$, and thus

$$\lim_{h\to 0^+} \frac{1}{2h} \int_{[x-h,x+h]} f(t)\ dt = f(x)$$

for all $x \in (a, b)$.

In these notes we explore the question of the extent to which these theorems continue to hold when the differentiability or integrability conditions on the various functions F, F', f are relaxed. Among the results proven in these notes are

(i) The *Lebesgue differentiation theorem*, which roughly speaking asserts that Corollary 1.6.10 continues to hold for *almost* every x if f is merely absolutely integrable, rather than continuous;

(ii) A number of *differentiation theorems*, which assert for instance that monotone, Lipschitz, or bounded variation functions in one dimension are almost everywhere differentiable; and

(iii) The second fundamental theorem of calculus for *absolutely continuous* functions.

1.6.1. The Lebesgue differentiation theorem in one dimension.

The main objective of this section is to show

Theorem 1.6.11 (Lebesgue differentiation theorem, one-dimensional case)**.** *Let $f\colon \mathbf{R} \to \mathbf{C}$ be an absolutely integrable function, and let $F\colon \mathbf{R} \to \mathbf{C}$ be the definite integral $F(x) := \int_{[-\infty,x]} f(t)\ dt$. Then F is continuous and almost everywhere differentiable, and $F'(x) = f(x)$ for almost every $x \in \mathbf{R}$.*

This can be viewed as a variant of Corollary 1.6.10; the hypotheses are weaker because f is only assumed to be absolutely integrable, rather than continuous (and can live on the entire real line, and not just on a compact interval); but the conclusion is weaker too, because F is only found to be almost everywhere differentiable, rather than everywhere differentiable. (But such a relaxation of the conclusion is necessary at this level of generality; consider for instance the example when $f = 1_{[0,1]}$.)

The continuity is an easy exercise:

Exercise 1.6.5. Let $f\colon \mathbf{R} \to \mathbf{C}$ be an absolutely integrable function, and let $F\colon \mathbf{R} \to \mathbf{C}$ be the definite integral $F(x) := \int_{[-\infty,x]} f(t)\ dt$. Show that F is continuous.

The main difficulty is to show that $F'(x) = f(x)$ for almost every $x \in \mathbf{R}$. This will follow from

Theorem 1.6.12 (Lebesgue differentiation theorem, second formulation)**.** *Let $f\colon \mathbf{R} \to \mathbf{C}$ be an absolutely integrable function. Then*

$$\lim_{h\to 0^+} \frac{1}{h}\int_{[x,x+h]} f(t)\ dt = f(x) \tag{1.20}$$

for almost every $x \in \mathbf{R}$, and

$$\lim_{h\to 0^+} \frac{1}{h}\int_{[x-h,x]} f(t)\ dt = f(x) \tag{1.21}$$

for almost every $x \in \mathbf{R}$.

Exercise 1.6.6. Show that Theorem 1.6.11 follows from Theorem 1.6.12.

We will just prove the first fact (1.20); the second fact (1.21) is similar (or can be deduced from (1.20) by replacing f with the reflected function $x \mapsto f(-x)$.

We are taking f to be complex valued, but it is clear from taking real and imaginary parts that it suffices to prove the claim when f is real-valued, and we shall thus assume this for the rest of the argument.

The conclusion (1.20) we want to prove is a convergence theorem - an assertion that for all functions f in a given class (in this case, the class of absolutely integrable functions $f\colon \mathbf{R} \to \mathbf{R}$), a certain sequence of linear expressions $T_h f$ (in this case, the right averages $T_h f(x) = \frac{1}{h}\int_{[x,x+h]} f(t)\ dt$) converge in some sense (in this case, pointwise almost everywhere) to a specified limit (in this case, f). There is a general and very useful argument to prove such convergence theorems, known as the *density argument*. This argument requires two ingredients, which we state informally as follows:

(i) A verification of the convergence result for some "dense subclass" of "nice" functions f, such as continuous functions, smooth functions, simple functions, etc.. By "dense", we mean that a general function f in the original class can be approximated to arbitrary accuracy in a suitable sense by a function in the nice subclass.

(ii) A quantitative estimate that upper bounds the maximal fluctuation of the linear expressions $T_h f$ in terms of the "size" of the function f (where the precise definition of "size" depends on the nature of the approximation in the first ingredient).

Once one has these two ingredients, it is usually not too hard to put them together to obtain the desired convergence theorem for general functions f (not just those in the dense subclass). We illustrate this with a simple example:

Proposition 1.6.13 (Translation is continuous in L^1)**.** *Let $f\colon \mathbf{R}^d \to \mathbf{C}$ be an absolutely integrable function, and for each $h \in \mathbf{R}^d$, let $f_h\colon \mathbf{R}^d \to \mathbf{C}$ be the shifted function*

$$f_h(x) := f(x-h).$$

Then f_h converges in L^1 norm to f as $h \to 0$, thus

$$\lim_{h\to 0} \int_{\mathbf{R}^d} |f_h(x) - f(x)|\, dx = 0.$$

Proof. We first verify this claim for a dense subclass of f, namely the functions f which are continuous and compactly supported (i.e. they vanish outside of a compact set). Such functions are continuous, and thus f_h converges uniformly to f as $h \to 0$. Furthermore, as f is compactly supported, the support of $f_h - f$ stays uniformly bounded for h in a bounded set. From this we see that f_h also converges to f in L^1 norm as required.

Next, we observe the quantitative estimate

$$\int_{\mathbf{R}^d} |f_h(x) - f(x)|\, dx \leq 2 \int_{\mathbf{R}^d} |f(x)|\, dx \tag{1.22}$$

for any $h \in \mathbf{R}^d$. This follows easily from the triangle inequality

$$\int_{\mathbf{R}^d} |f_h(x) - f(x)|\, dx \leq \int_{\mathbf{R}^d} |f_h(x)|\, dx + \int_{\mathbf{R}^d} |f(x)|\, dx$$

together with the translation invariance of the Lebesgue integral:

$$\int_{\mathbf{R}^d} |f_h(x)|\, dx = \int_{\mathbf{R}^d} |f(x)|\, dx.$$

Now we put the two ingredients together. Let $f \colon \mathbf{R}^d \to \mathbf{C}$ be absolutely integrable, and let $\varepsilon > 0$ be arbitrary. Applying Littlewood's second principle (Theorem 1.3.20(iii)) to the absolutely integrable function F', we can find a continuous, compactly supported function $g \colon \mathbf{R}^d \to \mathbf{C}$ such that

$$\int_{\mathbf{R}^d} |f(x) - g(x)|\ dx \le \varepsilon.$$

Applying (1.22), we conclude that

$$\int_{\mathbf{R}^d} |(f-g)_h(x) - (f-g)(x)|\ dx \le 2\varepsilon,$$

which we rearrange as

$$\int_{\mathbf{R}^d} |(f_h - f)_h(x) - (g_h - g)(x)|\ dx \le 2\varepsilon.$$

By the dense subclass result, we also know that

$$\int_{\mathbf{R}^d} |g_h(x) - g(x)|\ dx \le \varepsilon$$

for all h sufficiently close to zero. From the triangle inequality, we conclude that

$$\int_{\mathbf{R}^d} |f_h(x) - f(x)|\ dx \le 3\varepsilon$$

for all h sufficiently close to zero, and the claim follows. □

Remark 1.6.14. In the above application of the density argument, we proved the required quantitative estimate directly for all functions f in the original class of functions. However, it is also possible to use the density argument a second time and initially verify the quantitative estimate just for functions f in a nice subclass (e.g. continuous functions of compact support). In many cases, one can then extend that estimate to the general case by using tools such as *Fatou's lemma* (Corollary1.4.47), which are particularly suited for showing that upper bound estimates are preserved with respect to limits.

Exercise 1.6.7. Let $f \colon \mathbf{R}^d \to \mathbf{C}$, $g \colon \mathbf{R}^d \to \mathbf{C}$ be Lebesgue measurable functions such that f is absolutely integrable and g is essentially bounded (i.e. bounded outside of a null set). Show that the *convolution* $f * g \colon \mathbf{R}^d \to \mathbf{C}$ defined by the formula

$$f * g(x) = \int_{\mathbf{R}^d} f(y)g(x-y)\ dy$$

is well-defined (in the sense that the integrand on the right-hand side is absolutely integrable) and that $f * g$ is a bounded, continuous function.

The above exercise is illustrative of a more general intuition, which is that convolutions tend to be *smoothing* in nature; the convolution $f * g$ of two functions is usually at least as regular as, and often more regular than, either of the two factors f, g.

This smoothing phenomenon gives rise to an important fact, namely the *Steinhaus theorem*:

Exercise 1.6.8 (Steinhaus theorem)**.** Let $E \subset \mathbf{R}^d$ be a Lebesgue measurable set of positive measure. Show that the set $E - E := \{x - y : x, y \in E\}$ contains an open neighbourhood of the origin. (*Hint:* reduce to the case when E is bounded, and then apply the previous exercise to the convolution $1_E * 1_{-E}$, where $-E := \{-y : y \in E\}$.)

Exercise 1.6.9. A *homomorphism* $f \colon \mathbf{R}^d \to \mathbf{C}$ is a map with the property that $f(x + y) = f(x) + f(y)$ for all $x, y \in \mathbf{R}^d$.

(i) Show that all measurable homomorphisms are continuous. (*Hint:* for any disk D centered at the origin in the complex plane, show that $f^{-1}(z + D)$ has positive measure for at least one $z \in \mathbf{C}$, and then use the Steinhaus theorem from the previous exercise.)

- Show that f is a measurable homomorphism if and only if it takes the form $f(x_1, \ldots, x_d) = x_1 z_1 + \ldots + x_d z_d$ for all $x_1, \ldots, x_d \in \mathbf{R}$ and some complex coefficients $z_1, \ldots, z_d$. (*Hint:* first establish this for rational $x_1, \ldots, x_d$, and then use the previous part of this exercise.)

(ii) (For readers familiar with *Zorn's lemma*, see §2.4 of *An epsilon of room, Vol. I*) Show that there exist homomorphisms $f \colon \mathbf{R}^d \to \mathbf{C}$ which are not of the form in the previous exercise. (*Hint:* view $\mathbf{R}^d$ (or $\mathbf{C}$) as a vector space over the rationals $\mathbf{Q}$, and use the fact (from Zorn's lemma) that every vector space - even an infinite-dimensional one - has at least one basis.) This gives an alternate construction of a non-measurable set to that given in previous notes.

Remark 1.6.15. One drawback with the density argument is it gives convergence results which are *qualitative* rather than quantitative - there is no explicit bound on the rate of convergence. For instance, in Proposition 1.6.13, we know that for any $\varepsilon > 0$, there exists $\delta > 0$ such that $\int_{\mathbf{R}^d} |f_h(x) - f(x)|\, dx \leq \varepsilon$ whenever $|h| \leq \delta$, but we do not know exactly how δ depends on ε and f. Actually, the proof does eventually give such a bound, but it depends on "how measurable" the function f is, or more precisely how "easy" it is to approximate f by a "nice" function. To illustrate this issue, let's work in one dimension and consider the function $f(x) := \sin(Nx)1_{[0,2\pi]}(x)$, where $N \geq 1$ is a large integer. On the one hand, f is bounded in the L^1 norm uniformly in N: $\int_{\mathbf{R}} |f(x)|\, dx \leq 2\pi$ (indeed, the left-hand side is equal to 2). On the other hand, it is not hard to see that $\int_{\mathbf{R}} |f_{\pi/N}(x) - f(x)|\, dx \geq c$ for some absolute constant $c > 0$. Thus, if one force $\int_{\mathbf{R}} |f_h(x) - f(x)|\, dx$ to drop below c, one has to make h at most π/N from the origin. Making N large, we thus see that the rate of convergence of $\int_{\mathbf{R}} |f_h(x) - f(x)|\, dx$ to zero can be arbitrarily slow, even though f is bounded in L^1. The problem is that as N gets large, it becomes increasingly difficult to approximate f well by a "nice" function, by which we mean a uniformly continuous function with a reasonable *modulus of continuity*, due to the increasingly oscillatory nature of f. See [**Ta2008**, §1.4] for some further discussion of this issue, and what quantitative substitutes are available for such qualitative results.

Now we return to the Lebesgue differentiation theorem, and apply the density argument. The dense subclass result is already contained in Corollary 1.6.10, which asserts that (1.20) holds for all continuous functions f. The quantitative estimate we will need is the following special case of the *Hardy-Littlewood maximal inequality*:

Lemma 1.6.16 (One-sided Hardy-Littlewood maximal inequality)**.** *Let $f\colon \mathbf{R} \to \mathbf{C}$ be an absolutely integrable function, and let $\lambda > 0$.*

Then

$$m(\{x \in \mathbf{R} : \sup_{h>0} \frac{1}{h} \int_{[x,x+h]} |f(t)| \; dt \geq \lambda\}) \leq \frac{1}{\lambda} \int_{\mathbf{R}} |f(t)| \; dt.$$

We will prove this lemma shortly, but let us first see how this, combined with the dense subclass result, will give the Lebesgue differentiation theorem. Let $f\colon \mathbf{R} \to \mathbf{C}$ be absolutely integrable, and let $\varepsilon, \lambda > 0$ be arbitrary. Then by Littlewood's second principle, we can find a function $g\colon \mathbf{R} \to \mathbf{C}$ which is continuous and compactly supported, with

$$\int_{\mathbf{R}} |f(x) - g(x)| \; dx \leq \varepsilon.$$

Applying the one-sided Hardy-Littlewood maximal inequality, we conclude that

$$m(\{x \in \mathbf{R} : \sup_{h>0} \frac{1}{h} \int_{[x,x+h]} |f(t) - g(t)| \; dt \geq \lambda\}) \leq \frac{\varepsilon}{\lambda}.$$

In a similar spirit, from Markov's inequality (Lemma 1.3.15) we have

$$m(\{x \in \mathbf{R} : |f(x) - g(x)| \geq \lambda\}) \leq \frac{\varepsilon}{\lambda}.$$

By subadditivity, we conclude that for all $x \in \mathbf{R}$ outside of a set E of measure at most $2\varepsilon/\lambda$, one has both

$$\frac{1}{h} \int_{[x,x+h]} |f(t) - g(t)| \; dt < \lambda \tag{1.23}$$

and

$$|f(x) - g(x)| < \lambda \tag{1.24}$$

for all $h > 0$.

Now let $x \in \mathbf{R} \backslash E$. From the dense subclass result (Corollary 1.6.10) applied to the continuous function g, we have

$$|\frac{1}{h} \int_{[x,x+h]} g(t) \; dt - g(x)| < \lambda$$

whenever h is sufficiently close to x. Combining this with (1.23), (1.24), and the triangle inequality, we conclude that

$$|\frac{1}{h} \int_{[x,x+h]} f(t) \; dt - f(x)| < 3\lambda$$

for all h sufficiently close to zero. In particular we have

$$\limsup_{h\to 0} |\frac{1}{h}\int_{[x,x+h]} f(t)\ dt - f(x)| < 3\lambda$$

for all x outside of a set of measure $2\varepsilon/\lambda$. Keeping λ fixed and sending ε to zero, we conclude that

$$\limsup_{h\to 0} |\frac{1}{h}\int_{[x,x+h]} f(t)\ dt - f(x)| < 3\lambda$$

for almost every $x \in \mathbf{R}$. If we then let λ go to zero along a countable sequence (e.g. $\lambda := 1/n$ for $n = 1, 2, \ldots$), we conclude that

$$\limsup_{h\to 0} |\frac{1}{h}\int_{[x,x+h]} f(t)\ dt - f(x)| = 0$$

for almost every $x \in \mathbf{R}$, and the claim follows.

The only remaining task is to establish the one-sided Hardy-Littlewood maximal inequality. We will do so by using the *rising sun lemma*:

Lemma 1.6.17 (Rising sun lemma)**.** *Let $[a,b]$ be a compact interval, and let $F\colon [a,b] \to \mathbf{R}$ be a continuous function. Then one can find an at most countable family of disjoint non-empty open intervals $I_n = (a_n, b_n)$ in $[a,b]$ with the following properties:*

(i) *For each n, either $F(a_n) = F(b_n)$, or else $a_n = a$ and $F(b_n) \geq F(a_n)$.*
(ii) *If $x \in [a,b]$ does not lie in any of the intervals I_n, then one must have $F(y) \leq F(x)$ for all $x \leq y \leq b$.*

Remark 1.6.18. To explain the name "rising sun lemma", imagine the graph $\{(x, F(x)) : x \in [a,b]\}$ of F as depicting a hilly landscape, with the sun shining horizontally from the rightward infinity $(+\infty, 0)$ (or rising from the east, if you will). Those x for which $F(y) \leq F(x)$ are the locations on the landscape which are illuminated by the sun. The intervals I_n then represent the portions of the landscape that are in shadow. The reader is encouraged to draw a picture[14] to illustrate this perspective.

[14] Author's note: I have deliberately omitted including such pictures in the text, as I feel that it is far more instructive and useful for the reader to directly create a personalised visual aid for these results.

This lemma is proven using the following basic fact:

Exercise 1.6.10. Show that any open subset U of $\mathbf{R}$ can be written as the union of at most countably many disjoint non-empty open intervals, whose endpoints lie outside of U. (*Hint:* first show that every x in U is contained in a maximal open subinterval (a, b) of U, and that these maximal open subintervals are disjoint, with each such interval containing at least one rational number.)

Proof. (Proof of rising sun lemma) Let U be the set of all $x \in (a, b)$ such that $F(y) > F(x)$ for at least one $x < y < b$. As F is continuous, U is open, and so U is the union of at most countably many disjoint non-empty open intervals $I_n = (a_n, b_n)$, with the endpoints a_n, b_n lying outside of U.

The second conclusion of the rising sun lemma is clear from construction, so it suffices to establish the first. Suppose first that $I_n = (a_n, b_n)$ is such that $a_n \neq a$. As the endpoint a_n does not lie in U, we must have $F(y) \leq F(a_n)$ for all $a_n \leq y \leq b$; similarly we have $F(y) \leq F(b_n)$ for all $b_n \leq y \leq b$. In particular we have $F(b_n) \leq F(a_n)$. By the continuity of F, it will then suffice to show that $F(b_n) \geq F(t)$ for all $a_n < t < b_n$.

Suppose for contradiction that there was $a_n < t < b_n$ with $F(b_n) < F(t)$. Let $A := \{s \in [t, b] : F(s) \geq F(t)\}$, then A is a closed set that contains t but not b. Set $t_* := \sup(A)$, then $t_* \in [t, b) \subset I_n \subset U$, and thus there exists $t_* < y \leq b$ such that $F(y) > F(t_*)$. Since $F(t_*) \geq F(t) > F(b_n)$, and $F(b_n) \geq F(z)$ for all $b_n \leq z \leq b$, we see that y cannot exceed b_n, and thus lies in A, but this contradicts the fact that t_* is the supremum of A.

The case when $a_n = a$ is similar and is left to the reader; the only difference is that we can no longer assert that $F(y) \leq F(a_n)$ for all $a_n \leq y \leq b$, and so do not have the upper bound $F(b_n) \leq F(a_n)$. □

Now we can prove the one-sided Hardy-Littlewood maximal inequality. By upwards monotonicity, it will suffice to show that

$$m(\{x \in [a, b] : \sup_{h>0; [x,x+h] \subset [a,b]} \frac{1}{h} \int_{[x,x+h]} |f(t)|\, dt \geq \lambda\}) \leq \frac{1}{\lambda} \int_{\mathbf{R}} |f(t)|\, dt$$

for any compact interval $[a,b]$. By modifying λ by an epsilon, we may replace the non-strict inequality here with strict inequality:
(1.25)
$$m(\{x \in [a,b] : \sup_{h>0; [x,x+h] \subset [a,b]} \frac{1}{h} \int_{[x,x+h]} |f(t)|\, dt > \lambda\}) \leq \frac{1}{\lambda} \int_{\mathbf{R}} |f(t)|\, dt$$

Fix $[a,b]$. We apply the rising sun lemma to the function $F\colon [a,b] \to \mathbf{R}$ defined as
$$F(x) := \int_{[a,x]} |f(t)|\ dt - (x-a)\lambda.$$
By Lemma 1.6.5, F is continuous, and so we can find an at most countable sequence of intervals $I_n = (a_n, b_n)$ with the properties given by the rising sun lemma. From the second property of that lemma, we observe that
$$\{x \in [a,b] : \sup_{h>0; [x,x+h] \subset [a,b]} \frac{1}{h} \int_{[x,x+h]} |f(t)|\ dt > \lambda\} \subset \bigcup_n I_n,$$
since the property $\frac{1}{h} \int_{[x,x+h]} |f(t)|\ dt > \lambda$ can be rearranged as $F(x+h) > F(x)$. By countable additivity, we may thus upper bound the left-hand side of (1.25) by $\sum_n (b_n - a_n)$. On the other hand, since $F(b_n) - F(a_n) \geq 0$, we have
$$\int_{I_n} |f(t)|\ dt \geq \lambda(b_n - a_n)$$
and thus
$$\sum_n (b_n - a_n) \leq \frac{1}{\lambda} \sum_n \int_{I_n} |f(t)|\ dt.$$
As the I_n are disjoint intervals in I, we may apply monotone convergence and monotonicity to conclude that
$$\sum_n \int_{I_n} |f(t)|\ dt \leq \int_{[a,b]} |f(t)|\ dt,$$
and the claim follows.

Exercise 1.6.11 (Two-sided Hardy-Littlewood maximal inequality)**.** Let $f\colon \mathbf{R} \to \mathbf{C}$ be an absolutely integrable function, and let $\lambda > 0$. Show that
$$m(\{x \in \mathbf{R} : \sup_{x \in I} \frac{1}{|I|} \int_I |f(t)|\ dt \geq \lambda\}) \leq \frac{2}{\lambda} \int_{\mathbf{R}} |f(t)|\ dt,$$

where the supremum ranges over all intervals I of positive length that contain x.

Exercise 1.6.12 (Rising sun inequality)**.** Let $f\colon \mathbf{R} \to \mathbf{R}$ be an absolutely integrable function, and let $f^*\colon \mathbf{R} \to \mathbf{R}$ be the one-sided signed Hardy-Littlewood maximal function

$$f^*(x) := \sup_{h>0} \frac{1}{h} \int_{[x,x+h]} f(t)\ dt.$$

Establish the *rising sun inequality*

$$\lambda m(\{f^*(x) > \lambda\}) \leq \int_{x: f^*(x)>\lambda} f(x)\ dx$$

for all real λ (note here that we permit λ to be zero or negative), and show that this inequality implies Lemma 1.6.16. (*Hint:* First do the $\lambda = 0$ case, by invoking the rising sun lemma.) See [**Ta2009**, §2.9] for some further discussion of inequalities of this type, and applications to ergodic theory (and in particular the *maximal ergodic theorem*).

Exercise 1.6.13. Show that the left and right-hand sides in Lemma 1.6.16 are in fact equal. (*Hint:* one may first wish to try this in the case when f has compact support, in which case one can apply the rising sun lemma to a sufficiently large interval containing the support of f.)

1.6.2. The Lebesgue differentiation theorem in higher dimensions. Now we extend the Lebesgue differentiation theorem to higher dimensions. Theorem 1.6.11 does not have an obvious high-dimensional analogue, but Theorem 1.6.12 does:

Theorem 1.6.19 (Lebesgue differentiation theorem in general dimension)**.** *Let $f\colon \mathbf{R}^d \to \mathbf{C}$ be an absolutely integrable function. Then for almost every $x \in \mathbf{R}^d$, one has*

$$\lim_{r\to 0} \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y) - f(x)|\ dy = 0 \tag{1.26}$$

and

$$\lim_{r\to 0} \frac{1}{m(B(x,r))} \int_{B(x,r)} f(y)\ dy = f(x),$$

where $B(x,r) := \{y \in \mathbf{R}^d : |x-y| < r\}$ *is the open ball of radius* r *centred at* x*.*

From the triangle inequality we see that

$$|\frac{1}{m(B(x,r))}\int_{B(x,r)} f(y)\ dy - f(x)| = |\frac{1}{m(B(x,r))}\int_{B(x,r)} f(y) - f(x)\ dy|$$
$$\leq \frac{1}{m(B(x,r))}\int_{B(x,r)} |f(y) - f(x)|\ dy,$$

so we see that the first conclusion of Theorem 1.6.19 implies the second. A point x for which (1.26) holds is called a *Lebesgue point* of f; thus, for an absolutely integrable function f, almost every point in $\mathbf{R}^d$ will be a Lebesgue point for $\mathbf{R}^d$.

Exercise 1.6.14. Call a function $f\colon \mathbf{R}^d \to \mathbf{C}$ *locally integrable* if, for every $x \in \mathbf{R}^d$, there exists an open neighbourhood of x on which f is absolutely integrable.

(i) Show that f is locally integrable if and only if $\int_{B(0,r)} |f(x)|\ dx < \infty$ for all $r > 0$.

(ii) Show that Theorem 1.6.19 implies a generalisation of itself in which the condition of absolute integrability of f is weakened to local integrability.

Exercise 1.6.15. For each $h > 0$, let E_h be a subset of $B(0,h)$ with the property that $m(E_h) \geq cm(B(0,h))$ for some $c > 0$ independent of h. Show that if $f\colon \mathbf{R}^d \to \mathbf{C}$ is locally integrable, and x is a Lebesgue point of f, then

$$\lim_{h\to 0} \frac{1}{m(E_h)}\int_{x+E_h} f(y)\ dy = f(x).$$

Conclude that Theorem 1.6.19 implies Theorem 1.6.12.

To prove Theorem 1.6.19, we use the density argument. The dense subclass case is easy:

Exercise 1.6.16. Show that Theorem 1.6.19 holds whenever f is continuous.

The quantitative estimate needed is the following:

Theorem 1.6.20 (Hardy-Littlewood maximal inequality)**.** *Let $f : \mathbf{R}^d \to \mathbf{C}$ be an absolutely integrable function, and let $\lambda > 0$. Then*

$$m(\{x \in \mathbf{R}^d : \sup_{r>0} \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y)|\ dy \geq \lambda\}) \leq \frac{C_d}{\lambda} \int_{\mathbf{R}} |f(t)|\ dt$$

for some constant $C_d > 0$ depending only on d.

Remark 1.6.21. The expression $\sup_{r>0} \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y)|\ dy \geq \lambda\}$ is known as the *Hardy-Littlewood maximal function* of f, and is often denoted $Mf(x)$. It is an important function in the field of (real-variable) *harmonic analysis*.

Exercise 1.6.17. Use the density argument to show that Theorem 1.6.20 implies Theorem 1.6.19.

In the one-dimensional case, this estimate was established via the rising sun lemma. Unfortunately, that lemma relied heavily on the ordered nature of $\mathbf{R}$, and does not have an obvious analogue in higher dimensions. Instead, we will use the following covering lemma. Given an open ball $B = B(x,r)$ in $\mathbf{R}^d$ and a real number $c > 0$, we write $cB := B(x, cr)$ for the ball with the same centre as B, but c times the radius. (Note that this is slightly different from the set $c \cdot B := \{cy : y \in B\}$ - why?) Note that $|cB| = c^d|B|$ for any open ball $B \subset \mathbf{R}^d$ and any $c > 0$.

Lemma 1.6.22 (Vitali-type covering lemma)**.** *Let $B_1, \ldots, B_n$ be a finite collection of open balls in $\mathbf{R}^d$ (not necessarily disjoint). Then there exists a subcollection $B'_1, \ldots, B'_m$ of* disjoint *balls in this collection, such that*

$$\bigcup_{i=1}^n B_i \subset \bigcup_{j=1}^m 3B'_j. \tag{1.27}$$

In particular, by finite subadditivity,

$$m(\bigcup_{i=1}^n B_i) \leq 3^d \sum_{j=1}^m m(B'_j).$$

Proof. We use a *greedy algorithm* argument, selecting the balls B'_i to be as large as possible while remaining disjoint. More precisely, we run the following algorithm:

Step 0. Initialise $m = 0$ (so that, initially, there are no balls $B'_1, \ldots, B'_m$ in the desired collection).

Step 1. Consider all the balls B_j that do not already intersect one of the $B'_1, \ldots, B'_m$ (so, initially, all of the balls $B_1, \ldots, B_n$ will be considered). If there are no such balls, STOP. Otherwise, go on to Step 2.

Step 2. Locate the largest ball B_j that does not already intersect one of the $B'_1, \ldots, B'_m$. (If there are multiple largest balls with exactly the same radius, break the tie arbitrarily.) Add this ball to the collection $B'_1, \ldots, B'_m$ by setting $B'_{m+1} := B_j$ and then incrementing m to $m+1$. Then return to Step 1.

Note that at each iteration of this algorithm, the number of available balls amongst the $B_1, \ldots, B_n$ drops by at least one (since each ball selected certainly intersects itself and so cannot be selected again). So this algorithm terminates in finite time. It is also clear from construction that the $B'_1, \ldots, B'_m$ are a subcollection of the $B_1, \ldots, B_n$ consisting of disjoint balls. So the only task remaining is to verify that (1.27) holds at the completion of the algorithm, i.e. to show that each ball B_i in the original collection is covered by the triples $3B'_j$ of the subcollection.

For this, we argue as follows. Take any ball B_i in the original collection. Because the algorithm only halts when there are no more balls that are disjoint from the $B'_1, \ldots, B'_m$, the ball B_i must intersect at least one of the balls B'_j in the subcollection. Let B'_j be the first ball with this property, thus B_i is disjoint from $B'_1, \ldots, B'_{j-1}$, but intersects B'_j. Because B'_j was chosen to be largest amongst all balls that did not intersect $B'_1, \ldots, B'_{j-1}$, we conclude that the radius of B_i cannot exceed that of B'_j. From the triangle inequality, this implies that $B_i \subset 3B'_j$, and the claim follows. □

Exercise 1.6.18. Technically speaking, the above algorithmic argument was not phrased in the standard language of formal mathematical deduction, because in that language, any mathematical object (such as the natural number m) can only be defined once, and not redefined multiple times as is done in most algorithms. Rewrite the above argument in a way that avoids redefining any variable.

(*Hint:* introduce a "time" variable t, and recursively construct families $B'_{1,t}, \dots, B'_{m_t,t}$ of balls that represent the outcome of the above algorithm after t iterations (or t_* iterations, if the algorithm halted at some previous time $t_* < t$). For this particular algorithm, there are also more *ad hoc* approaches that exploit the relatively simple nature of the algorithm to allow for a less notationally complicated construction.) More generally, it is possible to use this time parameter trick to convert any construction involving a provably terminating algorithm into a construction that does not redefine any variable. (It is however dangerous to work with any algorithm that has an infinite run time, unless one has a suitably strong convergence result for the algorithm that allows one to take limits, either in the classical sense or in the more general sense of jumping to *limit ordinals*; in the latter case, one needs to use *transfinite induction* in order to ensure that the use of such algorithms is rigorous; see §2.4 of *An epsilon of room, Vol. I.*)

Remark 1.6.23. The actual *Vitali covering lemma*[**Vi1908**] is slightly different to this one, but we will not need it here. Actually there is a family of related covering lemmas which are useful for a variety of tasks in harmonic analysis, see for instance [**deG1981**] for further discussion.

Now we can prove the Hardy-Littlewood inequality, which we will do with the constant $C_d := 3^d$. It suffices to verify the claim with strict inequality,

$$m(\{x \in \mathbf{R}^d : \sup_{r>0} \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y)|\, dy > \lambda\}) \leq \frac{C_d}{\lambda} \int_{\mathbf{R}} |f(t)|\, dt$$

as the non-strict case then follows by perturbing λ slightly and then taking limits.

Fix f and λ. By inner regularity, it suffices to show that

$$m(K) \leq \frac{3^d}{\lambda} \int_{\mathbf{R}} |f(t)|\, dt$$

whenever K is a compact set that is contained in

$$\{x \in \mathbf{R}^d : \sup_{r>0} \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y)|\, dy > \lambda\}.$$

By construction, for every $x \in K$, there exists an open ball $B(x,r)$ such that

$$\frac{1}{m(B(x,r))}\int_{B(x,r)} |f(y)|\,dy > \lambda. \tag{1.28}$$

By compactness of K, we can cover K by a finite number $B_1,\dots,B_n$ of such balls. Applying the Vitali-type covering lemma, we can find a subcollection $B_1',\dots,B_m'$ of disjoint balls such that

$$m(\bigcup_{i=1}^n B_i) \le 3^d \sum_{j=1}^m m(B_j').$$

By (1.28), on each ball B_j' we have

$$m(B_j') < \frac{1}{\lambda}\int_{B_j'} |f(y)|\,dy;$$

summing in j and using the disjointness of the B_j' we conclude that

$$m(\bigcup_{i=1}^n B_i) \le \frac{3^d}{\lambda}\int_{\mathbf{R}^d} |f(y)|\,dy.$$

Since the $B_1,\dots,B_n$ cover K, we obtain Theorem 1.6.20 as desired.

Exercise 1.6.19. Improve the constant 3^d in the Hardy-Littlewood maximal inequality to 2^d. (*Hint:* observe that with the construction used to prove the Vitali covering lemma, the *centres* of the balls B_i are contained in $\bigcup_{j=1}^m 2B_j'$ and not just in $\bigcup_{j=1}^m 3B_j'$. To exploit this observation one may need to first create an epsilon of room, as the centers are not by themselves sufficient to cover the required set.)

Remark 1.6.24. The optimal value of C_d is not known in general, although a fairly recent result of Melas[**Me2003**] gives the surprising conclusion that the optimal value of C_1 is $C_1 = \frac{11+\sqrt{61}}{12} = 1.56\ldots$. It is known that C_d grows at most linearly in d, thanks to a result of Stein and Strömberg[**StSt1983**], but it is not known if C_d is bounded in d or grows as $d \to \infty$.

Exercise 1.6.20 (Dyadic maximal inequality)**.** If $f\colon \mathbf{R}^d \to \mathbf{C}$ is an absolutely integrable function, establish the dyadic Hardy-Littlewood maximal inequality

$$m(\{x \in \mathbf{R}^d : \sup_{x\in Q}\frac{1}{|Q|}\int_Q |f(y)|\,dy \ge \lambda\}) \le \frac{1}{\lambda}\int_{\mathbf{R}} |f(t)|\,dt$$

where the supremum ranges over all dyadic cubes Q that contain x. (*Hint:* the nesting property of dyadic cubes will be useful when it comes to the covering lemma stage of the argument, much as it was in Exercise 1.1.14.)

Exercise 1.6.21 (Besicovich covering lemma in one dimension)**.** Let $I_1, \ldots, I_n$ be a finite family of open intervals in $\mathbf{R}$ (not necessarily disjoint). Show that there exist a subfamily $I'_1, \ldots, I'_m$ of intervals such that

(i) $\bigcup_{i=1}^n I_n = \bigcup_{j=1}^m I'_m$; and

(ii) Each point $x \in \mathbf{R}$ is contained in at most two of the I'_m.

(*Hint:* First refine the family of intervals so that no interval I_i is contained in the union of the the other intervals. At that point, show that it is no longer possible for a point to be contained in three of the intervals.) There is a variant of this lemma that holds in higher dimensions, known as the *Besicovitch covering lemma*.

Exercise 1.6.22. Let μ be a Borel measure (i.e. a countably additive measure on the Borel σ-algebra) on $\mathbf{R}$, such that $0 < \mu(I) < \infty$ for every interval I of positive length. Assume that μ is *inner regular*, in the sense that $\mu(E) = \sup_{K \subset E, \text{ compact}} \mu(K)$ for every Borel measurable set E. (As it turns out, from the theory of *Radon measures*, all locally finite Borel measures have this property, but we will not prove this here; see §1.10 of *An epsilon of room, Vol. I.*) Establish the Hardy-Littlewood maximal inequality

$$\mu(\{x \in \mathbf{R} : \sup_{x \in I} \frac{1}{\mu(I)} \int_I |f(y)|\, d\mu(y) \geq \lambda\}) \leq \frac{2}{\lambda} \int_{\mathbf{R}} |f(y)|\, d\mu(y)$$

for any absolutely integrable function $f \in L^1(\mu)$, where the supremum ranges over all open intervals I that contain x. Note that this essentially generalises Exercise 1.6.11, in which μ is replaced by Lebesgue measure. (*Hint:* Repeat the proof of the usual Hardy-Littlewood maximal inequality, but use the Besicovich covering lemma in place of the Vitali-type covering lemma. Why do we need the former lemma here instead of the latter?)

Exercise 1.6.23 (Cousin's theorem)**.** Prove *Cousin's theorem*: given any function $\delta\colon [a,b] \to (0,+\infty)$ on a compact interval $[a,b]$ of positive length, there exists a partition $a = t_0 < t_1 < \ldots < t_k = b$ with $k \geq 1$, together with real numbers $t_j^* \in [t_{j-1}, t_j]$ for each $1 \leq j \leq k$ and $t_j - t_{j-1} \leq \delta(t_j^*)$. (*Hint:* use the Heine-Borel theorem, which asserts that any open cover of $[a,b]$ has a finite subcover, followed by the Besicovitch covering lemma.) This theorem is useful in a variety of applications related to the second fundamental theorem of calculus, as we shall see below. The positive function δ is known as a *gauge function*.

Now we turn to consequences of the Lebesgue differentiation theorem. Given a Lebesgue measurable set $E \subset \mathbf{R}^d$, call a point $x \in \mathbf{R}^d$ a *point of density* for E if $\frac{m(E\cap B(x,r))}{m(B(x,r))} \to 1$ as $r \to 0$. Thus, for instance, if $E = [-1,1]\backslash\{0\}$, then every point in $(-1,1)$ (including the boundary point 0) is a point of density for E, but the endpoints $-1, 1$ (as well as the exterior of E) are not points of density. One can think of a point of density as being an "almost interior" point of E; it is not necessarily the case that one can fit an small ball $B(x,r)$ centred at x inside of E, but one can fit *most* of that small ball inside E.

Exercise 1.6.24. If $E \subset \mathbf{R}^d$ is Lebesgue measurable, show that almost every point in E is a point of density for E, and almost every point in the complement of E is not a point of density for E.

Exercise 1.6.25. Let $E \subset \mathbf{R}^d$ be a measurable set of positive measure, and let $\varepsilon > 0$.

(i) Using Exercise 1.6.15 and Exercise 1.6.24, show that there exists a cube $Q \subset \mathbf{R}^d$ of positive sidelength such that $m(E\cap Q) > (1-\varepsilon)m(Q)$.

(ii) Give an alternate proof of the above claim that avoids the Lebesgue differentiation theorem. (*Hint:* reduce to the case when E is bounded, then approximate E by an almost disjoint union of cubes.)

(iii) Use the above result to give an alternate proof of the Steinhaus theorem (Exercise 1.6.8).

Of course, one can replace cubes here by other comparable shapes, such as balls. (Indeed, a good principle to adopt in analysis is that cubes and balls are "equivalent up to constants", in that a cube of some sidelength can be contained in a ball of comparable radius, and vice versa. This type of mental equivalence is analogous to, though not identical with, the famous dictum that a topologist cannot distinguish a doughnut from a coffee cup.)

Exercise 1.6.26.

(i) Give an example of a compact set $K \subset \mathbf{R}$ of positive measure such that $m(K \cap I) < |I|$ for every interval I of positive length. (*Hint:* first construct an open dense subset of $[0, 1]$ of measure strictly less than 1.)

(ii) Give an example of a measurable set $E \subset \mathbf{R}$ such that $0 < m(E \cap I) < |I|$ for every interval I of positive length. (*Hint:* first work in a bounded interval, such as $(-1, 2)$. The complement of the set K in the first example is the union of at most countably many open intervals, thanks to Exercise 1.6.10. Now fill in these open intervals and iterate.)

Exercise 1.6.27 (Approximations to the identity)**.** Define a *good kernel*[15] to be a measurable function $P\colon \mathbf{R}^d \to \mathbf{R}^+$ which is non-negative, radial (which means that there is a function $\tilde P\colon [0, +\infty) \to \mathbf{R}^+$ such that $P(x) = \tilde P(|x|)$), radially non-increasing (so that $\tilde P$ is a non-increasing function), and has total mass $\int_{\mathbf{R}^d} P(x)\ dx$ equal to 1. The functions $P_t(x) := \frac{1}{t^d} P(\frac{x}{t})$ for $t > 0$ are then said to be a *good family of approximations to the identity*.

(i) Show that the *heat kernels*[16] $P_t(x) := \frac{1}{(4\pi t^2)^{d/2}} e^{-|x|^2/4t^2}$ and *Poisson kernels* $P_t(x) := c_d \frac{t}{(t^2+|x|^2)^{(d+1)/2}}$ are good families of approximations to the identity, if the constant $c_d > 0$ is chosen correctly (in fact one has $c_d = \Gamma((d+1)/2)/\pi^{(d+1)/2}$, but you are not required to establish this).

[15] Different texts have slightly different notions of what a good kernel is; the "right" class of kernels to consider depends to some extent on what type of convergence results one is interested in (e.g. almost everywhere convergence, convergence in L^1 or L^∞ norm, etc.), and on what hypotheses one wishes to place on the original function f.

[16] Note that we have modified the usual formulation of the heat kernel by replacing t with t^2 in order to make it conform to the notational conventions used in this exercise.

(ii) Show that if P is a good kernel, then

$$c_d < \sum_{n=-\infty}^{\infty} 2^{dn} \tilde P(2^n) \leq C_d$$

for some constants $0 < c_d < C_d$ depending only on d. (*Hint:* compare P with such "horizontal wedding cake" functions as $\sum_{n=-\infty}^\infty 1_{2^{n-1} < |x| \leq 2^n} \tilde P(2^n)$.)

(iii) Establish the quantitative upper bound

$$|\int_{\mathbf{R}^d} f(y) P_t(x-y)\ dy| \leq C'_d \sup_{r>0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y)|\ dy$$

for any absolutely integrable function f and some constant $C'_d > 0$ depending only on d.

(iv) Show that if $f\colon \mathbf{R}^d \to \mathbf{C}$ is absolutely integrable and x is a Lebesgue point of f, then the convolution

$$f * P_t(x) := \int_{\mathbf{R}^d} f(y) P_t(x-y)\ dy$$

converges to $f(x)$ as $t \to 0$. (*Hint:* split $f(y)$ as the sum of $f(x)$ and $f(y) - f(x)$.) In particular, $f * P_t$ converges pointwise almost everywhere to f.

1.6.3. Almost everywhere differentiability. As we see in undergraduate real analysis, not every continuous function $f\colon \mathbf{R} \to \mathbf{R}$ is differentiable, with the standard example being the absolute value function $f(x) := |x|$, which is continuous not differentiable at the origin $x = 0$. Of course, this function is still almost everywhere differentiable. With a bit more effort, one can construct continuous functions that are in fact *nowhere differentiable*:

Exercise 1.6.28 (Weierstrass function)**.** Let $F\colon \mathbf{R} \to \mathbf{R}$ be the function

$$F(x) := \sum_{n=1}^\infty 4^{-n} \sin(8^n \pi x).$$

(i) Show that F is well-defined (in the sense that the series is absolutely convergent) and that F is a bounded continuous function.

(ii) Show that for every 8-dyadic interval $[\frac{j}{8^n}, \frac{j+1}{8^n}]$ with $n \geq 1$, one has $|F(\frac{j+1}{8^n}) - F(\frac{j}{8^n})| \geq c4^{-n}$ for some absolute constant $c > 0$.

(iii) Show that F is not differentiable at any point $x \in \mathbf{R}$. (*Hint:* argue by contradiction and use the previous part of this exercise.) Note that it is *not* enough to formally differentiate the series term by term and observe that the resulting series is divergent - why not?

The difficulty here is that a continuous function can still contain a large amount of *oscillation*, which can lead to breakdown of differentiability. However, if one can somehow limit the amount of oscillation present, then one can often recover a fair bit of differentiability. For instance, we have

Theorem 1.6.25 (Monotone differentiation theorem)**.** *Any function* $F \colon \mathbf{R} \to \mathbf{R}$ *which is monotone (either monotone non-decreasing or monotone non-increasing) is differentiable almost everywhere.*

Exercise 1.6.29. Show that every monotone function is measurable.

To prove this theorem, we just treat the case when F is monotone non-decreasing, as the non-increasing case is similar (and can be deduced from the non-decreasing case by replacing F with $-F$).

We also first focus on the case when F is continuous, as this allows us to use the rising sun lemma. To understand the differentiability of F, we introduce the four *Dini derivatives* of F at x:

(i) The upper right derivative $\overline{D^+}F(x) := \limsup_{h \to 0^+} \frac{F(x+h)-F(x)}{h}$;

(ii) The lower right derivative $\underline{D^+}F(x) := \liminf_{h \to 0^+} \frac{F(x+h)-F(x)}{h}$;

(iii) The upper left derivative $\overline{D^-}F(x) := \limsup_{h \to 0^-} \frac{F(x+h)-F(x)}{h}$;

(iv) The lower right derivative $\underline{D^-}F(x) := \liminf_{h \to 0^-} \frac{F(x+h)-F(x)}{h}$.

Regardless of whether F is differentiable or not (or even whether F is continuous or not), the four Dini derivatives always exist and take values in the extended real line $[-\infty, \infty]$. (If F is only defined on an interval $[a, b]$, rather than on the endpoints, then some of the Dini derivatives may not exist at the endpoints, but this is a measure zero set and will not impact our analysis.)

Exercise 1.6.30. If F is monotone, show that the four Dini derivatives of F are measurable. (*Hint:* the main difficulty is to reformulate the derivatives so that h ranges over a countable set rather than an uncountable one.)

A function F is differentiable at x precisely when the four derivatives are equal and finite:

$$\overline{D^+}F(x) = \underline{D^+}F(x) = \overline{D^-}F(x) = \underline{D^-}F(x) \in (-\infty, +\infty). \tag{1.29}$$

We also have the trivial inequalities

$$\underline{D^+}F(x) \leq \overline{D^+}F(x); \quad \underline{D^-}F(x) \leq \overline{D^-}F(x).$$

If F is non-decreasing, all these quantities are non-negative, thus

$$0 \leq \underline{D^+}F(x) \leq \overline{D^+}F(x); \quad 0 \leq \underline{D^-}F(x) \leq \overline{D^-}F(x).$$

The one-sided Hardy-Littlewood maximal inequality has an analogue in this setting:

Lemma 1.6.26 (One-sided Hardy-Littlewood inequality)**.** *Let $F\colon [a,b] \to \mathbf{R}$ be a continuous monotone non-decreasing function, and let $\lambda > 0$. Then we have*

$$m(\{x \in [a,b] : \overline{D^+}F(x) \geq \lambda\}) \leq \frac{F(b) - F(a)}{\lambda}.$$

Similarly for the other three Dini derivatives of F.

If F is not assumed to be continuous, then we have the weaker inequality

$$m(\{x \in [a,b] : \overline{D^+}F(x) \geq \lambda\}) \leq C\frac{F(b) - F(a)}{\lambda}$$

for some absolute constant $C > 0$.

Remark 1.6.27. Note that if one naively applies the fundamental theorems of calculus, one can *formally* see that the first part of Lemma 1.6.26 is equivalent to Lemma 1.6.16. We cannot however use this argument rigorously because we have not established the necessary fundamental theorems of calculus to do this. Nevertheless, we can borrow the *proof* of Lemma 1.6.16 without difficulty to use here, and this is exactly what we will do.

Proof. We just prove the continuous case and leave the discontinuous case as an exercise.

It suffices to prove the claim for $\overline{D^+}F$; by reflection (replacing $F(x)$ with $-F(-x)$, and $[a,b]$ with $[-b,-a]$), the same argument works for $\overline{D^-}F$, and then this trivially implies the same inequalities for $\underline{D^+}F$ and $\underline{D^-}F$. By modifying λ by an epsilon, and dropping the endpoints from $[a,b]$ as they have measure zero, it suffices to show that

$$m(\{x \in (a,b) : \overline{D^+}F(x) > \lambda\}) \leq \frac{F(b)-F(a)}{\lambda}$$

We may apply the rising sun lemma (Lemma 1.6.17) to the continuous function $G(x) := F(x) - \lambda x$. This gives us an at most countable family of intervals $I_n = (a_n, b_n)$ in (a,b), such that $G(b_n) \geq G(a_n)$ for each n, and such that $G(y) \leq G(x)$ whenever $a \leq x \leq y \leq b$ and x lies outside of all of the I_n.

Observe that if $x \in (a,b)$, and $G(y) \leq G(x)$ for all $x \leq y \leq b$, then $\overline{D^+}F(x) \leq \lambda$. Thus we see that the set $\{x \in (a,b) : \overline{D^+}F(x) > \lambda\}$ is contained in the union of the I_n, and so by countable additivity

$$m(\{x \in (a,b) : \overline{D^+}F(x) > \lambda\}) \leq \sum_n b_n - a_n.$$

But we can rearrange the inequality $G(b_n) \leq G(a_n)$ as $b_n - a_n \leq \frac{F(b_n)-F(a_n)}{\lambda}$. From telescoping series and the monotone nature of F we have $\sum_n F(b_n) - F(a_n) \leq F(b) - F(a)$ (this is easiest to prove by first working with a finite subcollection of the intervals (a_n, b_n), and then taking suprema), and the claim follows.

The discontinuous case is left as an exercise. □

Exercise 1.6.31. Prove Lemma 1.6.26 in the discontinuous case. (*Hint:* the rising sun lemma is no longer available, but one can use either the Vitali-type covering lemma (which will give $C = 3$) or the Besicovitch lemma (which will give $C = 2$), by modifying the proof of Theorem 1.6.20.

Sending $\lambda \to \infty$ in the above lemma (cf. Exercise 1.3.18), and then sending $[a,b]$ to $\mathbf{R}$, we conclude as a corollary that all the four Dini derivatives of a continuous monotone non-decreasing function are finite almost everywhere. So to prove Theorem 1.6.25 for continuous monotone non-decreasing functions, it suffices to show that (1.29) holds for almost every x. In view of the trivial inequalities, it suffices to show that $\overline{D_+}F(x) \leq \underline{D_-}F(x)$ and $\overline{D_-}F(x) \leq \underline{D_+}F(x)$ for almost every x. We will just show the first inequality, as the second follows by replacing F with its reflection $x \mapsto -F(-x)$. It will suffice to show that for every pair $0 < r < R$ of real numbers, the set

$$E = E_{r,R} := \{x \in \mathbf{R} : \overline{D_+}F(x) > R > r > \underline{D_-}F(x)\}$$

is a null set, since by letting R, r range over rationals with $R > r > 0$ and taking countable unions, we would conclude that the set $\{x \in \mathbf{R} : \overline{D_+}F(x) > \underline{D_-}F(x)\}$ is a null set (recall that the Dini derivatives are all non-negative when F is non-decreasing), and the claim follows.

Clearly E is a measurable set. To prove that it is null, we will establish the following estimate:

Lemma 1.6.28 (E has density less than one)**.** *For any interval* $[a, b]$ *and any* $0 < r < R$*, one has* $m(E_{r,R} \cap [a, b]) \leq \frac{r}{R}|b - a|$.

Indeed, this lemma implies that E has no points of density, which by Exercise 1.6.24 forces E to be a null set.

Proof. We begin by applying the rising sun lemma to the function $G(x) := rx + F(-x)$ on $[-b, -a]$; the large number of negative signs present here is needed in order to properly deal with the lower left Dini derivative $\underline{D_-}F$. This gives an at most countable family of disjoint intervals $-I_n = (-b_n, -a_n)$ in $(-b, -a)$, such that $G(-a_n) \geq G(-b_n)$ for all n, and such that $G(-x) \leq G(-y)$ whenever $-x \leq -y \leq -a$ and $-x \in (-b, -a)$ lies outside of all of the $-I_n$. Observe that if $x \in (a, b)$, and $G(-x) \leq G(-y)$ for all $-x \leq -y \leq -a$, then $\underline{D_-}F(x) \geq r$. Thus we see that $E_{r,R}$ is contained inside the union of the intervals $I_n = (a_n, b_n)$. On the other hand, from the first part of Lemma 1.6.26 we have

$$m(E_{r,R} \cap (a_n, b_n)) \leq \frac{F(b_n) - F(a_n)}{R}.$$

But we can rearrange the inequality $G(-a_n) \leq G(-b_n)$ as $F(b_n) - F(a_n) \leq r(b_n - a_n)$. From countable additivity, one thus has

$$m(E_{r,R}) \leq \frac{r}{R} \sum_n b_n - a_n.$$

But the (a_n, b_n) are disjoint inside (a, b), so from countable additivity again, we have $\sum_n b_n - a_n \leq b - a$, and the claim follows. $\square$

Remark 1.6.29. Note if F was not assumed to be continuous, then one would lose a factor of C here from the second part of Lemma 1.6.26, and one would then be unable to prevent $\overline{D^+ F}$ from being up to C times as large as $\underline{D_- F}$. So sometimes, even when all one is seeking is a qualitative result such as differentiability, it is still important to keep track of constants. (But this is the exception rather than the rule: for a large portion of arguments in analysis, the constants are not terribly important.)

This concludes the proof of Theorem 1.6.25 in the continuous monotone non-decreasing case. Now we work on removing the continuity hypothesis (which was needed in order to make the rising sun lemma work properly). If we naively try to run the density argument as we did in previous sections, then (for once) the argument does *not* work very well, as the space of continuous monotone functions are not sufficiently dense in the space of all monotone functions in the relevant sense (which, in this case, is in the *total variation* sense, which is what is needed to invoke such tools as Lemma 1.6.26.). To bridge this gap, we have to supplement the continuous monotone functions with another class of monotone functions, known as the *jump functions*.

Definition 1.6.30 (Jump function)**.** A *basic jump function* J is a function of the form

$$J(x) := \begin{cases} 0 & \text{when } x < x_0 \\ \theta & \text{when } x = x_0 \\ 1 & \text{when } x > x_0 \end{cases}$$

for some real numbers $x_0 \in \mathbf{R}$ and $0 \leq \theta \leq 1$; we call x_0 the *point of discontinuity* for J and θ the *fraction*. Observe that such functions are monotone non-decreasing, but have a discontinuity at one point. A *jump function* is any absolutely convergent combination of basic jump functions, i.e. a function of the form $F = \sum_n c_n J_n$, where n ranges over an at most countable set, each J_n is a basic jump function, and the c_n are positivereals with $\sum_n c_n < \infty$. If there are only finitely many n involved, we say that F is a *piecewise constant jump function*.

Thus, for instance, if $q_1, q_2, q_3, \ldots$ is any enumeration of the rationals, then $\sum_{n=1}^\infty 2^{-n} 1_{[q_n,+\infty)}$ is a jump function.

Clearly, all jump functions are monotone non-decreasing. From the absolute convergence of the c_n we see that every jump function is the uniform limit of piecewise constant jump functions, for instance $\sum_{n=1}^\infty c_n J_n$ is the uniform limit of $\sum_{n=1}^N c_n J_n$. One consequence of this is that the points of discontinuity of a jump function $\sum_{n=1}^\infty c_n J_n$ are precisely those of the individual summands $c_n J_n$, i.e. of the points x_n where each J_n jumps.

The key fact is that these functions, together with the continuous monotone functions, essentially generate all monotone functions, at least in the bounded case:

Lemma 1.6.31 (Continuous-singular decomposition for monotone functions)**.** *Let $F\colon \mathbf{R} \to \mathbf{R}$ be a monotone non-decreasing function.*

(i) *The only discontinuities of F are jump discontinuities. More precisely, if x is a point where F is discontinuous, then the limits $\lim_{y \to x^-} F(y)$ and $\lim_{y \to x^+} F(y)$ both exist, but are unequal, with $\lim_{y\to x^-} F(y) < \lim_{y \to x^+} F(y)$.*

(ii) *There are at most countably many discontinuities of F.*

(iii) *If F is bounded, then F can be expressed as the sum of a continuous monotone non-decreasing function F_c and a jump function F_{pp}.*

Remark 1.6.32. This decomposition is part of the more general *Lebesgue decomposition*, discussed in §1.2 of *An epsilon of room, Vol. I*.

Proof. By monotonicity, the limits $F_-(x) := \lim_{y \to x^-} F(y)$ and $F^+(x) := \lim_{y\to x^+} F(y)$ always exist, with $F_-(x) \leq F(x) \leq F_+(x)$ for all x. This gives (i).

By (i), whenever there is a discontinuity x of F, there is at least one rational number q_x strictly between $F_-(x)$ and $F_+(x)$, and from monotonicity, each rational number can be assigned to at most one discontinuity. This gives (ii).

Now we prove (iii). Let A be the set of discontinuities of F, thus A is at most countable. For each $x \in A$, we define the jump $c_x := F_+(x) - F_-(x) > 0$, and the fraction $\theta_x := \frac{F(x)-F_-(x)}{F_+(x)-F_-(x)} \in [0,1]$. Thus

$$F_+(x) = F_-(x) + c_x \text{ and } F(x) = F_-(x) + \theta_x c_x.$$

Note that c_x is the measure of the interval $(F_-(x), F_+(x))$. By monotonicity, these intervals are disjoint; by the boundedness of F, their union is bounded. By countable additivity, we thus have $\sum_{x \in A} c_x < \infty$, and so if we let J_x be the basic jump function with point of discontinuity x and fraction θ_x, then the function

$$F_{pp} := \sum_{x \in A} c_x J_x$$

is a jump function.

As discussed previously, G is discontinuous only at A, and for each $x \in A$ one easily checks that

$$(F_{pp})_+(x) = (F_{pp})_-(x) + c_x \text{ and } F_{pp}(x) = (F_{pp})_-(x) + \theta_x c_x$$

where $(F_{pp})_-(x) := \lim_{y \to x^-} F_{pp}(y)$, and $(F_{pp})_+(x) := \lim_{y \to x^+} F_{pp}(y)$. We thus see that the difference $F_c := F - F_{pp}$ is continuous. The only remaining task is to verify that F_c is monotone non-decreasing, thus we need

$$F_{pp}(b) - F_{pp}(a) \leq F(b) - F(a)$$

for all $a < b$. But the left-hand side can be rewritten as $\sum_{x \in A \cap [a,b]} c_x$. As each c_x is the measure of the interval $(F_-(x), F_+(x))$, and these intervals for $x \in A \cap [a,b]$ are disjoint and lie in $(F(a), F(b))$, the claim follows from countable additivity. □

Exercise 1.6.32. Show that the decomposition of a bounded monotone non-decreasing function F into continuous F_c and jump components F_{pp} given by the above lemma is unique.

Exercise 1.6.33. Find a suitable generalisation of the notion of a jump function that allows one to extend the above decomposition to unbounded monotone functions, and then prove this extension. (*Hint:* the notion to shoot for here is that of a "locally jump function".)

Now we can finish the proof of Theorem 1.6.25. As noted previously, it suffices to prove the claim for monotone non-decreasing functions. As differentiability is a local condition, we can easily reduce to the case of *bounded* monotone non-decreasing functions, since to test differentiability of a monotone non-decreasing function F in any compact interval $[a, b]$ we may replace F by the bounded monotone non-decreasing function $\max(\min(F, F(b)), F(a))$ with no change in the differentiability in $[a, b]$ (except perhaps at the endpoints a, b, but these form a set of measure zero). As we have already proven the claim for continuous functions, it suffices by Lemma 1.6.31 (and linearity of the derivative) to verify the claim for jump functions.

Now, finally, we are able to use the density argument, using the piecewise constant jump functions as the dense subclass, and using the second part of Lemma 1.6.26 for the quantitative estimate; fortunately for us, the density argument does not particularly care that there is a loss of a constant factor in this estimate.

For piecewise constant jump functions, the claim is clear (indeed, the derivative exists and is zero outside of finitely many discontinuities). Now we run the density argument. Let F be a bounded jump function, and let $\varepsilon > 0$ and $\lambda > 0$ be arbitrary. As every jump function is the uniform limit of piecewise constant jump functions, we can find a piecewise constant jump function F_ε such that $|F(x) - F_\varepsilon(x)| \leq \varepsilon$ for all x. Indeed, by taking F_ε to be a partial sum of the basic jump functions that make up F, we can ensure that $F - F_\varepsilon$ is also a monotone non-decreasing function. Applying the second part of Lemma 1.6.26, we have

$$\{x \in \mathbf{R} : \overline{D^+}(F - F_\varepsilon)(x) \geq \lambda\} \leq \frac{2C\varepsilon}{\lambda}$$

for some absolute constant C, and similarly for the other four Dini derivatives. Thus, outside of a set of measure at most $8C\varepsilon/\lambda$, all of the Dini derivatives of $F - F_\varepsilon$ are less than λ. Since F'_ε is almost everywhere differentiable, we conclude that outside of a set of measure at most $8C\varepsilon/\lambda$, all the Dini derivatives of $F(x)$ lie within λ of $F'_\varepsilon(x)$, and in particular are finite and lie within 2λ of each other. Sending ε to zero (holding λ fixed), we conclude that for almost every x, the Dini derivatives of F are finite and lie within 2λ of each other. If we then send λ to zero, we see that for almost every x, the Dini derivatives of F agree with each other and are finite, and the claim follows. This concludes the proof of Theorem 1.6.25.

Just as the integration theory of unsigned functions can be used to develop the integration theory of the absolutely convergent functions (see Section 1.3.4), the differentiation theory of monotone functions can be used to develop a parallel differentiation theory for the class of functions of *bounded variation*:

Definition 1.6.33 (Bounded variation)**.** Let $F\colon \mathbf{R} \to \mathbf{R}$ be a function. The *total variation* $\|F\|_{TV(\mathbf{R})}$ (or $\|F\|_{TV}$ for short) of F is defined to be the supremum

$$\|F\|_{TV(\mathbf{R})} := \sup_{x_0 < \ldots < x_n} \sum_{i=1}^{n} |F(x_i) - F(x_{i+1})|$$

where the supremum ranges over all finite increasing sequences $x_0, \ldots, x_n$ of real numbers with $n \geq 0$; this is a quantity in $[0, +\infty]$. We say that F has *bounded variation* (on $\mathbf{R}$) if $\|F\|_{TV(\mathbf{R})}$ is finite. (In this case, $\|F\|_{TV(\mathbf{R})}$ is often written as $\|F\|_{BV(\mathbf{R})}$ or just $\|F\|_{BV}$.)

Given any interval $[a, b]$, we define the total variation $\|F\|_{TV([a,b])}$ of F on $[a, b]$ as

$$\|F\|_{TV([a,b])} := \sup_{a \leq x_0 < \ldots < x_n \leq b} \sum_{i=1}^{n} |F(x_i) - F(x_{i+1})|;$$

thus the definition is the same, but the points $x_0, \ldots, x_n$ are restricted to lie in $[a, b]$. Thus for instance $\|F\|_{TV(\mathbf{R})} = \sup_{N \to \infty} \|F\|_{TV([-N,N])}$. We say that a function F has *bounded variation* on $[a, b]$ if $\|F\|_{BV([a,b])}$ is finite.

Exercise 1.6.34. If $F\colon \mathbf{R} \to \mathbf{R}$ is a monotone function, show that $\|F\|_{TV([a,b])} = |F(b) - F(a)|$ for any interval $[a, b]$, and that F has bounded variation on $\mathbf{R}$ if and only if it is bounded.

Exercise 1.6.35. For any functions $F, G\colon \mathbf{R} \to \mathbf{R}$, establish the triangle property $\|F + G\|_{TV(\mathbf{R})} \leq \|F\|_{TV(\mathbf{R})} + \|G\|_{TV(\mathbf{R})}$ and the homogeneity property $\|cF\|_{TV(\mathbf{R})} = |c|\|F\|_{TV(\mathbf{R})}$ for any $c \in \mathbf{R}$. Also show that $\|F\|_{TV} = 0$ if and only if F is constant.

Exercise 1.6.36. If $F\colon \mathbf{R} \to \mathbf{R}$ is a function, show that $\|F\|_{TV([a,b])} + \|F\|_{TV([b,c])} = \|F\|_{TV([a,c])}$ whenever $a \leq b \leq c$.

Exercise 1.6.37.
(i) Show that every function $f\colon \mathbf{R} \to \mathbf{R}$ of bounded variation is bounded, and that the limits $\lim_{x\to+\infty} f(x)$ and $\lim_{x\to-\infty} f(x)$, are well-defined.
(ii) Give a counterexample of a bounded, continuous, compactly supported function f that is not of bounded variation.

Exercise 1.6.38. Let $f\colon \mathbf{R} \to \mathbf{R}$ be an absolutely integrable function, and let $F\colon \mathbf{R} \to \mathbf{R}$ be the indefinite integral $F(x) := \int_{[-\infty,x]} f(x)$. Show that F is of bounded variation, and that $\|F\|_{TV(\mathbf{R})} = \|f\|_{L^1(\mathbf{R})}$. (*Hint:* the upper bound $\|F\|_{TV(\mathbf{R})} \leq \|f\|_{L^1(\mathbf{R})}$ is relatively easy to establish. To obtain the lower bound, use the density argument.)

Much as an absolutely integrable function can be expressed as the difference of its positive and negative parts, a bounded variation function can be expressed as the difference of two bounded monotone functions:

Proposition 1.6.34. *A function $F\colon \mathbf{R} \to \mathbf{R}$ is of bounded variation if and only if it is the difference of two bounded monotone functions.*

Proof. It is clear from Exercises 1.6.34, 1.6.35 that the difference of two bounded monotone functions is bounded. Now define the *positive variation* $F^+\colon \mathbf{R} \to \mathbf{R}$ of F by the formula

$$F^+(x) := \sup_{x_0 < \ldots < x_n \leq x} \sum_{i=1}^{n} \max(F(x_{i+1}) - F(x_i), 0). \tag{1.30}$$

It is clear from construction that this is a monotone increasing function, taking values between 0 and $\|F\|_{TV(\mathbf{R})}$, and is thus bounded. To conclude the proposition, it suffices to (by writing $F = F_+ - (F_+ - F_-)$ to show that $F_+ - F$ is non-decreasing, or in other words to show that

$$F^+(b) \geq F^+(a) + F(b) - F(a).$$

If $F(b) - F(a)$ is negative then this is clear from the monotone non-decreasing nature of F^+, so assume that $F(b) - F(a) \geq 0$. But then the claim follows because any sequence of real numbers $x_0 < \ldots < x_n \leq a$ can be extended by one or two elements by adding a and b, thus increasing the sum $\sup_{x_0<\ldots<x_n} \sum_{i=1}^n \max(F(x_i) - F(x_{i+1}), 0)$ by at least $F(b) - F(a)$. □

Exercise 1.6.39. Let $F\colon \mathbf{R} \to \mathbf{R}$ be of bounded variation. Define the positive variation F^+ by (1.30), and the negative variation F^- by

$$F^-(x) := \sup_{x_0<\ldots<x_n\leq x} \sum_{i=1}^n \max(-F(x_{i+1}) + F(x_i), 0).$$

Establish the identities

$$F(x) = F(-\infty) + F^+(x) - F^-(x),$$

$$\|F\|_{TV[a,b]} = F^+(b) - F^+(a) + F^-(b) - F^-(a),$$

and

$$\|F\|_{TV} = F^+(+\infty) + F^-(+\infty)$$

for every interval $[a, b]$, where $F(-\infty) := \lim_{x\to-\infty} F(x)$, $F^+(+\infty) := \lim_{x\to+\infty} F^+(x)$, and $F^-(+\infty) := \lim_{x\to+\infty} F^-(x)$. (*Hint:* The main difficulty comes from the fact that a partition $x_0 < \ldots < x_n \leq x$ that is good for F^+ need not be good for F^-, and vice versa. However, this can be fixed by taking a good partition for F^+ and a good partition for F^- and combining them together into a common refinement.)

From Proposition 1.6.34 and Theorem 1.6.25 we immediately obtain

Corollary 1.6.35 (BV differentiation theorem)**.** *Every bounded variation function is differentiable almost everywhere.*

Exercise 1.6.40. Call a function *locally of bounded variation* if it is of bounded variation on every compact interval $[a, b]$. Show that every function that is locally of bounded variation is differentiable almost everywhere.

Exercise 1.6.41 (Lipschitz differentiation theorem, one-dimensional case)**.** A function $f\colon \mathbf{R} \to \mathbf{R}$ is said to be *Lipschitz continuous* if there exists a constant $C > 0$ such that $|f(x) - f(y)| \leq C|x - y|$ for all $x, y \in \mathbf{R}$; the smallest C with this property is known as the *Lipschitz constant* of f. Show that every Lipschitz continuous function

F is locally of bounded variation, and hence differentiable almost everywhere. Furthermore, show that the derivative F', when it exists, is bounded in magnitude by the Lipschitz constant of F.

Remark 1.6.36. The same result is true in higher dimensions, and is known as the *Rademacher differentiation theorem*, but we will defer the proof of this theorem to Section 2.2, when we have the powerful tool of the Fubini-Tonelli theorem (Corollary 1.7.23) available, that is particularly useful for deducing higher-dimensional results in analysis from lower-dimensional ones.

Exercise 1.6.42. A function $f\colon \mathbf{R} \to \mathbf{R}$ is said to be *convex* if one has $f((1-t)x+ty) \leq (1-t)f(x)+tf(y)$ for all $x < y$ and $0 < t < 1$. Show that if f is convex, then it is continuous and almost everywhere differentiable, and its derivative f' is equal almost everywhere to a monotone non-decreasing function, and so is itself almost everywhere differentiable. (*Hint:* Drawing the graph of f, together with a number of chords and tangent lines, is likely to be very helpful in providing visual intuition.) Thus we see that in some sense, convex functions are "almost everywhere twice differentiable". Similar claims also hold for concave functions, of course.

1.6.4. The second fundamental theorem of calculus.

We are now finally ready to attack the second fundamental theorem of calculus in the cases where F is not assumed to be continuously differentiable. We begin with the case when $F\colon [a,b] \to \mathbf{R}$ is monotone non-decreasing. From Theorem 1.6.25 (extending F to the rest of the real line if needed), this implies that F is differentiable almost everywhere in $[a,b]$, so F' is defined a.e.; from monotonicity we see that F' is non-negative whenever it is defined. Also, an easy modification of Exercise 1.6.1 shows that F' is measurable.

One half of the second fundamental theorem is easy:

Proposition 1.6.37 (Upper bound for second fundamental theorem)**.** *Let $F\colon [a,b] \to \mathbf{R}$ be monotone non-decreasing (so that, as discussed above, F' is defined almost everywhere, is unsigned, and is measurable). Then*

$$\int_{[a,b]} F'(x)\ dx \leq F(b) - F(a).$$

In particular, F' is absolutely integrable.

Proof. It is convenient to extend F to all of $\mathbf{R}$ by declaring $F(x) := F(b)$ for $x > b$ and $F(x) := F(a)$ for $x < a$, then F is now a bounded monotone function on $\mathbf{R}$, and F' vanishes outside of $[a, b]$. As F is almost everywhere differentiable, the Newton quotients

$$f_n(x) := \frac{F(x + 1/n) - F(x)}{1/n}$$

converge pointwise almost everywhere to F'. Applying Fatou's lemma (Corollary1.4.47), we conclude that

$$\int_{[a,b]} F'(x)\ dx \leq \liminf_{n\to\infty} \int_{[a,b]} \frac{F(x + 1/n) - F(x)}{1/n}\ dx.$$

The right-hand side can be rearranged as

$$\liminf_{n\to\infty} n(\int_{[a+1/n,b+1/n]} F(y)\ dy - \int_{[a,b]} F(x)\ dx)$$

which can be rearranged further as

$$\liminf_{n\to\infty} n(\int_{[b,b+1/n]} F(x)\ dx - \int_{[a,a+1/n]} F(x)\ dx).$$

Since F is equal to $F(b)$ for the first integral and is at least $F(a)$ for the second integral, this expression is at most

$$\leq \liminf_{n\to\infty} n(F(b)/n - F(a)/n) = F(b) - F(a)$$

and the claim follows. □

Exercise 1.6.43. Show that any function of bounded variation has an (almost everywhere defined) derivative that is absolutely integrable.

In the Lipschitz case, one can do better:

Exercise 1.6.44 (Second fundamental theorem for Lipschitz functions)**.** Let $F\colon [a, b] \to \mathbf{R}$ be Lipschitz continuous. Show that $\int_{[a,b]} F'(x)\, dx = F(b) - F(a)$. (*Hint:* Argue as in the proof of Proposition 1.6.37, but use the dominated convergence theorem (Theorem 1.4.49) in place of Fatou's lemma (Corollary1.4.47).)

Exercise 1.6.45 (Integration by parts formula)**.** Let $F, G\colon [a,b] \to \mathbf{R}$ be Lipschitz continuous functions. Show that

$$\int_{[a,b]} F'(x)G(x)\ dx = F(b)G(b) - F(a)G(a) - \int_{[a,b]} F(x)G'(x)\ dx.$$

(*Hint:* first show that the product of two Lipschitz continuous functions on $[a,b]$ is again Lipschitz continuous.)

Now we return to the monotone case. Inspired by the Lipschitz case, one may hope to recover equality in Proposition 1.6.37 for such functions F. However, there is an important obstruction to this, which is that all the variation of F may be concentrated in a set of measure zero, and thus undetectable by the Lebesgue integral of F'. This is most obvious in the case of a discontinuous monotone function, such as the (appropriately named) *Heaviside function* $F := 1_{[0,+\infty)}$; it is clear that F' vanishes almost everywhere, but $F(b) - F(a)$ is not equal to $\int_{[a,b]} F'(x)\ dx$ if b and a lie on opposite sides of the discontinuity at 0. In fact, the same problem arises for all jump functions:

Exercise 1.6.46. Show that if F is a jump function, then F' vanishes almost everywhere. (*Hint:* use the density argument, starting from piecewise constant jump functions and using Proposition 1.6.37 as the quantitative estimate.)

One may hope that jump functions - in which all the fluctuation is concentrated in a countable set - are the only obstruction to the second fundamental theorem of calculus holding for monotone functions, and that as long as one restricts attention to *continuous* monotone functions, that one can recover the second fundamental theorem. However, this is still not true, because it is possible for all the fluctuation to now be concentrated, not in a countable collection of jump discontinuities, but instead in an uncountable set of zero measure, such as the middle thirds Cantor set (Exercise 1.2.9). This can be illustrated by the key counterexample of the *Cantor function*, also known as the *Devil's staircase function*. The construction of this function is detailed in the exercise below.

Exercise 1.6.47 (Cantor function)**.** Define the functions $F_0, F_1, F_2, \ldots : [0,1] \to \mathbf{R}$ recursively as follows:

1. Set $F_0(x) := x$ for all $x \in [0,1]$.
2. For each $n = 1, 2, \ldots$ in turn, define
$$F_n(x) := \begin{cases} \frac{1}{2}F_{n-1}(3x) & \text{if } x \in [0,1/3]; \\ \frac{1}{2} & \text{if } x \in (1/3,2/3); \\ \frac{1}{2} + \frac{1}{2}F_{n-1}(3x-2) & \text{if } x \in [2/3,1] \end{cases}$$

(i) Graph F_0, F_1, F_2, and F_3 (preferably on a single graph).

(ii) Show that for each $n = 0, 1, \ldots$, F_n is a continuous monotone non-decreasing function with $F_n(0) = 0$ and $F_n(1) = 1$. (*Hint:* induct on n.)

(iii) Show that for each $n = 0, 1, \ldots$, one has $|F_{n+1}(x) - F_n(x)| \leq 2^{-n}$ for each $x \in [0,1]$. Conclude that the F_n converge uniformly to a limit $F\colon [0,1] \to \mathbf{R}$. This limit is known as the *Cantor function*.

(iv) Show that the Cantor function F is continuous and monotone non-decreasing, with $F(0) = 0$ and $F(1) = 1$.

(v) Show that if $x \in [0,1]$ lies outside the middle thirds Cantor set (Exercise 1.2.9), then F is constant in a neighbourhood of x, and in particular $F'(x) = 0$. Conclude that $\int_{[0,1]} F'(x)\ dx = 0 \neq 1 = F(1) - F(0)$, so that the second fundamental theorem of calculus fails for this function.

(vi) Show that $F(\sum_{n=1}^\infty a_n 3^{-n}) = \sum_{n=1}^\infty \frac{a_n}{2} 2^{-n}$ for any digits $a_1, a_2, \ldots \in \{0,2\}$. Thus the Cantor function, in some sense, converts base three expansions to base two expansions.

(1) Let $I = [\sum_{i=1}^n \frac{a_i}{3^i}, \sum_{i=1}^n \frac{a_i}{3^i} + \frac{1}{3^n}]$ be one of the intervals used in the n^{th} cover I_n of C (see Exercise 1.2.9), thus $n \geq 0$ and $a_1, \ldots, a_n \in \{0,2\}$. Show that I is an interval of length 3^{-n}, but $F(I)$ is an interval of length 2^{-n}.

(2) Show that F is not differentiable at any element of the Cantor set C.

Remark 1.6.38. This example shows that the classical derivative $F'(x) := \lim_{h \to 0; h \neq 0} \frac{F(x+h)-F(x)}{h}$ of a function has some defects; it cannot "see" some of the variation of a continuous monotone function such as the Cantor function. In §1.13 of *An epsilon of room, Vol. I*, this will be rectified by introducing the concept of the *weak derivative* of a function, which despite the name, is more able than the strong derivative to detect this type of singular variation behaviour. (We will also encounter in Section 1.7.3 the *Lebesgue-Stieltjes integral*, which is another (closely related) way to capture all of the variation of a monotone function, and which is related to the classical derivative via the *Lebesgue-Radon-Nikodym theorem*, see §1.2 of *An epsilon of room, Vol. I*.)

In view of this counterexample, we see that we need to add an additional hypothesis to the continuous monotone non-increasing function F before we can recover the second fundamental theorem. One such hypothesis is *absolute continuity*. To motivate this definition, let us recall two existing definitions:

(i) A function $F\colon \mathbf{R} \to \mathbf{R}$ is *continuous* if, for every $\varepsilon > 0$ and $x_0 \in \mathbf{R}$, there exists a $\delta > 0$ such that $|F(b) - F(a)| \leq \varepsilon$ whenever (a, b) is an interval of length at most δ that contains x_0.

(ii) A function $F\colon \mathbf{R} \to \mathbf{R}$ is *uniformly continuous* if, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $|F(b) - F(a)| \leq \varepsilon$ whenever (a, b) is an interval of length at most δ.

Definition 1.6.39. A function $F\colon \mathbf{R} \to \mathbf{R}$ is said to be *absolutely continuous* if, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $\sum_{j=1}^n |F(b_j) - F(a_j)| \leq \varepsilon$ whenever $(a_1, b_1), \ldots, (a_n, b_n)$ is a finite collection of disjoint intervals of total length $\sum_{j=1}^n b_j - a_j$ at most δ.

We define absolute continuity for a function $F\colon [a, b] \to \mathbf{R}$ defined on an interval $[a, b]$ similarly, with the only difference being that the intervals $[a_j, b_j]$ are of course now required to lie in the domain $[a, b]$ of F.

The following exercise places absolute continuity in relation to other regularity properties:

Exercise 1.6.48. (i) Show that every absolutely continuous function is uniformly continuous and therefore continuous.

(ii) Show that every absolutely continuous function is of bounded variation on every compact interval $[a, b]$. (*Hint:* first show this is true for any sufficiently small interval.) In particular (by Exercise 1.6.40), absolutely continuous functions are differentiable almost everywhere.

(iii) Show that every Lipschitz continuous function is absolutely continuous.

(iv) Show that the function $x \mapsto \sqrt{x}$ is absolutely continuous, but not Lipschitz continuous, on the interval $[0, 1]$.

(v) Show that the Cantor function from Exercise 1.6.47 is continuous, monotone, and uniformly continuous, but *not* absolutely continuous, on $[0, 1]$.

(vi) If $f : \mathbf{R} \to \mathbf{R}$ is absolutely integrable, show that the indefinite integral $F(x) := \int_{[-\infty,x]} f(y)\ dy$ is absolutely continuous, and that F is differentiable almost everywhere with $F'(x) = f(x)$ for almost every x.

(vii) Show that the sum or product of two absolutely continuous functions on an interval $[a, b]$ remains absolutely continuous. What happens if we work on $\mathbf{R}$ instead of on $[a, b]$?

Exercise 1.6.49. (i) Show that absolutely continuous functions map null sets to null sets, i.e. if $F : \mathbf{R} \to \mathbf{R}$ is absolutely continuous and E is a null set then $F(E) := \{F(x) : x \in E\}$ is also a null set.

(ii) Show that the Cantor function does *not* have this property.

For absolutely continuous functions, we can recover the second fundamental theorem of calculus:

Theorem 1.6.40 (Second fundamental theorem for absolutely continuous functions)**.** *Let $F : [a, b] \to \mathbf{R}$ be absolutely continuous. Then $\int_{[a,b]} F'(x)\ dx = F(b) - F(a)$.*

Proof. Our main tool here will be *Cousin's theorem* (Exercise 1.6.23).

By Exercise 1.6.43, F' is absolutely integrable. By Exercise 1.5.10, F' is thus uniformly integrable. Now let $\varepsilon > 0$. By Exercise 1.5.13, we can find $\kappa > 0$ such that $\int_U |F'(x)|\ dx \leq \varepsilon$ whenever $U \subset [a, b]$ is a measurable set of measure at most κ. (Here we adopt the convention that F' vanishes outside of $[a,b]$.) By making κ small enough, we may also assume from absolute continuity that $\sum_{j=1}^n |F(b_j)-F(a_j)| \leq \varepsilon$ whenever $(a_1,b_1),\ldots,(a_n,b_n)$ is a finite collection of disjoint intervals of total length $\sum_{j=1}^n b_j - a_j$ at most κ.

Let $E \subset [a,b]$ be the set of points x where F is not differentiable, together with the endpoints a,b, as well as the points where x is not a Lebesgue point of F'. thus E is a null set. By outer regularity (or the definition of outer measure) we can find an open set U containing E of measure $m(U) < \kappa$. In particular, $\int_U |F'(x)|\ dx \leq \varepsilon$.

Now define a gauge function $\delta\colon [a,b] \to (0,+\infty)$ as follows.

(i) If $x \in E$, we define $\delta(x) > 0$ to be small enough that the open interval $(x-\delta(x), x+\delta(x))$ lies in U.

(ii) If $x \notin E$, then F is differentiable at x and x is a Lebesgue point of F'. We let $\delta(x) > 0$ be small enough that $|F(y) - F(x) - (y-x)F'(x)| \leq \varepsilon|y-x|$ holds whenever $|y-x| \leq \delta(x)$, and such that $|\frac{1}{|I|}\int_I F'(y)\ dy - F'(x)| \leq \varepsilon$ whenever I is an interval containing x of length at most $\delta(x)$; such a $\delta(x)$ exists by the definition of differentiability, and of Lebesgue point. We rewrite these properties using *big-O notation*[17] as $F(y)-F(x) = (y-x)F'(x) + O(\varepsilon|y-x|)$ and $\int_I F'(y)\ dy = |I|F'(x) + O(\varepsilon|I|)$.

Applying Cousin's theorem, we can find a partition $a = t_0 < t_1 < \ldots < t_k = b$ with $k \geq 1$, together with real numbers $t_j^* \in [t_{j-1}, t_j]$ for each $1 \leq j \leq k$ and $t_j - t_{j-1} \leq \delta(t_j^*)$.

We can express $F(b) - F(a)$ as a telescoping series

$$F(b) - F(a) = \sum_{j=1}^k F(t_j) - F(t_{j-1}).$$

To estimate the size of this sum, let us first consider those j for which $t_j^* \in E$. Then, by construction, the intervals (t_{j-1}, t_j) are disjoint in

[17] In this notation, we use $O(X)$ to denote a quantity Y whose magnitude $|Y|$ is at most CX for some absolute constant C. This notation is convenient for managing error terms when it is not important to keep track of the exact value of constants such as C, due to such rules as $O(X) + O(X) = O(X)$.

U. By construction of κ, we thus have

$$\sum_{j:t_j^* \in E} |F(t_j) - F(t_{j-1})| \leq \varepsilon$$

and thus

$$\sum_{j:t_j^* \in E} F(t_j) - F(t_{j-1}) = O(\varepsilon).$$

Next, we consider those j for which $t_j^* \notin E$. By construction, for those j we have

$$F(t_j) - F(t_j^*) = (t_j - t_j^*)F'(t_j^*) + O(\varepsilon|t_j - t_j^*|)$$

and

$$F(t_j^*) - F(t_{j-1}) = (t_j^* - t_{j-1})F'(t_j^*) + O(\varepsilon|t_j^* - t_{j-1}|)$$

and thus

$$F(t_j) - F(t_{j-1}) = (t_j - t_{j-1})F'(t_j^*) + O(\varepsilon|t_j - t_{j-1}|).$$

On the other hand, from construction again we have

$$\int_{[t_{j-1},t_j]} F'(y)\ dy = (t_j - t_{j-1})F'(t_j^*) + O(\varepsilon|t_j - t_{j-1}|)$$

and thus

$$F(t_j) - F(t_{j-1}) = \int_{[t_{j-1},t_j]} F'(y)\ dy + O(\varepsilon|t_j - t_{j-1}|).$$

Summing in j, we conclude that

$$\sum_{j:t_j^* \notin E} F(t_j) - F(t_{j-1}) = \int_S F'(y)\ dy + O(\varepsilon(b-a)),$$

where S is the union of all the $[t_{j-1}, t_j]$ with $t_j^* \notin E$. By construction, this set is contained in $[a,b]$ and contains $[a,b]\backslash U$. Since $\int_U |F'(x)|\ dx \leq \varepsilon$, we conclude that

$$\int_S F'(y)\ dy = \int_{[a,b]} F'(y)\ dy + O(\varepsilon).$$

Putting everything together, we conclude that

$$F(b) - F(a) = \int_{[a,b]} F'(y)\ dy + O(\varepsilon) + O(\varepsilon|b-a|).$$

Since $\varepsilon > 0$ was arbitrary, the claim follows. $\square$

Combining this result with Exercise 1.6.48, we obtain a satisfactory classification of the absolutely continuous functions:

Exercise 1.6.50. Show that a function $F\colon [a,b] \to \mathbf{R}$ is absolutely continuous if and only if it takes the form $F(x) = \int_{[a,x]} f(y)\ dy + C$ for some absolutely integrable $f\colon [a,b] \to \mathbf{R}$ and a constant C.

Exercise 1.6.51 (Compatibility of the strong and weak derivatives in the absolutely continuous case)**.** Let $F\colon [a,b] \to \mathbf{R}$ be an absolutely continuous function, and let $\phi\colon [a,b] \to \mathbf{R}$ be a continuously differentiable function supported in a compact subset of (a,b). Show that $\int_{[a,b]} F'\phi(x)\ dx = -\int_{[a,b]} F\phi'(x)\ dx$.

Inspecting the proof of Theorem 1.6.40, we see that the absolute continuity was used primarily in two ways: firstly, to ensure the almost everywhere existence, and to control an exceptional null set E. It turns out that one can achieve the latter control by making a different hypothesis, namely that the function F is *everywhere* differentiable rather than merely almost everywhere differentiable. More precisely, we have

Proposition 1.6.41 (Second fundamental theorem of calculus, again)**.** *Let $[a,b]$ be a compact interval of positive length, let $F\colon [a,b] \to \mathbf{R}$ be a differentiable function, such that F' is absolutely integrable. Then the Lebesgue integral $\int_{[a,b]} F'(x)\ dx$ of F' is equal to $F(b) - F(a)$.*

Proof. This will be similar to the proof of Theorem 1.6.40, the one main new twist being that we need several open sets U instead of just one. Let $E \subset [a,b]$ be the set of points x which are not Lebesgue points of F', together with the endpoints a, b. This is a null set. Let $\varepsilon > 0$, and then let $\kappa > 0$ be small enough that $\int_U |F'(x)|\ dx \le \varepsilon$ whenever U is measurable with $m(U) \le \kappa$. We can also ensure that $\kappa \le \varepsilon$.

For every natural number $m = 1, 2, \ldots$ we can find an open set U_m containing E of measure $m(U_m) \le \kappa/4^m$. In particular we see that $m(\bigcup_{m=1}^\infty U_m) \le \kappa$ and thus $\int_{\bigcup_{m=1}^\infty U_m} |F'(x)|\ dx \le \varepsilon$.

Now define a gauge function $\delta\colon [a,b] \to (0,+\infty)$ as follows.

(i) If $x \in E$, we define $\delta(x) > 0$ to be small enough that the open interval $(x - \delta(x), x + \delta(x))$ lies in U_m, where m is the first natural number such that $|F'(x)| \leq 2^m$, and also small enough that $|F(y) - F(x) - (y - x)F'(x)| \leq \varepsilon|y - x|$ holds whenever $|y - x| \leq \delta(x)$. (Here we crucially use the *everywhere* differentiability to ensure that $f'(x)$ exists and is finite here.)

(ii) If $x \notin E$, we let $\delta(x) > 0$ be small enough that $|F(y)-F(x)-(y - x)F'(x)| \leq \varepsilon|y - x|$ holds whenever $|y - x| \leq \delta(x)$, and such that $|\frac{1}{|I|} \int_I F'(y)\ dy - F'(x)| \leq \varepsilon$ whenever I is an interval containing x of length at most $\delta(x)$, exactly as in the proof of Theorem 1.6.40.

Applying Cousin's theorem, we can find a partition $a = t_0 < t_1 < \ldots < t_k = b$ with $k \geq 1$, together with real numbers $t_j^* \in [t_{j-1}, t_j]$ for each $1 \leq j \leq k$ and $t_j - t_{j-1} \leq \delta(t_j^*)$.

As before, we express $F(b) - F(a)$ as a telescoping series

$$F(b) - F(a) = \sum_{j=1}^{k} F(t_j) - F(t_{j-1}).$$

For the contributions of those j with $t_j^* \notin E$, we argue exactly as in the proof of Theorem 1.6.40 to conclude eventually that

$$\sum_{j:t_j^* \notin E} F(t_j) - F(t_{j-1}) = \int_S F'(y)\ dy + O(\varepsilon(b - a)),$$

where S is the union of all the $[t_{j-1}, t_j]$ with $t_j^* \notin E$. Since

$$\int_{[a,b]\backslash S} |F'(x)|\ dx \leq \int_{\bigcup_{m=1}^{\infty} U_m} |F'(x)|\ dx \leq \varepsilon$$

we thus have

$$\int_S F'(y)\ dy = \int_{[a,b]} F'(y)\ dy + O(\varepsilon).$$

Now we turn to those j with $t_j^* \in E$. By construction, we have

$$F(t_j) - F(t_{j-1}) = (t_j - t_{j-1})F'(t_j^*) + O(\varepsilon|t_j - t_{j-1}|)$$

fir these intervals, and so

$$\sum_{j:t_j^*\in E} F(t_j)-F(t_{j-1}) = (\sum_{j:t_j^*\in E}(t_j-t_{j-1})F'(t_j^*)) + O(\varepsilon(b-a)).$$

Next, for each j we have $F'(t_j^*) \leq 2^m$ and $[t_{j-1},t_j]\subset U_m$ for some natural number $m=1,2,\ldots$, by construction. By countable additivity, we conclude that

$$(\sum_{j:t_j^*\in E}(t_j-t_{j-1})F'(t_j^*)) \leq \sum_{m=1}^\infty 2^m m(U_m) \leq \sum_{m=1}^\infty 2^m\varepsilon/4^m = O(\varepsilon).$$

Putting all this together, we again have

$$F(b)-F(a) = \int_{[a,b]} F'(y)\ dy + O(\varepsilon) + O(\varepsilon|b-a|).$$

Since $\varepsilon>0$ was arbitrary, the claim follows. □

Remark 1.6.42. The above proposition is yet another illustration of how the property of everywhere differentiability is significantly better than that of almost everywhere differentiability. In practice, though, the above proposition is not as useful as one might initially think, because there are very few methods that establish the everywhere differentiability of a function that do not also establish continuous differentiability (or at least Riemann integrability of the derivative), at which point one could just use Theorem 1.6.7 instead.

Exercise 1.6.52. Let $F\colon[-1,1]\to\mathbf{R}$ be the function defined by setting $F(x) := x^2\sin(\frac{1}{x^3})$ when x is non-zero, and $F(0):=0$. Show that F is everywhere differentiable, but the deriative F' is not absolutely integrable, and so the second fundamental theorem of calculus does not apply in this case (at least if we interpret $\int_{[a,b]} F'(x)\ dx$ using the absolutely convergent Lebesgue integral). See however the next exercise.

Exercise 1.6.53 (Henstock-Kurzweil integral)**.** Let $[a,b]$ be a compact interval of positive length. We say that a function $f\colon[a,b]\to\mathbf{R}$ is *Henstock-Kurzweil integrable* with integral $L\in\mathbf{R}$ if for every $\varepsilon>0$ there exists a gauge function $\delta\colon[a,b]\to(0,+\infty)$ such that one has

$$|\sum_{j=1}^k f(t_j^*)(t_j-t_{j-1}) - L| \leq \varepsilon$$

whenever $k \geq 1$ and $a = t_0 < t_1 < \ldots < t_k = b$ and $t_1^*, \ldots, t_k^*$ are such that $t_j^* \in [t_{j-1}, t_j]$ and $|t_j - t_{j-1}| \leq \delta(t_j^*)$ for every $1 \leq j \leq k$. When this occurs, we call L the *Henstock-Kurzweil integral* of f and write it as $\int_{[a,b]} f(x)\ dx$.

(i) Show that if a function is Henstock-Kurzweil integrable, it has a unique Henstock-Kurzweil integral. (*Hint:* use Cousin's theorem.)

(ii) Show that if a function is Riemann integrable, then it is Henstock-Kurzweil integrable, and the Henstock-Kurzweil integral $\int_{[a,b]} f(x)\ dx$ is equal to the Riemann integral $\int_a^b f(x)\ dx$.

(iii) Show that if a function $f\colon [a, b] \to \mathbf{R}$ is everywhere defined, everywhere finite, and is absolutely integrable, then it is Henstock-Kurzweil integrable, and the Henstock-Kurzweil integral $\int_{[a,b]} f(x)\ dx$ is equal to the Lebesgue integral $\int_{[a,b]} f(x)\ dx$. (*Hint:* this is a variant of the proof of Theorem 1.6.40 or Proposition 1.6.41.)

(iv) Show that if $F\colon [a, b] \to \mathbf{R}$ is everywhere differentiable, then F' is Henstock-Kurzweil integrable, and the Henstock-Kurzweil integral $\int_{[a,b]} F'(x)\ dx$ is equal to $F(b) - F(a)$. (*Hint:* this is a variant of the proof of Theorem 1.6.40 or Proposition 1.6.41.)

(v) Explain why the above results give an alternate proof of Exercise 1.6.4 and of Proposition 1.6.41.

Remark 1.6.43. As the above exercise indicates, the Henstock-Kurzweil integral (also known as the *Denjoy integral* or *Perron integral*) extends the Riemann integral and the absolutely convergent Lebesgue integral, at least as long as one restricts attention to functions that are defined and are finite everywhere (in contrast to the Lebesgue integral, which is willing to tolerate functions being infinite or undefined so long as this only occurs on a null set). It is the notion of integration that is most naturally associated with the fundamental theorem of calculus for everywhere differentiable functions, as seen in part 4 of the above exercise; it can also be used as a unified framework for all the proofs in this section that invoked Cousin's theorem.

The Henstock-Kurzweil integral can also integrate some (highly oscillatory) functions that the Lebesgue integral cannot, such as the derivative F' of the function F appearing in Exercise 1.6.52. This is analogous to how conditional summation $\lim_{N\to\infty}\sum_{n=1}^N a_n$ can sum conditionally convergent series $\sum_{n=1}^\infty a_n$, even if they are not absolutely integrable. However, much as conditional summation is not always well-behaved with respect to rearrangement, the Henstock-Kurzweil integral does not always react well to changes of variable; also, due to its reliance on the order structure of the real line $\mathbf{R}$, it is difficult to extend the Henstock-Kurzweil integral to more general spaces, such as the Euclidean space $\mathbf{R}^d$, or to abstract measure spaces.

1.7. Outer measures, pre-measures, and product measures

In this text so far, we have focused primarily on one specific example of a countably additive measure, namely *Lebesgue measure*. This measure was constructed from a more primitive concept of *Lebesgue outer measure*, which in turn was constructed from the even more primitive concept of *elementary measure*.

It turns out that both of these constructions can be abstracted. In this section, we will give the *Carathéodory extension theorem*, which constructs a countably additive measure from any abstract outer measure; this generalises the construction of Lebesgue measure from Lebesgue outer measure. One can in turn construct outer measures from another concept known as a *pre-measure*, of which elementary measure is a typical example.

With these tools, one can start constructing many more measures, such as *Lebesgue-Stieltjes measures*, *product measures*, and *Hausdorff measures*. With a little more effort, one can also establish the *Kolmogorov extension theorem*, which allows one to construct a variety of measures on infinite-dimensional spaces, and is of particular importance in the foundations of probability theory, as it allows one to set up probability spaces associated to both discrete and continuous random processes, even if they have infinite length.

The most important result about product measure, beyond the fact that it exists, is that one can use it to evaluate iterated integrals, and to interchange their order, provided that the integrand is either unsigned or absolutely integrable. This fact is known as the *Fubini-Tonelli theorem*, and is an absolutely indispensable tool for computing integrals, and for deducing higher-dimensional results from lower-dimensional ones.

In this section we will however omit a very important way to construct measures, namely the *Riesz representation theorem*, which is discussed in §1.10 of *An epsilon of room, Vol. I.*

1.7.1. Outer measures and the Carathéodory extension theorem.

We begin with the abstract concept of an outer measure.

Definition 1.7.1 (Abstract outer measure)**.** Let X be a set. An *abstract outer measure* (or *outer measure* for short) is a map $\mu^*: 2^X \to [0, +\infty]$ that assigns an unsigned extended real number $\mu^*(E) \in [0, +\infty]$ to *every* set $E \subset X$ which obeys the following axioms:

(i) (Empty set) $\mu^*(\emptyset) = 0$.

(ii) (Monotonicity) If $E \subset F$, then $\mu^*(E) \leq \mu^*(F)$.

(iii) (Countable subadditivity) If $E_1, E_2, \ldots \subset X$ is a countable sequence of subsets of X, then $\mu^*(\bigcup_{n=1}^\infty E_n) \leq \sum_{n=1}^\infty \mu^*(E_n)$.

Outer measures are also known as *exterior measures.*

Thus, for instance, Lebesgue outer measure m^* is an outer measure (see Exercise 1.2.3). On the other hand, Jordan outer measure $m^{*,(J)}$ is only finitely subadditive rather than countably subadditive and thus is not, strictly speaking, an outer measure; for this reason this concept is often referred to as *Jordan outer content* rather than *Jordan outer measure.*

Note that outer measures are weaker than measures in that they are merely countably subadditive, rather than countably additive. On the other hand, they are able to measure *all* subsets of X, whereas measures can only measure a σ-algebra of measurable sets.

In Definition 1.2.2, we used Lebesgue outer measure together with the notion of an open set to define the concept of Lebesgue measurability. This definition is not available in our more abstract setting, as we do not necessarily have the notion of an open set. An alternative definition of measurability was put forth in Exercise 1.2.17, but this still required the notion of a box or an elementary set, which is still not available in this setting. Nevertheless, we can modify that definition to give an abstract definition of measurability:

Definition 1.7.2 (Carathéodory measurability)**.** Let μ^* be an outer measure on a set X. A set $E \subset X$ is said to be *Carathéodory measurable* with respect to μ^* if one has

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

for *every* set $A \subset X$.

Exercise 1.7.1 (Null sets are Carathéodory measurable)**.** Suppose that E is a null set for an outer measure μ^* (i.e. $\mu^*(E) = 0$). Show that E is Carathéodory measurable with respect to μ^*.

Exercise 1.7.2 (Compatibility with Lebesgue measurability)**.** Show that a set $E \subset \mathbf{R}^d$ is Carathéodory measurable with respect to Lebesgue outer measurable if and only if it is Lebesgue measurable. (*Hint:* one direction follows from Exercise 1.2.17. For the other direction, first verify simple cases, such as when E is a box, or when E or A are bounded.)

The construction of Lebesgue measure can then be abstracted as follows:

Theorem 1.7.3 (Carathéodory extension theorem)**.** *Let* $\mu^* \colon 2^X \to [0, +\infty]$ *be an outer measure on a set* X*, let* $\mathcal{B}$ *be the collection of all subsets of* X *that are Carathéodory measurable with respect to* μ^**, and let* $\mu \colon \mathcal{B} \to [0, +\infty]$ *be the restriction of* μ^* *to* $\mathcal{B}$ *(thus* $\mu(E) := \mu^*(E)$ *whenever* $E \in \mathcal{B}$*). Then* $\mathcal{B}$ *is a* σ*-algebra, and* μ *is a measure.*

Proof. We begin with the σ-algebra property. It is easy to see that the empty set lies in $\mathcal{B}$, and that the complement of a set in $\mathcal{B}$ lies in $\mathcal{B}$ also. Next, we verify that $\mathcal{B}$ is closed under finite unions (which will make $\mathcal{B}$ a Boolean algebra). Let $E, F \in \mathcal{B}$, and let $A \subset X$ be arbitrary. By definition, it suffices to show that

$$\mu^*(A) = \mu^*(A \cap (E \cup F)) + \mu^*(A \backslash (E \cup F)). \tag{1.31}$$

To simplify the notation, we partition A into the four disjoint sets

$$\begin{aligned} A_{00} &:= A \backslash (E \cup F); \\ A_{10} &:= (A \backslash F) \cap E; \\ A_{01} &:= (A \backslash E) \cap F; \\ A_{11} &:= A \cap E \cap F \end{aligned}$$

(the reader may wish to draw a Venn diagram here to understand the nature of these sets). Thus (1.31) becomes

$$\mu^*(A_{00} \cup A_{01} \cup A_{10} \cup A_{11}) = \mu^*(A_{01} \cup A_{10} \cup A_{11}) + \mu^*(A_{00}). \tag{1.32}$$

On the other hand, from the Carathéodory measurability of E, one has

$$\mu^*(A_{00} \cup A_{01} \cup A_{10} \cup A_{11}) = \mu^*(A_{00} \cup A_{01}) + \mu^*(A_{10} \cup A_{11})$$

and

$$\mu^*(A_{01} \cup A_{10} \cup A_{11}) = \mu^*(A_{01}) + \mu^*(A_{10} \cup A_{11})$$

while from the Carathéodory measurability of F one has

$$\mu^*(A_{00} \cup A_{01}) = \mu^*(A_{00}) + \mu^*(A_{01});$$

putting these identities together we obtain (1.32). (Note that no subtraction is employed here, and so the arguments still work when some sets have infinite outer measure.)

Now we verify that $\mathcal{B}$ is a σ-algebra. As it is already a Boolean algebra, it suffices (see Exercise 1.7.3 below) to verify that $\mathcal{B}$ is closed with respect to countable disjoint unions. Thus, let $E_1, E_2, \ldots$ be a disjoint sequence of Carathéodory-measurable sets, and let A be arbitrary. We wish to show that

$$\mu^*(A) = \mu^*(A \cap \bigcup_{n=1}^\infty E_n) + \mu^*(A \backslash \bigcup_{n=1}^\infty E_n).$$

In view of subadditivity, it suffices to show that

$$\mu^*(A) \geq \mu^*(A \cap \bigcup_{n=1}^\infty E_n) + \mu^*(A \backslash \bigcup_{n=1}^\infty E_n).$$

For any $N \geq 1$, $\bigcup_{n=1}^N E_n$ is Carathéodory measurable (as $\mathcal{B}$ is a Boolean algebra), and so

$$\mu^*(A) \geq \mu^*(A \cap \bigcup_{n=1}^N E_n) + \mu^*(A \backslash \bigcup_{n=1}^N E_n).$$

By monotonicity, $\mu^*(A \backslash \bigcup_{n=1}^N E_n) \geq \mu^*(A \backslash \bigcup_{n=1}^\infty E_n)$. Taking limits as $N \to \infty$, it thus suffices to show that

$$\mu^*(A \cap \bigcup_{n=1}^\infty E_n) \leq \lim_{N \to \infty} \mu^*(A \cap \bigcup_{n=1}^N E_n).$$

But by the Carathéodory measurability of $\bigcup_{n=1}^N E_n$, we have

$$\mu^*(A \cap \bigcup_{n=1}^{N+1} E_n) = \mu^*(A \cap \bigcup_{n=1}^N E_n) + \mu^*(A \cap E_{N+1} \backslash \bigcup_{n=1}^N E_n)$$

for any $N \geq 0$, and thus on iteration

$$\lim_{N \to \infty} \mu^*(A \cap \bigcup_{n=1}^N E_n) = \sum_{N=0}^\infty \mu^*(A \cap E_{N+1} \backslash \bigcup_{n=1}^N E_n)$$

On the other hand, from countable subadditivity one has

$$\mu^*(A \cap \bigcup_{n=1}^\infty E_n) \leq \sum_{N=0}^\infty \mu^*(A \cap E_{N+1} \backslash \bigcup_{n=1}^N E_n)$$

and the claim follows.

Finally, we show that μ is a measure. It is clear that $\mu(\emptyset) = 0$, so it suffices to establish countable additivity, thus we need to show that

$$\mu^*(\bigcup_{n=1}^\infty E_n) = \sum_{n=1}^\infty \mu^*(E_n)$$

whenever $E_1, E_2, \ldots$ are Carathéodory-measurable and disjoint. By subadditivity it suffices to show that

$$\mu^*(\bigcup_{n=1}^\infty E_n) \geq \sum_{n=1}^\infty \mu^*(E_n).$$

By monotonicity it suffices to show that

$$\mu^*(\bigcup_{n=1}^N E_n) = \sum_{n=1}^N \mu^*(E_n)$$

for any finite N. But from the Carathéodory measurability of $\bigcup_{n=1}^N E_n$ one has

$$\mu^*(\bigcup_{n=1}^{N+1} E_n) = \mu^*(\bigcup_{n=1}^{N} E_n) + \mu^*(E_{N+1})$$

for any $N \geq 0$, and the claim follows from induction. □

Exercise 1.7.3. Let $\mathcal{B}$ be a Boolean algebra on a set X. Show that $\mathcal{B}$ is a σ-algebra if and only if it is closed under countable *disjoint* unions, which means that $\bigcup_{n=1}^\infty E_n \in \mathcal{B}$ whenever $E_1, E_2, E_3, \ldots \in \mathcal{B}$ are a countable sequence of *disjoint* sets in $\mathcal{B}$.

Remark 1.7.4. Note that the above theorem, combined with Exercise 1.7.2 gives a slightly alternate way to construct Lebesgue measure from Lebesgue outer measure than the construction given in Section 1.2. This is arguably a more efficient way to proceed, but is also less geometrically intuitive than the approach taken in Section 1.2.

Remark 1.7.5. From Exercise 1.7.1 we see that the measure μ constructed by the Carathéodory extension theorem is automatically *complete* (see Definition 1.4.31).

Remark 1.7.6. In §1.15 of *An epsilon of room, Vol. I*, an important example of a measure constructed by Carathéodory's theorem is given, namely the d-dimensional *Hausdorff measure* $\mathcal{H}^d$ on $\mathbf{R}^n$ that is good for measuring the size of d-dimensional subsets of $\mathbf{R}^n$.

1.7.2. Pre-measures.

In previous notes, we saw that finitely additive measures, such as elementary measure or Jordan measure, could be extended to a countably additive measure, namely Lebesgue measure. It is natural to ask whether this property is true in general. In other words, given a finitely additive measure $\mu_0 \colon \mathcal{B}_0 \to [0, +\infty]$ on a Boolean algebra $\mathcal{B}_0$, is it possible to find a σ-algebra $\mathcal{B}$ refining $\mathcal{B}_0$, and a countably additive measure $\mu \colon \mathcal{B} \to [0, +\infty]$ that extends μ_0?

There is an obvious necessary condition in order for μ_0 to have a countably additive extension, namely that μ_0 already has to be countably additive *within* $\mathcal{B}_0$. More precisely, suppose that $E_1, E_2, E_3, \ldots \in \mathcal{B}_0$ were disjoint sets such that their union $\bigcup_{n=1}^\infty E_n$ was also in $\mathcal{B}_0$. (Note that this latter property is not automatic as $\mathcal{B}_0$ is merely a Boolean algebra rather than a σ-algebra.) Then, in order for μ_0 to be extendible to a countably additive measure, it is clearly necessary that

$$\mu_0(\bigcup_{n=1}^\infty E_n) = \sum_{n=1}^\infty \mu_0(E_n).$$

Using the Carathéodory extension theorem, we can show that this necessary condition is also sufficient. More precisely, we have

Definition 1.7.7 (Pre-measure)**.** A *pre-measure* on a Boolean algebra $\mathcal{B}_0$ is a finitely additive measure $\mu_0 \colon \mathcal{B}_0 \to [0, +\infty]$ with the property that $\mu_0(\bigcup_{n=1}^\infty E_n) = \sum_{n=1}^\infty \mu_0(E_n)$ whenever $E_1, E_2, E_3, \ldots \in \mathcal{B}_0$ are disjoint sets such that $\bigcup_{n=1}^\infty E_n$ is in $\mathcal{B}_0$.

Exercise 1.7.4.

(i) Show that the requirement that μ_0 is finitely additive can be relaxed to the condition that $\mu_0(\emptyset) = 0$ without affecting the definition of a pre-measure.

(ii) Show that the condition $\mu_0(\bigcup_{n=1}^\infty E_n) = \sum_{n=1}^\infty \mu_0(E_n)$ can be relaxed to $\mu_0(\bigcup_{n=1}^\infty E_n) \leq \sum_{n=1}^\infty \mu_0(E_n)$ without affecting the definition of a pre-measure.

(iii) On the other hand, give an example to show that if one performs both of the above two relaxations at once, one starts admitting objects μ_0 that are not pre-measures.

Exercise 1.7.5. Without using the theory of Lebesgue measure, show that elementary measure (on the elementary Boolean algebra) is a pre-measure. (*Hint:* use Lemma 1.2.6. Note that one has to also deal with co-elementary sets as well as elementary sets in the elementary Boolean algebra.)

Exercise 1.7.6. Construct a finitely additive measure $\mu_0 \colon \mathcal{B}_0 \to [0, +\infty]$ that is not a pre-measure. (*Hint:* take X to be the natural numbers, take $\mathcal{B}_0 = 2^{\mathbf{N}}$ to be the discrete algebra, and define μ_0 separately for finite and infinite sets.)

Theorem 1.7.8 (Hahn-Kolmogorov theorem)**.** *Every pre-measure $\mu_0 \colon \mathcal{B}_0 \to [0, +\infty]$ on a Boolean algebra $\mathcal{B}_0$ in X can be extended to a countably additive measure $\mu \colon \mathcal{B} \to [0, +\infty]$.*

Proof. We mimic the construction of Lebesgue measure from elementary measure. Namely, for any set $E \subset X$, define the *outer measure* $\mu^*(E)$ of E to be the quantity

$$\mu^*(E) := \inf\{\sum_{n=1}^{\infty} \mu_0(E_n) \colon E \subset \bigcup_{n=1}^{\infty} E_n; E_n \in \mathcal{B}_0 \text{ for all } n\}.$$

It is easy to verify (cf. Exercise 1.2.3) that μ^* is indeed an outer measure. Let $\mathcal{B}$ be the collection of all sets $E \subset X$ that are Carathéodory measurable with respect to μ^*, and let μ be the restriction of μ^* to $\mathcal{B}$. By the Carathéodory extension theorem, $\mathcal{B}$ is a σ-algebra and μ is a countably additive measure.

It remains to show that $\mathcal{B}$ contains $\mathcal{B}_0$ and that μ extends μ_0. Thus, let $E \in \mathcal{B}_0$; we need to show that E is Carathéodory measurable with respect to μ^* and that $\mu^*(E) = \mu_0(E)$. To prove the first claim, let $A \subset X$ be arbitrary. We need to show that

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

by subadditivity, it suffices to show that

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

We may assume that $\mu^*(A)$ is finite, since the claim is trivial otherwise.

Fix $\varepsilon > 0$. By definition of μ^*, one can find $E_1, E_2, \ldots \in \mathcal{B}_0$ covering A such that

$$\sum_{n=1}^{\infty} \mu_0(E_n) \leq \mu^*(A) + \varepsilon.$$

The sets $E_n \cap E$ lie in $\mathcal{B}_0$ and cover $A \cap E$ and thus

$$\mu^*(A \cap E) \leq \sum_{n=1}^{\infty} \mu_0(E_n \cap E).$$

Similarly we have

$$\mu^*(A \backslash E) \leq \sum_{n=1}^{\infty} \mu_0(E_n \backslash E).$$

Meanwhile, from finite additivity we have

$$\mu_0(E_n \cap E) + \mu_0(E_n \backslash E) = \mu_0(E_n).$$

Combining all of these estimates, we obtain

$$\mu^*(A \cap E) + \mu^*(A \backslash E) \leq \mu^*(A) + \varepsilon;$$

since $\varepsilon > 0$ was arbitrary, the claim follows.

Finally, we have to show that $\mu^*(E) = \mu_0(E)$. Since E covers itself, we certainly have $\mu^*(E) \leq \mu_0(E)$. To show the converse inequality, it suffices to show that

$$\sum_{n=1}^{\infty} \mu_0(E_n) \geq \mu_0(E)$$

whenever $E_1, E_2, \ldots \in \mathcal{B}_0$ cover E. By replacing each E_n with the smaller set $E_n \backslash \bigcup_{m=1}^{n-1} E_m$ (which still lies in $\mathcal{B}_0$, and still covers E), we may assume without loss of generality (thanks to the monotonicity of μ_0) that the E_n are disjoint. Similarly, by replacing each E_n with the smaller set $E_n \cap E$ we may assume without loss of generality that the union of the E_n is exactly equal to E. But then the claim follows from the hypothesis that μ_0 is a pre-measure (and not merely a finitely additive measure). □

Let us call the measure μ constructed in the above proof the *Hahn-Kolmogorov extension* of the pre-measure μ_0. Thus, for instance, from Exercise 1.7.2, the Hahn-Kolmogorov extension of elementary measure (with the convention that co-elementary sets have infinite elementary measure) is Lebesgue measure. This is not quite the unique extension of μ_0 to a countably additive measure, though. For instance, one could restrict Lebesgue measure to the Borel σ-algebra, and this would still be a countably additive extension of elementary measure. However, the extension is unique within its own σ-algebra:

Exercise 1.7.7. Let $\mu_0 \colon \mathcal{B}_0 \to [0, +\infty]$ be a pre-measure, let $\mu \colon \mathcal{B} \to [0, +\infty]$ be the Hahn-Kolmogorov extension of μ_0, and let $\mu' \colon \mathcal{B}' \to [0, +\infty]$ be another countably additive extension of μ_0. Suppose also that μ_0 is *σ-finite*, which means that one can express the whole space X as the countable union of sets $E_1, E_2, \ldots \in \mathcal{B}_0$ for which $\mu_0(E_n) < \infty$ for all n. Show that μ and μ' agree on their common domain of definition. In other words, show that $\mu(E) = \mu'(E)$ for all $E \in \mathcal{B} \cap \mathcal{B}'$. (*Hint:* first show that $\mu'(E) \leq \mu^*(E)$ for all $E \in \mathcal{B}'$.)

Exercise 1.7.8. The purpose of this exercise is to show that the σ-finite hypothesis in Exercise 1.7.7 cannot be removed. Let $\mathcal{A}$ be the collection of all subsets in $\mathbf{R}$ that can be expressed as finite unions of half-open intervals $[a, b)$. Let $\mu_0 \colon \mathcal{A} \to [0, +\infty]$ be the function such that $\mu_0(E) = +\infty$ for non-empty E and $\mu_0(\emptyset) = 0$.

(i) Show that μ_0 is a pre-measure.

(ii) Show that $\langle \mathcal{A} \rangle$ is the Borel σ-algebra $\mathcal{B}[\mathbf{R}]$.

(iii) Show that the Hahn-Kolmogorov extension $\mu \colon \mathcal{B}[\mathbf{R}] \to [0, +\infty]$ of μ_0 assigns an infinite measure to any non-empty Borel set.

(iv) Show that counting measure $\#$ (or more generally, $c\#$ for any $c \in (0, +\infty]$) is another extension of μ_0 on $\mathcal{B}[\mathbf{R}]$.

Exercise 1.7.9. Let $\mu_0 \colon \mathcal{B}_0 \to [0, +\infty]$ be a pre-measure which is σ-finite (thus X is the countable union of sets in $\mathcal{B}_0$ of finite μ_0-measure), and let $\mu \colon \mathcal{B} \to [0, +\infty]$ be the Hahn-Kolmogorov extension of μ_0.

(i) Show that if $E \in \mathcal{B}$, then there exists $F \in \langle \mathcal{B}_0 \rangle$ containing E such that $\mu(F \backslash E) = 0$ (thus F consists of the union of E and a null set). Furthermore, show that F can be chosen to be a countable intersection $F = \bigcap_{n=1}^\infty F_n$ of sets F_n, each of which is a countable union $F_n = \bigcup_{m=1}^\infty F_{n,m}$ of sets $F_{n,m}$ in $\mathcal{B}_0$.

(ii) If $E \in \mathcal{B}$ has finite measure (i.e. $\mu(E) < \infty$), and $\varepsilon > 0$, show that there exists $F \in \mathcal{B}_0$ such that $\mu(E \Delta F) \leq \varepsilon$.

(iii) Conversely, if E is a set such that for every $\varepsilon > 0$ there exists $F \in \mathcal{B}_0$ such that $\mu^*(E \Delta F) \leq \varepsilon$, show that $E \in \mathcal{B}$.

1.7.3. Lebesgue-Stieltjes measure.

Now we use the Hahn-Kolmogorov extension theorem to construct a variety of measures. We begin with Lebesgue-Stieltjes measure.

Theorem 1.7.9 (Existence of Lebesgue-Stieltjes measure). *Let $F \colon \mathbf{R} \to \mathbf{R}$ be a monotone non-decreasing function, and define the left and right limits*

$$F_-(x) := \sup_{y<x} F(y); \quad F_+(x) := \inf_{y>x} F(y),$$

thus one has $F_-(x) \leq F(x) \leq F_+(x)$ for all x. Let $\mathcal{B}[\mathbf{R}]$ be the Borel σ-algebra on $\mathbf{R}$. Then there exists a unique Borel measure

$\mu_F\colon \mathcal{B}[\mathbf{R}] \to [0,+\infty]$ *such that*

$$\mu_F([a,b]) = F_+(b) - F_-(a), \mu_F([a,b)) = F_-(b) - F_-(a), \tag{1.33}$$

$$\mu_F((a,b]) = F_+(b) - F_+(a), \mu_F((a,b)) = F_-(b) - F_+(a)$$

for all $-\infty < b < a < \infty$*, and*

$$\mu_F(\{a\}) = F_+(a) - F_-(a) \tag{1.34}$$

for all $a \in \mathbf{R}$.

Proof. (Sketch) For this proof, we will deviate from our previous notational conventions, and allow intervals to be unbounded, thus in particular including the half-infinite intervals $[a,+\infty)$, $(a,+\infty)$, $(-\infty,a]$, $(-\infty,a)$ and the doubly infinite interval $(-\infty,+\infty)$ as intervals.

Define the F-volume $|I|_F \in [0,+\infty]$ of any interval I, adopting the obvious conventions that $F_-(+\infty) = \sup_{y\in\mathbf{R}} F(y)$ and $F_+(-\infty) = \inf_{y\in\mathbf{R}} F(y)$, and also adopting the convention that the empty interval $\emptyset$ has zero F-volume, $|\emptyset|_F = 0$. Note that $F_-(+\infty)$ could equal $+\infty$ and $F_+(-\infty)$ could equal $-\infty$, but in all circumstances the F-volume $|I|_F$ is well-defined and takes values in $[0,+\infty]$, after adopting the obvious conventions to evaluate expressions such as $+\infty - (-\infty)$.

A somewhat tedious case check (Exercise!) gives the additivity property

$$|I \cup J|_F = |I|_F + |J|_F$$

whenever I, J are disjoint intervals that share a common endpoint. As a corollary, we see that if a interval I is partitioned into finitely many disjoint sub-intervals $I_1,\ldots,I_k$, we have $|I| = |I_1| + \ldots + |I_k|$.

Let $\mathcal{B}_0$ be the Boolean algebra generated by the (possibly infinite) intervals, then $\mathcal{B}_0$ consists of those sets that can be expressed as a finite union of intervals. (This is slightly larger than the elementary algebra, as it allows for half-infinite intervals such as $[0,+\infty)$, whereas the elementary algebra does not.) We can define a measure μ_0 on this algebra by declaring

$$\mu_0(E) = |I_1|_F + \ldots + |I_k|_F$$

whenever $E = I_1 \cup \ldots \cup I_k$ is the disjoint union of finitely many intervals. One can check (Exercise!) that this measure is well-defined (in the sense that it gives a unique value to $\mu_0(E)$ for each $E \in \mathcal{B}_0$) and is finitely additive. We now claim that μ_0 is a pre-measure: thus we suppose that $E = \mathcal{B}_0$ is the disjoint union of countably many sets $E_1, E_2, \ldots \in \mathcal{B}_0$, and wish to show that

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

By splitting up E into intervals and then intersecting each of the E_n with these intervals and using finite additivity, we may assume that E is a single interval. By splitting up the E_n into their component intervals and using finite additivity, we may assume that the E_n are also individual intervals. By subadditivity, it suffices to show that

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

By the definition of $\mu_0(E)$, one can check that

$$\mu_0(E) = \sup_{K \subset E} \mu_0(K) \tag{1.35}$$

where K ranges over all compact intervals contained in E (Exercise!). Thus, it suffices to show that

$$\mu_0(K) \leq \sum_{n=1}^{\infty} \mu_0(E_n)$$

for each compact sub-interval K of E. In a similar spirit, one can show that

$$\mu_0(E_n) = \inf_{U \supset E_n} \mu_0(E_n)$$

where U ranges over all open intervals containing E_n (Exercise!). Using the $\varepsilon/2^n$ trick, it thus suffices to show that

$$\mu_0(K) \leq \sum_{n=1}^{\infty} \mu_0(U_n)$$

whenever U_n is an open interval containing E_n. But by the Heine-Borel theorem, one can cover K by a finite number $\bigcup_{n=1}^{N} U_n$ of the U_n, hence by finite subadditivity

$$\mu_0(K) \leq \sum_{n=1}^{N} \mu_0(U_n)$$

and the claim follows.

As μ_0 is now verified to be a pre-measure, we may use the Hahn-Kolmogorov extension theorem to extend it to a countably additive measure μ on a σ-algebra $\mathcal{B}$ that contains $\mathcal{B}_0$. In particular, $\mathcal{B}$ contains all the elementary sets and hence (by Exercise 1.4.14) contains the Borel σ-algebra. Restricting μ to the Borel σ-algebra we obtain the existence claim.

Finally, we establish uniqueness. If μ' is another Borel measure with the stated properties, then $\mu'(K) = |K|_F$ for every compact interval K, and hence by (1.35) and upward monotone convergence, one has $\mu'(I) = |I|_F$ for every interval (including the unbounded ones). This implies that μ' agrees with μ_0 on $\mathcal{B}_0$, and thus (by Exercise 1.7.7, noting that μ_0 is σ-finite) agrees with μ on Borel measurable sets. $\square$

Exercise 1.7.10. Verify the claims marked "Exercise!" in the above proof.

The measure μ_F given by the above theorem is known as the *Lebesgue-Stieltjes measure* μ_F of F. (In some texts, this measure is only defined when F is right-continuous, or equivalently if $F = F_+$.)

Exercise 1.7.11. Define a *Radon measure* on $\mathbf{R}$ to be a Borel measure μ obeying the following additional properties:

(i) (Local finiteness) $\mu(K) < \infty$ for every compact K.

(ii) (Inner regularity) One has $\mu(E) = \sup_{K \subset E, K \text{ compact}} \mu(K)$ for every Borel set E.

(iii) (Outer regularity) One has $\mu(E) = \inf_{U \supset E, U \text{ open}} \mu(U)$ for every Borel set E.

Show that for every monotone function $F\colon \mathbf{R} \to \mathbf{R}$, the Lebesgue-Stieltjes measure μ_F is a Radon measure on $\mathbf{R}$; conversely, if μ is a Radon measure on $\mathbf{R}$, show that there exists a monotone function $F\colon \mathbf{R} \to \mathbf{R}$ such that $\mu = \mu_F$.

Radon measures are studied in more detail in §1.10 of *An epsilon of room, Vol. I.*

Exercise 1.7.12 (Near uniqueness)**.** If $F, F' : \mathbf{R} \to \mathbf{R}$ are monotone non-decreasing functions, show that $\mu_F = \mu_{F'}$ if and only if there exists a constant $C \in \mathbf{R}$ such that $F_+(x) = F'_+(x) + C$ and $F_-(x) = F'_-(x) + C$ for all $x \in \mathbf{R}$. Note that this implies that the value of F at its points of discontinuity are irrelevant for the purposes of determining the Lebesgue-Stieltjes measure μ_F; in particular, $\mu_F = \mu_{F_+} = \mu_{F_-}$.

In the special case when $F_+(-\infty) = 0$ and $F_-(+\infty) = 1$, then μ_F is a probability measure, and $F_+(x) = \mu_F((-\infty, x])$ is known as the *cumulative distribution function* of μ_F.

Now we give some examples of Lebesgue-Stieltjes measure.

Exercise 1.7.13 (Lebesgue-Stieltjes measure, absolutely continuous case)**.**

(i) If $F: \mathbf{R} \to \mathbf{R}$ is the identity function $F(x) = x$, show that μ_F is equal to Lebesgue measure m.

(ii) If $F: \mathbf{R} \to \mathbf{R}$ is monotone non-decreasing and absolutely continuous (which in particular implies that F' exists and is absolutely integrable, show that $\mu_F = m_{F'}$ in the sense of Exercise 1.4.49, thus

$$\mu_F(E) = \int_E F'(x)\ dx$$

for any Borel measurable E, and

$$\int_{\mathbf{R}} f(x)\ d\mu_F(x) = \int_{\mathbf{R}} f(x) F'(x)\ dx$$

for any unsigned Borel measurable $f: \mathbf{R} \to [0, +\infty]$.

In view of the above exercise, the integral $\int_{\mathbf{R}} f\ d\mu_F$ is often abbreviated $\int_{\mathbf{R}} f\ dF$, and referred to as the *Lebesgue-Stieltjes integral* of f with respect to F. In particular, observe the identity

$$\int_{[a,b]} dF = F_+(b) - F_-(a)$$

for any monotone non-decreasing $F: \mathbf{R} \to \mathbf{R}$ and any $-\infty < b < a < +\infty$, which can be viewed as yet another formulation of the fundamental theorem of calculus.

Exercise 1.7.14 (Lebesgue-Stieltjes measure, pure point case)**.**

(i) If $H \colon \mathbf{R} \to \mathbf{R}$ is the *Heaviside function* $H := 1_{[0,+\infty)}$, show that μ_H is equal to the Dirac measure δ_0 at the origin (defined in Example 1.4.22).

(ii) If $F = \sum_n c_n J_n$ is a jump function (as defined in Definition 1.6.30), show that μ_F is equal to the linear combination $\sum c_n \delta_{x_n}$ of delta functions (as defined in Exercise 1.4.22), where x_n is the point of discontinuity for the basic jump function J_n.

Exercise 1.7.15 (Lebesgue-Stieltjes measure, singular continuous case)**.**

(i) If $F \colon \mathbf{R} \to \mathbf{R}$ is a monotone non-decreasing function, show that F is continuous if and only if $\mu_F(\{x\}) = 0$ for all $x \in \mathbf{R}$.

(ii) If F is the Cantor function (defined in Exercise 1.6.47), show that μ_F is a probability measure supported on the middle-thirds Cantor set (see Exercise 1.2.9) in the sense that $\mu_F(\mathbf{R} \backslash C) = 0$. The measure μ_F is known as *Cantor measure*.

(iii) If μ_F is Cantor measure, establish the self-similarity properties $\mu(\frac{1}{3} \cdot E) = \frac{1}{2}\mu(E)$ and $\mu(\frac{1}{3} \cdot E + \frac{2}{3}) = \frac{1}{2}\mu(E)$ for every Borel-measurable $E \subset [0,1]$, where $\frac{1}{3} \cdot E := \{\frac{1}{3}x : x \in E\}$.

Exercise 1.7.16 (Connection with Riemann-Stieltjes integral)**.** Let $F \colon \mathbf{R} \to \mathbf{R}$ be monotone non-decreasing, let $[a,b]$ be a compact interval, and let $f \colon [a,b] \to \mathbf{R}$ be continuous. Suppose that F is continuous at the endpoints a, b of the interval. Show that for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|\sum_{i=1}^n f(t_i^*)(F(t_i) - F(t_{i-1})) - \int_{[a,b]} f\ dF| \leq \varepsilon$$

whenever $a = t_0 < t_1 < \ldots < t_n = b$ and $t_i^* \in [t_{i-1}, t_i]$ for $1 \leq i \leq n$ are such that $\sup_{1 \leq i \leq n} |t_i - t_{i-1}| \leq \delta$. In the language of the *Riemann-Stieltjes integral*, this result asserts that the Lebesgue-Stieltjes integral extends the Riemann-Stieltjes integral.

Exercise 1.7.17 (Integration by parts formula)**.** Let $F, G\colon \mathbf{R} \to \mathbf{R}$ be monotone non-decreasing and continuous. Show that

$$\int_{[a,b]} F\ dG = -\int_{[a,b]} G\ dF + F(b)G(b) - F(a)G(a)$$

for any compact interval $[a,b]$. (*Hint:* use Exercise 1.7.16.) This formula can be partially extended to the case when one or both of F, G have discontinuities, but care must be taken when F and G are simultaneously discontinuous at the same location.

1.7.4. Product measure. Given two sets X and Y, one can form their Cartesian product $X \times Y = \{(x,y) : x \in X, y \in Y\}$. This set is naturally equipped with the coordinate projection maps $\pi_X\colon X \times Y \to X$ and $\pi_Y\colon X \times Y \to Y$ defined by setting $\pi_X(x,y) := x$ and $\pi_Y(x,y) := y$. One can certainly take Cartesian products $X_1 \times \ldots \times X_d$ of more than two sets, or even take an infinite product $\prod_{\alpha \in A} X_\alpha$, but for simplicity we will only discuss the theory for products of two sets for now.

Now suppose that $(X, \mathcal{B}_X)$ and $(Y, \mathcal{B}_Y)$ are measurable spaces. Then we can still form the Cartesian product $X \times Y$ and the projection maps $\pi_X\colon X \times Y \to X$ and $\pi_Y\colon X \times Y \to Y$. But now we can also form the *pullback* σ-algebras

$$\pi_X^*(\mathcal{B}_X) := \{\pi_X^{-1}(E) : E \in \mathcal{B}_X\} = \{E \times Y : E \in \mathcal{B}_X\}$$

and

$$\pi_Y^*(\mathcal{B}_Y) := \{\pi_Y^{-1}(E) : E \in \mathcal{B}_Y\} = \{X \times F : F \in \mathcal{B}_Y\}.$$

We then define the *product σ-algebra* $\mathcal{B}_X \times \mathcal{B}_Y$ to be the σ-algebra generated by the union of these two σ-algebras:

$$\mathcal{B}_X \times \mathcal{B}_Y := \langle \pi_X^*(\mathcal{B}_X) \cup \pi_Y^*(\mathcal{B}_Y) \rangle.$$

This definition has several equivalent formulations:

Exercise 1.7.18. Let $(X, \mathcal{B}_X)$ and $(Y, \mathcal{B}_Y)$ be measurable spaces.

(i) Show that $\mathcal{B}_X \times \mathcal{B}_Y$ is the σ-algebra generated by the sets $E \times F$ with $E \in \mathcal{B}_X$, $Y \in \mathcal{B}_Y$. In other words, $\mathcal{B}_X \times \mathcal{B}_Y$ is the coarsest σ-algebra on $X \times Y$ with the property that the product of a $\mathcal{B}_X$-measurable set and a $\mathcal{B}_Y$-measurable set is always $\mathcal{B}_X \times \mathcal{B}_Y$ measurable.

(ii) Show that $\mathcal{B}_X \times \mathcal{B}_Y$ is the coarsest σ-algebra on $X \times Y$ that makes the projection maps π_X, π_Y both measurable morphisms (see Remark 1.4.33).

(iii) If $E \in \mathcal{B}_X \times \mathcal{B}_Y$, show that the sets $E_x := \{y \in Y : (x, y) \in E\}$ lie in $\mathcal{B}_Y$ for every $x \in X$, and similarly that the sets $E^y := \{x \in X : (x, y) \in E\}$ lie in $\mathcal{B}_X$ for every $y \in Y$.

(iv) If $f \colon X \times Y \to [0, +\infty]$ is measurable (with respect to $\mathcal{B}_X \times \mathcal{B}_Y$), show that the function $f_x : y \mapsto f(x, y)$ is $\mathcal{B}_Y$-measurable for every $x \in X$, and similarly that the function $f^y : x \mapsto f(x, y)$ is $\mathcal{B}_X$-measurable for every $y \in Y$.

(v) If $E \in \mathcal{B}_X \times \mathcal{B}_Y$, show that the slices $E_x := \{y \in Y : (x, y) \in E\}$ lie in a countably generated σ-algebra. In other words, show that there exists an at most countable collection $\mathcal{A} = \mathcal{A}_E$ of sets (which can depend on E) such that $\{E_x : x \in X\} \subset \langle \mathcal{A} \rangle$. Conclude in particular that the number of distinct slices E_x is at most c, the cardinality of the continuum. (The last part of this exercise is only suitable for students who are comfortable with cardinal arithmetic.)

Exercise 1.7.19.

(i) Show that the product of two trivial σ-algebras (on two different spaces X, Y) is again trivial.

(ii) Show that the product of two atomic σ-algebras is again atomic.

(iii) Show that the product of two finite σ-algebras is again finite.

(iv) Show that the product of two Borel σ-algebras (on two Euclidean spaces $\mathbf{R}^d, \mathbf{R}^{d'}$ with $d, d' \geq 1$) is again the Borel σ-algebra (on $\mathbf{R}^d \times \mathbf{R}^{d'} \equiv \mathbf{R}^{d+d'}$).

(v) Show that the product of two Lebesgue σ-algebras (on two Euclidean spaces $\mathbf{R}^d, \mathbf{R}^{d'}$ with $d, d' \geq 1$) is *not* the Lebesgue σ-algebra. (*Hint:* argue by contradiction and use Exercise 1.7.18(iii).)

(vi) However, show that the Lebesgue σ-algebra on $\mathbf{R}^{d+d'}$ is the *completion* (see Exercise 1.4.26) of the product of the Lebesgue σ-algebras of $\mathbf{R}^d$ and $\mathbf{R}^{d'}$ with respect to $d+d'$-dimensional Lebesgue measure.

(vii) This part of the exercise is only for students who are comfortable with cardinal arithmetic. Give an example to show that the product of two discrete σ-algebras is not necessarily discrete.

(viii) On the other hand, show that the product of two discrete σ-algebras $2^X, 2^Y$ is again a discrete σ-algebra if at least one of the domains X, Y is at most countably infinite.

Now suppose we have two measure spaces $(X, \mathcal{B}_X, \mu_X)$ and $(Y, \mathcal{B}_Y, \mu_Y)$. Given that we can multiply together the sets X and Y to form a product set $X \times Y$, and can multiply the σ-algebras $\mathcal{B}_X$ and $\mathcal{B}_Y$ together to form a product σ-algebra $\mathcal{B}_X \times \mathcal{B}_Y$, it is natural to expect that we can multiply the two measures $\mu_X \colon \mathcal{B}_X \to [0, +\infty]$ and $\mu_Y \colon \mathcal{B}_Y \to [0, +\infty]$ to form a product measure $\mu_X \times \mu_Y \colon \mathcal{B}_X \times \mathcal{B}_Y \to [0, +\infty]$. In view of the "base times height formula" that one learns in elementary school, one expects to have

$$\mu_X \times \mu_Y(E \times F) = \mu_X(E)\mu_Y(F) \tag{1.36}$$

whenever $E \in \mathcal{B}_X$ and $F \in \mathcal{B}_Y$.

To construct this measure, it is convenient to make the assumption that both spaces are *σ-finite*:

Definition 1.7.10 (σ-finite)**.** A measure space $(X, \mathcal{B}, \mu)$ is *σ-finite* if X can be expressed as the countable union of sets of finite measure.

Thus, for instance, $\mathbf{R}^d$ with Lebesgue measure is σ-finite, as $\mathbf{R}^d$ can be expressed as the union of (for instance) the balls $B(0, n)$ for $n = 1, 2, 3, \ldots$, each of which has finite measure. On the other hand, $\mathbf{R}^d$ with counting measure is not σ-finite (why?). But most measure spaces that one actually encounters in analysis (including, clearly, all probability spaces) are σ-finite. It is possible to partially extend the theory of product spaces to the non-σ-finite setting, but there are a number of very delicate technical issues that arise and so we will not discuss them here.

As long as we restrict attention to the σ-finite case, product measure always exists and is unique:

Proposition 1.7.11 (Existence and uniqueness of product measure)**.** *Let $(X, \mathcal{B}_X, \mu_X)$ and $(Y, \mathcal{B}_Y, \mu_Y)$ be σ-finite measure spaces. Then there exists a unique measure $\mu_X \times \mu_Y$ on $\mathcal{B}_X \times \mathcal{B}_Y$ that obeys $\mu_X \times \mu_Y(E \times F) = \mu_X(E)\mu_Y(F)$ whenever $E \in \mathcal{B}_X$ and $F \in \mathcal{B}_Y$.*

Proof. We first show existence. Inspired by the fact that Lebesgue measure is the Hahn-Kolmogorov completion of elementary (pre-)measure, we shall first construct an "elementary product pre-measure" that we will then apply Theorem 1.7.8 to.

Let $\mathcal{B}_0$ be the collection of all finite unions

$$S := (E_1 \times F_1) \cup \ldots \cup (E_k \times F_k)$$

of Cartesian products of $\mathcal{B}_X$-measurable sets $E_1, \ldots, E_k$ and $\mathcal{B}_Y$-measurable sets $F_1, \ldots, F_k$. (One can think of such sets as being somewhat analogous to elementary sets in Euclidean space, although the analogy is not perfectly exact.) It is not difficult to verify that this is a Boolean algebra (though it is not, in general, a σ-algebra). Also, any set in $\mathcal{B}_0$ can be easily decomposed into a *disjoint* union of product sets $E_1 \times F_1, \ldots, E_k \times F_k$ of $\mathcal{B}_X$-measurable sets and $\mathcal{B}_Y$-measurable sets (cf. Exercise 1.1.2). We then define the quantity $\mu_0(S)$ associated such a disjoint union S by the formula

$$\mu_0(S) := \sum_{j=1}^{k} \mu_X(E_j)\mu_Y(F_j)$$

whenever S is the disjoint union of products $E_1 \times F_1, \ldots, E_k \times F_k$ of $\mathcal{B}_X$-measurable sets and $\mathcal{B}_Y$-measurable sets. One can show that this definition does not depend on exactly how S is decomposed, and gives a finitely additive measure $\mu_0 \colon \mathcal{B}_0 \to [0, +\infty]$ (cf. Exercise 1.1.2 and Exercise 1.4.33).

Now we show that μ_0 is a pre-measure. It suffices to show that if $S \in \mathcal{B}_0$ is the countable disjoint union of sets $S_1, S_2, \ldots \in \mathcal{B}_0$, then $\mu_0(S) = \sum_{n=1}^{\infty} \mu(S_n)$.

Splitting S up into disjoint product sets, and restricting the S_n to each of these product sets in turn, we may assume without loss of generality (using the finite additivity of μ_0) that $S = E \times F$ for some $E \in \mathcal{B}_X$ and $F \in \mathcal{B}_Y$. In a similar spirit, by breaking each S_n up into component product sets and using finite additivity again, we may assume without loss of generality that each S_n takes the form $S_n = E_n \times F_n$ for some $E_n \in \mathcal{B}_X$ and $F_n \in \mathcal{B}_Y$. By definition of μ_0, our objective is now to show that

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

To do this, first observe from construction that we have the pointwise identity

$$1_E(x)1_F(y) = \sum_{n=1}^{\infty} 1_{E_n}(x)1_{F_n}(y)$$

for all $x \in X$ and $y \in Y$. We fix $x \in X$, and integrate this identity in y (noting that both sides are measurable and unsigned) to conclude that

$$\int_Y 1_E(x)1_F(y)\ d\mu_Y(y) = \int_Y \sum_{n=1}^{\infty} 1_{E_n}(x)1_{F_n}(y)\ d\mu_Y(y).$$

The left-hand side simplifies to $1_E(x)\mu_Y(F)$. To compute the right-hand side, we use the monotone convergence theorem (Theorem 1.4.44) to interchange the summation and integration, and soon see that the right-hand side is $\sum_{n=1}^{\infty} 1_{E_n}(x)\mu_Y(F_n)$, thus

$$1_E(x)\mu_Y(F) = \sum_{n=1}^{\infty} 1_{E_n}(x)\mu_Y(F_n)$$

for all x. Both sides are measurable and unsigned in x, so we may integrate in X and conclude that

$$\int_X 1_E(x)\mu_Y(F)\ d\mu_X = \int_X \sum_{n=1}^{\infty} 1_{E_n}(x)\mu_Y(F_n)\ d\mu_X(x).$$

The left-hand side here is $\mu_X(E)\mu_Y(F)$. Using monotone convergence as before, the right-hand side simplifies to $\sum_{n=1}^{\infty} \mu_X(E_n)\mu_Y(F_n)$, and the claim follows.

Now that we have established that μ_0 is a pre-measure, we may apply Theorem 1.7.8 to extend this measure to a countably additive measure $\mu_X \times \mu_Y$ on a σ-algebra containing $\mathcal{B}_0$. By Exercise 1.7.18(2),

$\mu_X \times \mu_Y$ is a countably additive measure on $\mathcal{B}_X \times \mathcal{B}_Y$, and as it extends μ_0, it will obey (1.36). Finally, to show uniqueness, observe from finite additivity that any measure $\mu_X \times \mu_Y$ on $\mathcal{B}_X \times \mathcal{B}_Y$ that obeys (1.36) must extend μ_0, and so uniqueness follows from Exercise 1.7.7. $\square$

Remark 1.7.12. When X, Y are not both σ-finite, then one can still construct at least one product measure, but it will, in general, not be unique. This makes the theory much more subtle, and we will not discuss it in these notes.

Example 1.7.13. From Exercise 1.2.22, we see that the product $m^d \times m^{d'}$ of the Lebesgue measures $m^d, m^{d'}$ on $(\mathbf{R}^d, \mathcal{L}[\mathbf{R}^d])$ and $(\mathbf{R}^d, \mathcal{L}[\mathbf{R}^{d'}])$ respectively will agree with Lebesgue measure $m^{d+d'}$ on the product space $\mathcal{L}[\mathbf{R}^d] \times \mathcal{L}[\mathbf{R}^{d'}]$, which as noted in Exercise 1.7.19 is a subalgebra of $\mathcal{L}[\mathbf{R}^{d+d'}]$. After taking the completion $\overline{m^d \times m^{d'}}$ of this product measure, one obtains the full Lebesgue measure $m^{d+d'}$.

Exercise 1.7.20. Let $(X, \mathcal{B}_X)$, $(Y, \mathcal{B}_Y)$ be measurable spaces.

(i) Show that the product of two Dirac measures on $(X, \mathcal{B}_X)$, $(Y, \mathcal{B}_Y)$ is a Dirac measure on $(X \times Y, \mathcal{B}_X \times \mathcal{B}_Y)$.

(ii) If X, Y are at most countable, show that the product of the two counting measures on $(X, \mathcal{B}_X)$, $(Y, \mathcal{B}_Y)$ is the counting measure on $(X \times Y, \mathcal{B}_X \times \mathcal{B}_Y)$.

Exercise 1.7.21 (Associativity of product)**.** Let $(X, \mathcal{B}_X, \mu_X)$, $(Y, \mathcal{B}_Y, \mu_Y)$, $(Z, \mathcal{B}_Z, \mu_Z)$ be σ-finite sets. We may identify the Cartesian products $(X \times Y) \times Z$ and $X \times (Y \times Z)$ with each other in the obvious manner. If we do so, show that $(\mathcal{B}_X \times \mathcal{B}_Y) \times \mathcal{B}_Z = \mathcal{B}_X \times (\mathcal{B}_Y \times \mathcal{B}_Z)$ and $(\mu_X \times \mu_Y) \times \mu_Z = \mu_X \times (\mu_Y \times \mu_Z)$.

Now we integrate using this product measure. We will need the following technical lemma. Define a *monotone class* in X is a collection $\mathcal{B}$ of subsets of X with the following two closure properties:

(i) If $E_1 \subset E_2 \subset \ldots$ are a countable increasing sequence of sets in $\mathcal{B}$, then $\bigcup_{n=1}^\infty E_n \in \mathcal{B}$.

(ii) If $E_1 \supset E_2 \supset \ldots$ are a countable decreasing sequence of sets in $\mathcal{B}$, then $\bigcap_{n=1}^\infty E_n \in \mathcal{B}$.

Lemma 1.7.14 (Monotone class lemma)**.** *Let $\mathcal{A}$ be a Boolean algebra on X. Then $\langle\mathcal{A}\rangle$ is the smallest monotone class that contains $\mathcal{A}$.*

Proof. Let $\mathcal{B}$ be the intersection of all the monotone classes that contain $\mathcal{A}$. Since $\langle\mathcal{A}\rangle$ is clearly one such class, $\mathcal{B}$ is a subset of $\langle\mathcal{A}\rangle$. Our task is then to show that $\mathcal{B}$ contains $\langle\mathcal{A}\rangle$.

It is also clear that $\mathcal{B}$ is a monotone class that contains $\mathcal{A}$. By replacing all the elements of $\mathcal{B}$ with their complements, we see that $\mathcal{B}$ is necessarily closed under complements.

For any $E \in \mathcal{A}$, consider the set $\mathcal{C}_E$ of all sets $F \in \mathcal{B}$ such that $F\backslash E$, $E\backslash F$, $F \cap E$, and $X\backslash(E \cup F)$ all lie in $\mathcal{B}$. It is clear that $\mathcal{C}_E$ contains $\mathcal{A}$; since $\mathcal{B}$ is a monotone class, we see that $\mathcal{C}_E$ is also. By definition of $\mathcal{B}$, we conclude that $\mathcal{C}_E = \mathcal{B}$ for all $E \in \mathcal{A}$.

Next, let $\mathcal{D}$ be the set of all $E \in \mathcal{B}$ such that $F\backslash E$, $E\backslash F$, $F \cap E$, and $X\backslash(E\cup F)$ all lie in $\mathcal{B}$ for all $F \in \mathcal{B}$. By the previous discussion, we see that $\mathcal{D}$ contains $\mathcal{A}$. One also easily verifies that $\mathcal{D}$ is a monotone class. By definition of $\mathcal{B}$, we conclude that $\mathcal{D} = \mathcal{B}$. Since $\mathcal{B}$ is also closed under complements, this implies that $\mathcal{B}$ is closed with respect to finite unions. Since this class also contains $\mathcal{A}$, which contains $\emptyset$, we conclude that $\mathcal{B}$ is a Boolean algebra. Since $\mathcal{B}$ is also closed under increasing countable unions, we conclude that it is closed under arbitrary countable unions, and is thus a σ-algebra. As it contains $\mathcal{A}$, it must also contain $\langle\mathcal{A}\rangle$. □

Theorem 1.7.15 (Tonelli's theorem, incomplete version)**.** *Let $(X, \mathcal{B}_X, \mu_X)$ and $(Y, \mathcal{B}_Y, \mu_Y)$ be σ-finite measure spaces, and let $f\colon X \times Y \to [0, +\infty]$ be measurable with respect to $\mathcal{B}_X \times \mathcal{B}_Y$. Then:*

(i) *The functions $x \mapsto \int_Y f(x, y)\ d\mu_Y(y)$ and $y \mapsto \int_X f(x, y)\ d\mu_X(x)$ (which are well-defined, thanks to Exercise 1.7.18) are measurable with respect to $\mathcal{B}_X$ and $\mathcal{B}_Y$ respectively.*

(ii) *We have*

$$\int_{X\times Y} f(x, y)\ d\mu_X \times \mu_Y(x, y)$$

$$= \int_X (\int_Y f(x, y)\ d\mu_Y(y))\ d\mu_X(x)$$

$$= \int_Y \Big(\int_X f(x,y)\ d\mu_X(x) \Big)\ d\mu_Y(y).$$

Proof. By writing the σ-finite space X as an increasing union $X = \bigcup_{n=1}^\infty X_n$ of finite measure sets, we see from several applications of the monotone convergence theorem (Theorem 1.4.44) that it suffices to prove the claims with X replaced by X_n. Thus we may assume without loss of generality that X has finite measure. Similarly we may assume Y has finite measure. Note from (1.36) that this implies that $X \times Y$ has finite measure also.

Every unsigned measurable function is the increasing limit of unsigned simple functions. By several applications of the monotone convergence theorem (Theorem 1.4.44), we thus see that it suffices to verify the claim when f is a simple function. By linearity, it then suffices to verify the claim when f is an indicator function, thus $f = 1_S$ for some $S \in \mathcal{B}_X \times \mathcal{B}_Y$.

Let $\mathcal{C}$ be the set of all $S \in \mathcal{B}_X \times \mathcal{B}_Y$ for which the claims hold. From the repeated applications of the monotone convergence theorem (Theorem 1.4.44) and the downward monotone convergence theorem (which is available in this finite measure setting) we see that $\mathcal{C}$ is a monotone class.

By direct computation (using (1.36)), we see that $\mathcal{C}$ contains as an element any product $S = E \times F$ with $E \in \mathcal{B}_X$ and $F \in \mathcal{B}_Y$. By finite additivity, we conclude that $\mathcal{C}$ also contains as an element any a disjoint finite union $S = E_1 \times F_1 \cup \ldots \cup E_k \times F_k$ of such products. This implies that $\mathcal{C}$ also contains the Boolean algebra $\mathcal{B}_0$ in the proof of Proposition 1.7.11, as such sets can always be expressed as the disjoint finite union of Cartesian products of measurable sets. Applying the monotone class lemma, we conclude that $\mathcal{C}$ contains $\langle \mathcal{B}_0 \rangle = \mathcal{B}_X \times \mathcal{B}_Y$, and the claim follows. $\square$

Remark 1.7.16. Note that Tonelli's theorem for sums (Theorem 0.0.2) is a special case of the above result when μ_X, μ_Y are counting measure. In a similar spirit, Corollary 1.4.46 is the special case when just one of μ_X, μ_Y is counting measure.

Corollary 1.7.17. *Let $(X, \mathcal{B}_X, \mu_X)$ and $(Y, \mathcal{B}_Y, \mu_Y)$ be σ-finite measure spaces, and let $E \in \mathcal{B}_X \times \mathcal{B}_Y$ be a null set with respect to $\mu_X \times \mu_Y$.*

Then for μ_X-almost every $x \in X$, the set $E_x := \{y \in Y : (x, y) \in E\}$ is a μ_Y-null set; and similarly, for μ_Y-almost every $y \in Y$, the set $E^y := \{x \in X : (x, y) \in E\}$ is a μ_X-null set.

Proof. Applying the Tonelli theorem to the indicator function 1_E, we conclude that

$$0 = \int_X (\int_Y 1_E(x, y)\ d\mu_Y(y))\ d\mu_X(x) = \int_Y (\int_X 1_E(x, y)\ d\mu_X(x))\ d\mu_Y(y)$$

and thus

$$0 = \int_X \mu_Y(E_x)\ d\mu_X(x) = \int_Y \mu_X(E^y)\ d\mu_Y(y),$$

and the claim follows. □

With this corollary, we can extend Tonelli's theorem to the *completion* $(X \times Y, \overline{\mathcal{B}_X \times \mathcal{B}_Y}, \overline{\mu_X \times \mu_Y})$ of the product space $(X \times Y, \mathcal{B}_X \times \mathcal{B}_Y, \mu_X \times \mu_Y)$, as constructed in Exercise 1.4.26. But we can easily extend the Tonelli theorem to this context:

Theorem 1.7.18 (Tonelli's theorem, complete version)**.** *Let $(X, \mathcal{B}_X, \mu_X)$ and $(Y, \mathcal{B}_Y, \mu_Y)$ be complete σ-finite measure spaces, and let $f \colon X \times Y \to [0, +\infty]$ be measurable with respect to $\overline{\mathcal{B}_X \times \mathcal{B}_Y}$. Then:*

(i) *For μ_X-almost every $x \in X$, the function $y \mapsto f(x, y)$ is $\mathcal{B}_Y$-measurable, and in particular $\int_Y f(x, y)\ d\mu_Y(y)$ exists. Furthermore, the (μ_X-almost everywhere defined) map $x \mapsto \int_Y f(x, y)\ d\mu_Y$ is $\mathcal{B}_X$-measurable.*

(ii) *For μ_Y-almost every $y \in Y$, the function $x \mapsto f(x, y)$ is $\mathcal{B}_X$-measurable, and in particular $\int_X f(x, y)\ d\mu_X(x)$ exists. Furthermore, the (μ_Y-almost everywhere defined) map $y \mapsto \int_X f(x, y)\ d\mu_X$ is $\mathcal{B}_Y$-measurable.*

(iii) *We have*

(1.37)
$$\int_{X \times Y} f(x, y)\ d\overline{\mu_X \times \mu_Y}(x, y) = \int_X (\int_Y f(x, y)\ d\mu_Y(y))\ d\mu_X(x)$$
$$= \int_X (\int_Y f(x, y)\ d\mu_Y(y))\ d\mu_X(x).$$

Proof. From Exercise 1.4.28, every measurable set in $\overline{\mathcal{B}_X \times \mathcal{B}_Y}$ is equal to a measurable set in $\mathcal{B}_X \times \mathcal{B}_Y$ outside of a $\mu_X \times \mu_Y$-null set. This implies that the $\overline{\mathcal{B}_X \times \mathcal{B}_Y}$-measurable function f agrees with a $\mathcal{B}_X \times \mathcal{B}_Y$-measurable function $\tilde f$ outside of a $\mu_X \times \mu_Y$-null set E (as can be seen by expressing f as the limit of simple functions). From Corollary 1.7.17, we see that for μ_X-almost every $x \in X$, the function $y \mapsto f(x,y)$ agrees with $y \mapsto \tilde f(x,y)$ outside of a μ_Y-null set (and is in particular measurable, as $(Y, \mathcal{B}_Y, \mu_Y)$ is complete); and similarly for μ_Y-almost every $y \in Y$, the function $x \mapsto f(x,y)$ agrees with $x \mapsto \tilde f(x,y)$ outside of a μ_X-null set and is measurable, and the claim follows. $\square$

Specialising to the case when f is an indicator function $f = 1_E$, we conclude

Corollary 1.7.19 (Tonelli's theorem for sets)**.** *Let $(X, \mathcal{B}_X, \mu_X)$ and $(Y, \mathcal{B}_Y, \mu_Y)$ be complete σ-finite measure spaces, and let $E \in \overline{\mathcal{B}_X \times \mathcal{B}_Y}$. Then:*

(i) *For μ_X-almost every $x \in X$, the set $E_x := \{y \in Y : (x,y) \in E\}$ lies in $\mathcal{B}_Y$, and the (μ_X-almost everywhere defined) map $x \mapsto \mu_Y(E_x)$ is $\mathcal{B}_X$-measurable.*

(ii) *For μ_Y-almost every $y \in Y$, the set $E^y := \{x \in X : (x,y) \in E\}$ lies in $\mathcal{B}_X$, and the (μ_Y-almost everywhere defined) map $y \mapsto \mu_X(E^y)$ is $\mathcal{B}_Y$-measurable.*

(iii) *We have*

$$\overline{\mu_X \times \mu_Y}(E) = \int_X \mu_Y(E_x)\ d\mu_X(x) \tag{1.38}$$

$$= \int_X \mu_X(E^y)\ d\mu_X(x).$$

Exercise 1.7.22. The purpose of this exercise is to demonstrate that Tonelli's theorem can fail if the σ-finite hypothesis is removed, and also that product measure need not be unique. Let X is the unit interval $[0,1]$ with Lebesgue measure m (and the Lebesgue σ-algebra $\mathcal{L}([0,1])$) and Y is the unit interval $[0,1]$ with counting measure (and the discrete σ-algebra $2^{[0,1]}$) $\#$. Let $f := 1_E$ be the indicator function of the diagonal $E := \{(x,x) : x \in [0,1]\}$.

(i) Show that f is measurable in the product σ-algebra.

(ii) Show that $\int_X(\int_Y f(x,y)\ d\#(y))dm(x) = 1$.

(iii) Show that $\int_Y(\int_X f(x,y)\ dm(x))d\#(y) = 0$.

(iv) Show that there is more than one measure μ on $\mathcal{L}([0,1]) \times 2^{[0,1]}$ with the property that $\mu(E \times F) = m(E)\#(F)$ for all $E \in \mathcal{L}([0,1])$ and $F \in 2^{[0,1]}$. (*Hint:* use the two different ways to perform a double integral to create two different measures.)

Remark 1.7.20. If f is not assumed to be measurable in the product space (or its completion), then of course the expression $\int_{X\times Y} f(x,y)\ d\overline{\mu_X \times \mu_Y}(x,y)$ does not make sense. Furthermore, in this case the remaining two expressions in (1.37) may become different as well (in some models of set theory, at least), even when X and Y are finite measure. For instance, let us assume the *continuum hypothesis*, which implies that the unit interval $[0,1]$ can be placed in one-to-one correspondence with the *first uncountable ordinal* ω_1. Let $\prec$ be the ordering of $[0,1]$ that is associated to this ordinal, let $E := \{(x,y) \in [0,1]^2 : x \prec y\}$, and let $f := 1_E$. Then, for any $y \in [0,1]$, there are at most countably many x such that $x \prec y$, and so $\int_{[0,1]} f(x,y)\ dx$ exists and is equal to zero for every y. On the other hand, for every $x \in [0,1]$, one has $x \prec y$ for all but countably many $y \in [0,1]$, and so $\int_{[0,1]} f(x,y)\ dy$ exists and is equal to one for every y, and so the last two expressions in (1.37) exist but are unequal. (In particular, Tonelli's theorem implies that E cannot be a Lebesgue measurable subset of $[0,1]^2$.) Thus we see that measurability in the product space is an important hypothesis. (There do however exist models of set theory (with the axiom of choice) in which such counterexamples cannot be constructed, at least in the case when X and Y are the unit interval with Lebesgue measure.)

Tonelli's theorem is for the unsigned integral, but it leads to an important analogue for the absolutely integral, known as *Fubini's theorem*:

Theorem 1.7.21 (Fubini's theorem)**.** *Let $(X, \mathcal{B}_X, \mu_X)$ and $(Y, \mathcal{B}_Y, \mu_Y)$ be complete σ-finite measure spaces, and let $f\colon X \times Y \to \mathbf{C}$ be absolutely integrable with respect to $\overline{\mathcal{B}_X \times \mathcal{B}_Y}$. Then:*

(i) *For μ_X-almost every $x \in X$, the function $y \mapsto f(x,y)$ is absolutely integrable with respect to μ_Y, and in particular $\int_Y f(x,y)\ d\mu_Y(y)$ exists. Furthermore, the (μ_X-almost everywhere defined) map $x \mapsto \int_Y f(x,y)\ d\mu_Y(y)$ is absolutely integrable with respect to μ_X.*

(ii) *For μ_Y-almost every $y \in Y$, the function $x \mapsto f(x,y)$ is absolutely integrable with respect to μ_X, and in particular $\int_X f(x,y)\ d\mu_X(x)$ exists. Furthermore, the (μ_Y-almost everywhere defined) map $y \mapsto \int_X f(x,y)\ d\mu_X(x)$ is absolutely integrable with respect to μ_Y.*

(iii) *We have*

$$\int_{X\times Y} f(x,y)\ d\overline{\mu_X \times \mu_Y}(x,y) = \int_X (\int_Y f(x,y)\ d\mu_Y(y))\ d\mu_X(x)$$
$$= \int_X (\int_Y f(x,y)\ d\mu_Y(y))\ d\mu_X(x).$$

Proof. By taking real and imaginary parts we may assume that f is real; by taking positive and negative parts we may assume that f is unsigned. But then the claim follows from Tonelli's theorem; note from (1.37) that $\int_X(\int_Y f(x,y)\ d\mu_Y(y))\ d\mu_X(x)$ is finite, and so $\int_Y f(x,y)\ d\mu_Y(y) < \infty$ for μ_X-almost every $x \in X$, and similarly $\int_X f(x,y)\ d\mu_X(x) < \infty$ for μ_Y-almost every $y \in Y$. $\square$

Exercise 1.7.23. Give an example of a Borel measurable function $f\colon [0,1]^2 \to \mathbf{R}$ such that the integrals $\int_{[0,1]} f(x,y)\ dy$ and $\int_{[0,1]} f(x,y)\ dx$ exist and are absolutely integrable for all $x \in [0,1]$ and $y \in [0,1]$ respectively, and that $\int_{[0,1]}(\int_{[0,1]} f(x,y)\ dy)\ dx$ and $\int_{[0,1]}(\int_{[0,1]} f(x,y)\ dy)\ dx$ exist and are absolutely integrable, but such that

$$\int_{[0,1]}(\int_{[0,1]} f(x,y)\ dy)\ dx \neq \int_{[0,1]}(\int_{[0,1]} f(x,y)\ dy)\ dx.$$

are unequal. (*Hint:* adapt the example from Remark 0.0.3.) Thus we see that Fubini's theorem fails when one drops the hypothesis that f is absolutely integrable with respect to the product space.

Remark 1.7.22. Despite the failure of Tonelli's theorem in the σ-finite setting, it is possible to (carefully) extend Fubini's theorem to the non-σ-finite setting, as the absolute integrability hypotheses, when combined with Markov's inequality (Exercise 1.4.36(vi)), can provide a substitute for the σ-finite property. However, we will not do so here, and indeed I would recommend proceeding with extreme caution when performing any sort of interchange of integrals or invoking of product measure when one is not in the σ-finite setting.

Informally, Fubini's theorem allows one to always interchange the order of two integrals, as long as the integrand is absolutely integrable in the product space (or its completion). In particular, specialising to Lebesgue measure, we have

$$\int_{\mathbf{R}^{d+d'}} f(x,y)\ d(x,y) = \int_{\mathbf{R}^d} (\int_{\mathbf{R}^{d'}} f(x,y)\ dy)\ dx = \int_{\mathbf{R}^{d'}} (\int_{\mathbf{R}^d} f(x,y)\ dx)\ dy$$

whenever $f\colon \mathbf{R}^{d+d'} \to \mathbf{C}$ is absolutely integrable. In view of this, we often write $dxdy$ (or $dydx$) for $d(x,y)$.

By combining Fubini's theorem with Tonelli's theorem, we can recast the absolute integrability hypothesis:

Corollary 1.7.23 (Fubini-Tonelli theorem)**.** *Let $(X, \mathcal{B}_X, \mu_X)$ and $(Y, \mathcal{B}_Y, \mu_Y)$ be complete σ-finite measure spaces, and let $f\colon X \times Y \to \mathbf{C}$ be measurable with respect to $\overline{\mathcal{B}_X \times \mathcal{B}_Y}$. If*

$$\int_X (\int_Y |f(x,y)|\ d\mu_Y(y))\ d\mu_X(x) < \infty$$

(note the left-hand side always exists, by Tonelli's theorem) then f is absolutely integrable with respect to $\overline{\mathcal{B}_X \times \mathcal{B}_Y}$, and in particular the conclusions of Fubini's theorem hold. Similarly if we use $\int_Y (\int_X |f(x,y)|\ d\mu_X(x))\ d\mu_Y(y)$ instead of $\int_X (\int_Y |f(x,y)|\ d\mu_Y)\ d\mu_X$.

The Fubini-Tonelli theorem is an indispensable tool for computing integrals. We give some basic examples below:

Exercise 1.7.24 (Area interpretation of integral)**.** Let $(X, \mathcal{B}, \mu)$ be a σ-finite measure space, and let $\mathbf{R}$ be equipped with Lebesgue measure m and the Borel σ-algebra $\mathcal{B}[\mathbf{R}]$. Show that if $f\colon X \to [0, +\infty]$ is measurable if and only if the set $\{(x,t) \in X \times \mathbf{R} : 0 \leq t \leq f(x)\}$ is measurable in $\mathcal{B} \times \mathcal{B}[\mathbf{R}]$, in which case we have

$$(\mu \times m)(\{(x,t) \in X \times \mathbf{R} : 0 \leq t \leq f(x)\}) = \int_X f(x)\ d\mu(x).$$

Similarly if we replace $\{(x,t) \in X \times \mathbf{R} : 0 \leq t \leq f(x)\}$ by $\{(x,t) \in X \times \mathbf{R} : 0 \leq t < f(x)\}$.

Exercise 1.7.25 (Distribution formula)**.** Let $(X, \mathcal{B}, \mu)$ be a σ-finite measure space, and let $f\colon X \to [0,+\infty]$ be measurable. Show that

$$\int_X f(x)\ d\mu(x) = \int_{[0,+\infty]} \mu(\{x \in X : f(x) \geq \lambda\})\ d\lambda.$$

(Note that the integrand on the right-hand side is monotone and thus Lebesgue measurable.) Similarly if we replace $\{x \in X : f(x) \geq \lambda\}$ by $\{x \in X : f(x) > \lambda\}$.

Exercise 1.7.26 (Approximations to the identity)**.** Let $P\colon \mathbf{R}^d \to \mathbf{R}^+$ be a good kernel (see Exercise 1.6.27), and let $P_t(x) := \frac{1}{t^d} P(\frac{x}{t})$ be the associated rescaled functions. Show that if $f\colon \mathbf{R}^d \to \mathbf{C}$ is absolutely integrable, that $f * P_t$ converges in L^1 norm to f as $t \to 0$. (*Hint:* use the density argument. You will need an upper bound on $\|f * P_t\|_{L^1(\mathbf{R}^d)}$ which can be obtained using Tonelli's theorem.)

Chapter 2

Related articles

2.1. Problem solving strategies

The purpose of this section is to list (in no particular order) a number of common problem solving strategies for attacking real analysis exercises such as that presented in this text. Some of these strategies are specific to real analysis type problems, but others are quite general and would be applicable to other mathematical exercises.

2.1.1. Split up equalities into inequalities. If one has to show that two numerical quantities X and Y are equal, try proving that $X \leq Y$ and $Y \leq X$ separately. Often one of these will be very easy, and the other one harder; but the easy direction may still provide some clue as to what needs to be done to establish the other direction. Exercise 1.1.6(iii) is a typical problem in which this strategy can be applied.

In a similar spirit, to show that two sets E and F are equal, try proving that $E \subset F$ and $F \subset E$. See for instance the proof of Lemma 1.2.11 for a simple example of this.

2.1.2. Give yourself an epsilon of room. If one has to show that $X \leq Y$, try proving that $X \leq Y + \varepsilon$ for any $\varepsilon > 0$. (This trick combines well with §2.1.1.) See for instance Lemma 1.2.5 for an example of this.

In a similar spirit:

- if one needs to show that a quantity X vanishes, try showing that $|X| \leq \varepsilon$ for every $\varepsilon > 0$. (Exercise 1.2.19 is a simple application of this strategy.)
- if one wishes to show that two functions f, g agree almost everywhere, try showing first that $|f(x) - g(x)| \leq \varepsilon$ holds for almost every x, or even just outside of a set of measure at most ε, for any given $\varepsilon > 0$. (See for instance the proof of Lemma 1.5.7 for an example of this.)
- if one wants to show that a sequence x_n of real numbers converges to zero, try showing that $\limsup_{n\to\infty} |x_n| \leq \varepsilon$ for every $\varepsilon > 0$. (The proof of the Lebesgue differentiation theorem, Theorem 1.6.12, is in this spirit.)

Don't be too focused on getting all your error terms adding up to exactly ε - usually, as long as the final error bound consists of terms that can all be made as small as one wishes by choosing parameters in a suitable way, that is enough. For instance, an error term such as 10ε is certainly OK, or even more complicated expressions such as $10\varepsilon/\delta + 4\delta$ if one has the ability to choose δ as small as one wishes, and then after δ is chosen, one can then also set ε as small as one wishes (in a manner that can depend on δ).

One caveat: for finite x, and any $\varepsilon > 0$, it is true that $x + \varepsilon > x$ and $x - \varepsilon < x$, but this statement is not true when x is equal to $+\infty$ (or $-\infty$). So remember to exercise some care with the epsilon of room trick when some quantities are infinite.

See also §2.7 of *An epsilon of room, Vol. I.*

2.1.3. Decompose (or approximate) a rough or general object into (or by) a smoother or simpler one. If one has to prove something about an unbounded (or infinite measure) set, consider proving it for bounded (or finite measure) sets first if this looks easier.

In a similar spirit:

- If one has to prove something about a measurable set, try proving it for open, closed, compact, bounded, or elementary sets first.
- If one has to prove something about a measurable function, try proving it for functions that are continuous, bounded, compactly supported, simple, absolutely integrable, etc..
- If one has to prove something about an infinite sum or sequence, try proving it first for finite truncations of that sum or sequence (but try to get all the bounds independent of the number of terms in that truncation, so that you can still pass to the limit!).
- If one has to prove something about a complex-valued function, try it for real-valued functions first.
- If one has to prove something about a real-valued function, try it for unsigned functions first.

- If one has to prove something about a simple function, try it for indicator functions first.

In order to pass back to the general case from these special cases, one will have to somehow decompose the general object into a combination of special ones, or approximate general objects by special ones (or as a limit of a sequence of special objects). In the latter case, one may need an epsilon of room (§2.1.2), and some sort of limiting analysis may be needed to deal with the errors in the approximation (it is not always enough to just "pass to the limit", as one has to justify that the desirable properties of the approximating object are preserved in the limit). Littlewood's principles (Section 1.3.5) and their variants are often useful for thus purpose.

Note: one should not do this blindly, as one might then be loading on a bunch of distracting but ultimately useless hypotheses that end up being a lot less help than one might hope. But they should be kept in mind as something to try if one starts having thoughts such as "Gee, it would be nice at this point if I could assume that f is continuous / real-valued / simple / unsigned / etc.".

In the more quantitative areas of analysis and PDE, one sees a common variant of the above technique, namely the method of *a priori estimates*. Here, one needs to prove an estimate or inequality for all functions in a large, rough class (e.g. all rough solutions to a PDE). One can often then first prove this inequality in a much smaller (but still "dense") class of "nice" functions, so that there is little difficulty justifying the various manipulations (e.g. exchanging integrals, sums, or limits, or integrating by parts) that one wishes to perform. Once one obtains these a priori estimates, one can then often take some sort of limiting argument to recover the general case.

2.1.4. If one needs to flip an upper bound to a lower bound or vice versa, look for a way to take reflections or complements. Sometimes one needs a lower bound for some quantity, but only has techniques that give upper bounds. In some cases, though, one can "reflect" an upper bound into a lower bound (or vice versa) by replacing a set E contained in some space X with its complement $X \backslash E$, or a function f with its negation $-f$ (or perhaps subtracting f

from some dominating function F to obtain $F - f$). This trick works best when the objects being reflected are contained in some sort of "bounded", "finite measure", or "absolutely integrable" container, so that one avoids having the dangerous situation of having to subtract infinite quantities from each other.

A typical example of this is when one deduces downward monotone convergence for sets from upward monotone convergence for sets (Exercise 1.2.11).

2.1.5. Uncountable unions can sometimes be replaced by countable or finite unions. Uncountable unions are not well-behaved in measure theory; for instance, an uncountable union of null sets need not be a null set (or even a measurable set). (On the other hand, the uncountable union of open sets remains open; this can often be important to know.) However, in many cases one can replace an uncountable union by a countable one. For instance, if one needs to prove a statement for all $\varepsilon > 0$, then there are an uncountable number of ε's one needs to check, which may threaten measurability; but in many cases it is enough to only work with a countable sequence of εs, such as the numbers $1/m$ for $m = 1, 2, 3, \ldots$. (Exercise 1.6.30 relies heavily on this trick.)

In a similar spirit, given a real parameter λ, this parameter initially ranges over uncountably many values, but in some cases one can get away with only working with a countable set of such values, such as the rationals. In a similar spirit, rather than work with all boxes (of which there are uncountably many), one might work with the dyadic boxes (of which there are only countably many; also, they obey nicer nesting properties than general boxes and so are often desirable to work with in any event).

If you are working on a compact set, then one can often replace even uncountable unions with finite ones, so long as one is working with open sets. (The proof of Theorem 1.6.20 is a good example of this strategy.) When this option is available, it is often worth spending an epsilon of measure (or whatever other resource is available to spend) to make one's sets open, just so that one can take advantage of compactness.

2.1.6. If it is difficult to work globally, work locally instead. A domain such as Euclidean space $\mathbf{R}^d$ has infinite measure, and this creates a number of technical difficulties when trying to do measure theory directly on such spaces. Sometimes it is best to work more locally, for instance working on a large ball $B(0, R)$ or even a small ball such as $B(x, \varepsilon)$ first, and then figuring out how to patch things together later. Compactness (or the closely related property of total boundedness) is often useful for patching together small balls to cover a large ball. Patching together large balls into the whole space tends to work well when the properties one are trying to establish are local in nature (such as continuity, or pointwise convergence) or behave well with respect to countable unions. For instance, to prove that a sequence of functions f_n converges pointwise almost everywhere to f on $\mathbf{R}^d$, it suffices to verify this pointwise almost everywhere convergence on the ball $B(0, R)$ for every $R > 0$ (which one can take to be an integer to get countability, see §2.1.5). The application of vertical truncation (as done, for instance, in the proof of Corollary 1.3.14) is an instance of this idea.

2.1.7. Be willing to throw away an exceptional set. The "Lebesgue philosophy" to measure theory is that null sets are often "irrelevant", and so one should be very willing to cut out a set of measure zero on which bad things are happening (e.g. a function is undefined or infinite, a sequence of functions is not converging, etc.). One should also be only slightly less willing to throw away sets of positive but small measure, e.g. sets of measure at most ε. If such sets can be made arbitrarily small in measure, this is often almost as good as just throwing away a null set.

Many things in measure theory improve after throwing away a small set. The most notable examples of this are Egorov's theorem (Theorem 1.3.26) and Lusin's theorem (Theorem 1.3.28); see also Exercise 1.3.25 for some other examples of this idea.

It is also common to see a similar trick[1] of throwing away most of a sequence and working with a subsequence instead. See §2.1.17 below.

2.1.8. Draw pictures and try to build counterexamples. Measure theory, particularly on Euclidean spaces, has a significant geometric aspect to it, and you should be exploiting your geometric intuition. Drawing pictures and graphs of all the objects being studied is a good way to start. These pictures need not be completely realistic; they should be just complicated enough to hint at the complexities of the problem, but not more. For instance, usually one- or two-dimensional pictures suffice for understanding problems in $\mathbf{R}^d$; drawing intricate 3D (or 4D, etc.) pictures does not often make things simpler. To indicate that a function is not continuous, one or two discontinuities or oscillations might suffice; make it too ornate and it becomes less clear what to do about that function. One should view these pictures as providing a "cartoon sketch" of the situation, which exaggerates key features and downplays others, rather than a photorealistic image of the situation; too much detail or accuracy in a picture may be a waste of time, or otherwise counterproductive.

A common mistake is to try to draw a picture in which *both* the hypotheses *and* conclusion of the problem hold. This is actually not all that useful, as it often does not reveal the causal relationship between the former and the latter. One should try instead to draw a picture in which the hypotheses hold but for which the conclusion does not - in other words, a counterexample to the problem. Of course, you should be expected to fail at this task, given that the statement of the problem is presumably true. However, the way in which your picture fails to accomplish this task is often very instructive, and can reveal vital clues as to how the solution to the problem is supposed to proceed.

I have deliberately avoided drawing pictures in this book. This is not because I feel that pictures are not useful - far from it - but because I have found that it is far more informative for a reader

[1] This trick can also be interpreted as "throwing away a small set", but to understand what "small" means in this context, one needs the language of *ultrafilters*, which will not be discussed here; see [**Ta2008**, §1.5] for a discussion.

to draw his or her own pictures of a given mathematical situation, rather than rely on the author's images (except in situations where the geometric situation is particularly complicated or subtle), as such pictures will naturally be adapted to the reader's mindset rather than the author's. Besides, the process of actually *drawing* the picture is at least as instructive as the picture itself.

2.1.9. Try simpler cases first. This advice of course extends well beyond measure theory, but if one is completely stuck on a problem, try making the problem simpler (while still capturing at least one of the difficulties of the problem that you cannot currently resolve). For instance, if faced with a problem in $\mathbf{R}^d$, try the one-dimensional case $d = 1$ first. Faced with a problem about a general measurable function f, try a much simpler case first, such as an indicator function $f = 1_E$. Faced with a problem about a general measurable set, try an elementary set first. Faced with a problem about a sequence of functions, try a monotone sequence of functions first. And so forth. (Note that this trick overlaps quite a bit with §2.1.3.)

The problem should not be made so simple that it becomes trivial, as this doesn't really gain you any new insight about the original problem; instead, one should try to keep the "essential" difficulties of the problem while throwing away those aspects that you think are less important (but are still serving to add to the overall difficulty level).

On the other hand, if the simplified problem is unexpectedly easy, but one cannot extend the methods to the general case (or somehow leverage the simplified case to the general case, as in §2.1.3), this is an indication that the true difficulty lies elsewhere. For instance, if a problem involving general functions could be solved easily for monotone functions, but one cannot then extend that argument to the general case, this suggests that the true enemy is oscillation, and perhaps one should try another simple case in which the function is allowed to be highly oscillatory (but perhaps simple in other ways, e.g. bounded with compact support).

2.1.10. Abstract away any information that you believe or suspect to be irrelevant. Sometimes one is faced with an embarrassment of riches when it comes to what choice of technique to use on a problem; there are so many different facts that one knows about the problem, and so many different pieces of theory that one could apply, that one doesn't quite know where to begin.

When this happens, abstraction can be a vital tool to clear away some of the conceptual clutter. Here, one wants to "forget" part of the setting that the problem is phrased in, and only keep the part that seems to be most relevant to the hypotheses and conclusions of the problem (and thus, presumably, to the solution as well).

For instance, if one is working in a problem that is set in Euclidean space $\mathbf{R}^d$, but the hypotheses and conclusions only involve measure-theoretic concepts (e.g. measurability, integrability, measure, etc.) rather than topological structure, metric structure, etc., then it may be worthwhile to try abstracting the problem to the more general setting of an abstract measure space, thus forgetting that one was initially working in $\mathbf{R}^d$. The point of doing this is that it cuts down on the number of possible ways to start attacking the problem. For instance, facts such as outer regularity (every measurable set can be approximated from above by an open set) do not hold in abstract measure spaces (which do not even have a meaningful notion of an open set), and so presumably will not play a role in the solution; similarly for any facts involving boxes. Instead, one should be trying to use general facts about measure, such as countable additivity, which are not specific to $\mathbf{R}^d$.

Remark 2.1.1. It is worth noting that sometimes this abstraction method does not always work; for instance, when viewed as a measure space, $\mathbf{R}^d$ is not completely arbitrary, but does have one or two features that distinguish it from a generic measure space, most notably the fact that it is σ-finite. So, even if the hypotheses and conclusion of a problem about $\mathbf{R}^d$ is purely measure-theoretic in nature, one may still need to use some measure-theoretic facts specific to $\mathbf{R}^d$. Here, it becomes useful to know a little bit about the classification of measure spaces to have some intuition as to how "generic" a measure space such as $\mathbf{R}^d$ really is. This intuition is hard to convey at this level of the subject, but in general, measure spaces form a very "non-rigid" category, with very few invariants, and so it is largely true that one measure space usually behaves much the same as any other.

Another example of abstraction: suppose that a problem involves a large number of sets (e.g. E_n and F_n) and their measures, but that the conclusion of the problem only involves the measures $m(E_n), m(F_n)$ of the sets, rather than the sets themselves. Then one can try to abstract the sets out of the problem, by trying to write down every single relationship between the numerical quantities $m(E_n), m(F_n)$ that one can easily deduce from the given hypotheses (together with basic properties of measure, such as monotonicity or countable additivity). One can then rename these quantities (e.g. $a_n := m(E_n)$ and $b_n := m(F_n)$) to "forget" that these quantities arose from a measure-theoretic context, and then work with a purely numerical problem, in which one is starting with hypotheses on some sequences a_n, b_n of numbers and trying to deduce a conclusion about such sequences. Such a problem is often easier to solve than the original problem due to the simpler context. Sometimes, this simplified problem will end up being false, but the counterexample will often be instructive, either in indicating the need to add an additional hypothesis connecting the a_n, b_n, or to indicate that one cannot work at this level of abstraction but must introduce some additional concrete ingredient.

Note that this trick is in many ways the antithesis of §2.1.9, because by passing to a special case, one often makes the problem more concrete, with more things that one is now able to start trying. However, the two tricks can work together. One particularly useful "advanced move" in mathematical problem solving is to first abstract the problem to a more general one, and then consider a special case of that more abstract problem which is not directly related to the original one, but is somehow simpler than the original while still capturing some of the essence of the difficulty. Attacking this alternate problem can then lead to some indirect but important ways to make progress on the original problem.

2.1.11. Exploit Zeno's paradox: a single epsilon can be cut up into countably many sub-epsilons. A particularly useful fact in measure theory is that one can cut up a single epsilon into countably many pieces, for instance by using the geometric series identity

$$\varepsilon = \varepsilon/2 + \varepsilon/4 + \varepsilon/8 + \ldots;$$

this observation arguably goes all the way back to Zeno. As such, even if one only has an epsilon of room budgeted for a problem, one can still use this budget to do a countably infinite number of things. This fact underlies many of the countable additivity and subadditivity properties in measure theory, and also makes the ability to approximate rough objects by smoother ones to be useful even when countably many rough objects need to be approximated. (Exercise 1.2.3 is a typical example in which this trick is used.)

In general, one should be alert to this sort of trick when one has to spend an epsilon or so on an infinite number of objects. If one was forced to spend the same epsilon on each object, one would soon end up with an unacceptable loss; but if one can get away with using a different epsilon each time, then Zeno's trick comes in very handy.

2.1.12. If you expand your way to a double sum, a double integral, a sum of an integral, or an integral of a sum, try interchanging the two operations. Or, to put it another way: "The Fubini-Tonelli theorem (Corollary 1.7.23) is your friend". Provided that one is either in the unsigned or absolutely convergent worlds, this theorem allows you to interchange sums and integrals with each other. In many cases, a double sum or integral that is difficult to sum in one direction can become easier to sum (or at least to upper bound, which is often all that one needs in analysis). In fact, if in the course of expanding an expression, you encounter such a double sum or integral, you should reflexively try interchanging the operations to see if the resulting expression looks any simpler.

Note that in some cases the parameters in the summation may be constrained, and one may have to take a little care to sum it properly.

For instance,

$$\sum_{n=-\infty}^{\infty} \sum_{m=n}^{\infty} a_{m,n} \tag{2.1}$$

interchanges (assuming that the $a_{n,m}$ are either unsigned or absolutely convergent) to

$$\sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{m} a_{m,n}$$

(why? try plotting the set of pairs (m,n) that appear in both). If one is having trouble interchanging constrained sums or integrals, one solution is to re-express the constraint using indicator functions. For instance, one can rewrite the constrained sum (2.1) as the unconstrained sum

$$\sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} 1_{m \geq n} a_{m,n}$$

(extending the domain of $a_{m,n}$ if necessary), at which point interchanging the summations is easily accomplished.

The following point is obvious, but bears mentioning explicitly: while the interchanging sums/integrals trick can be very powerful, one should *not* apply it twice in a row to the same double sum or double operation, unless one is doing something genuinely non-trivial in between those two applications. So, after one exchanges a sum or integral, the next move should be something other than another exchange (unless one is dealing with a triple or higher sum or integral).

A related move (not so commonly used in measure theory, but occurring in other areas of analysis, particularly those involving the geometry of Euclidean spaces) is to merge two sums or integrals into a single sum or integral over the product space, in order to use some additional feature of the product space (e.g. rotation symmetry) that is not readily visible in the factor spaces. The classic example of this trick is the evaluation of the gaussian integral $\int_{-\infty}^{\infty} e^{-x^2}\ dx$ by squaring it, rewriting that square as the two-dimensional gaussian integral $\int_{\mathbf{R}^2} e^{-x^2-y^2}\ dxdy$, and then switching to polar coordinates.

2.1.13. Pointwise control, uniform control, and integrated (average) control are all partially convertible to each other.

There are three main ways to control functions (or sequences of functions, etc.) in analysis. Firstly, there is *pointwise control*, in which one can control the function at every point (or almost every point), but in a non-uniform way. Then there is *uniform control*, where one can control the function in the same way at most points (possibly throwing out a set of zero measure, or small measure). Finally, there is *integrated control* (or control "on the average"), in which one controls the integral of a function, rather than the pointwise values of that function.

It is important to realise that control of one type can often be partially converted to another type. Simple examples include the deduction of pointwise convergence from uniform convergence, or integrating a pointwise bound $f(x) \leq g(x)$ to obtain an integrated bound $\int f \leq \int g$. Of course, these conversions are not reversible and thus lose some information; not every pointwise convergent sequence is uniformly convergent, and an integral bound does not imply a pointwise bound. However, one can partially reverse such implications if one is willing to throw away an exceptional set (§2.1.7). For instance, Egorov's theorem (Theorem 1.3.26) lets one convert pointwise convergence to (local) uniform convergence after throwing away an exceptional set, and Markov's inequality (Exercise 1.4.36(vi)) lets one convert integral bounds to pointwise bounds, again after throwing away an exceptional set.

2.1.14. If the conclusion and hypotheses look particularly close to each other, just expand out all the definitions and follow your nose. This trick is particularly useful when building the most basic foundations of a theory. Here, one may not need to experiment too much with generalisations, abstractions, or special cases, or try to marshall a lot of possibly relevant facts about the objects being studied: sometimes, all one has to do is go back to first principles, write out all the definitions with their epsilons and deltas, and start plugging away at the problem.

Knowing when to just follow one's nose, and when to instead look for a more high-level approach to a problem, can require some judgement or experience. A direct approach tends to work best when the conclusion and hypothesis already look quite similar to each other (e.g. they both state that a certain set or family of sets is measurable, or they both state that a certain function or family of functions is continuous, etc.). But when the conclusion looks quite different from the hypotheses (e.g. the conclusion is some sort of integral identity, and the hypotheses involve measurability or convergence properties), then one may need to use more sophisticated tools than what one can easily get from using first principles.

2.1.15. Don't worry too much about exactly what ε (or δ, or N, etc.) needs to be. It can usually be chosen or tweaked later if necessary. Often in the middle of an argument, you will want to use some fact that involves a parameter, such as ε, that you are completely free to choose (subject of course to reasonable constraints, such as requiring ε to be positive). For instance, you may have a measurable set and decide to approximate it from above by an open set of at most ε more measure. But it may not be obvious exactly what value to give this parameter ε; you have so many choices available that you don't know which one to pick!

In many cases, one can postpone thinking about this problem by leaving ε undetermined for now, and continuing on with one's argument, which will gradually start being decorated with ε's all over the place. At some point, one will need ε to do something (and, in the particular case of ε, "doing something" almost always means "being small enough"), e.g. one may need $3n\varepsilon$ to be less than δ, where n, δ are some other positive quantities in one's problem that do not depend on ε. At this point, one could now set ε to be whatever is needed to get past this step in the argument, e.g. one could set ε to equal $\delta/4n$. But perhaps one still wishes to retain the freedom to set ε because it might come in handy later. In that case, one sets aside the requirement "$3n\varepsilon < \delta$" and keeps going. Perhaps a bit later on, one might need ε to do something else; for instance, one might also need $5\varepsilon \leq 2^{-n}$. Once one has compiled the complete "wish list" of everything one wishes one's parameters to do, then one can finally make the decision of what value to set those parameters equal to. For instance, if the above two inequalities are the only inequalities required of ε, one can choose ε equal to $\min(\delta/4n, 2^{-n}/5)$. This may be a choice of ε which was not obvious at the start of the argument, but becomes so as the argument progresses.

There is however one big caveat when adopting this "choose parameters later" approach, which is that one needs to avoid a circular dependence of constants. For instance, it is perfectly fine to have two arbitrary parameters ε and δ floating around unspecified for most of the argument, until at some point you realise that you need ε to be smaller than δ, and so one chooses ε accordingly (e.g. one sets it to equal $\delta/2$). Or, perhaps instead one needs δ to be smaller than ε, and so sets δ equal to $\varepsilon/2$. One can execute either of these two choices separately, but of course one cannot perform them simultaneously; this sets up an inconsistent circular dependency in which ε needs to be defined after δ is chosen, and δ can only be chosen after ε is fixed. So, if one is going to delay choosing a parameter such as ε until later, it becomes important to mentally keep track of what objects in one's argument depend on ε, and which ones are independent of ε. One can choose ε in terms of the latter quantities, but one usually cannot do so in terms of the former quantities (unless one takes the care to show that the interlinked constraints between the quantities are still consistent, and thus simultaneously satisfiable).

2.1.16. Once one has started to lose some constants, don't be hesitant to lose some more. Many techniques in analysis end up giving inequalities that are inefficient by a constant factor. For instance, any argument involving dyadic decomposition and powers of two tends to involve losses of factors of 2. When arguing using balls in Euclidean space, one sometimes loses factors involving the volume of the unit ball (although this factor often cancels itself out if one tracks it more carefully). And so forth. However, in many cases these constant factors end up being of little importance: an upper bound of 2ε or 100ε is often just as good as an upper bound of ε for the purposes of analysis (cf. §2.1.15). So it is often best not to invest too much energy in carefully computing and optimising these constants; giving these constants a symbol such as C, and not worrying about their exact value, is often the simplest approach. (One can also use asymptotic notation, such as $O()$, which is very convenient to use once you know how it works.)

Now there are some cases in which one really does not want to lose any constants at all. For instance, if one is using §2.1.1 to prove that $X = Y$, it is not enough to show that $X \leq 2Y$ and $Y \leq 2X$; one really needs to show $X \leq Y$ and $Y \leq X$ without losing any constants. (But proving $X \leq (1+\varepsilon)Y$ and $Y \leq (1+\varepsilon)X$ is OK, by §2.1.2.) But once one has already performed one step that loses a constant, there is little further to be lost by losing more; there can be a big difference between $X \leq Y$ and $X \leq 2Y$, but there is little difference in practice between $X \leq 2Y$ and $X \leq 100Y$, at least for the purposes of mathematical analysis. At that stage, one should put oneself in the mental mode of thought where "constants don't matter", which can lead to some simplifications. For instance, if one has to estimate a sum $X+Y$ of two positive quantities, one can start using such estimates as

$$\max(X,Y) \leq X+Y \leq 2\max(X,Y),$$

which says that, up to a factor of 2, $X+Y$ is the same thing as $\max(X,Y)$. In some cases the latter is easier to work with (e.g. $\max(X,Y)^n$ is equal to $\max(X^n,Y^n)$, whereas the formula for $(X+Y)^n$ is messier).

2.1.17. One can often pass to a subsequence to improve the convergence properties. In real analysis, one often ends up possessing a sequence of objects, such as a sequence of functions f_n, which may converge in some rather slow or weak fashion to a limit f. Often, one can improve the convergence of this sequence by passing to a subsequence. For instance:

- In a metric space, if a sequence x_n converges to a limit x, then one can find a subsequence x_{n_j} which converges quickly to the same limit x, for instance one can ensure that $d(x_{n_j},x) \leq 2^{-j}$ (or one can replace 2^{-j} with any other positive expression depending on j). In particular, one can make $\sum_{j=1}^\infty d(x_{n_j},x)$ and $\sum_{j=1}^\infty d(x_{n_j},x_{n_{j+1}})$ absolutely convergent, which is sometimes useful.
- A sequence of functions that converges in L^1 norm or in measure can be refined to a subsequence that converges pointwise almost everywhere as well.

- A sequence in a (sequentially) compact space may not converge at all, but some subsequence of it will always converge.
- The pigeonhole principle: A sequence which takes only finitely many values has a subsequence that is constant. More generally, a sequence which lives in the union of finitely many sets has a subsequence that lives in just one of these sets.

Often, the subsequence is good enough for one's applications, and there are also a number of ways to get back from a subsequence to the original sequence, such as:

- In a metric space, if you know that x_n is a Cauchy sequence, and some subsequence of x_n already converges to x, then this drags the entire sequence with it, i.e. x_n converges to x also.
- The *Urysohn subsequence principle*: in a topological space, if every subsequence of a sequence x_n itself has a subsequence that converges to a limit x, then the entire sequence converges to x.

2.1.18. A real limit can be viewed as a meeting of the limit superior and limit inferior. A sequence x_n of real numbers does not necessarily have a limit $\lim_{n\to\infty} x_n$, but the limit superior $\limsup_{n\to\infty} x_n := \inf_N \sup_{n>N} x_n$ and the limit inferior $\liminf_{n\to\infty} x_n = \sup_N \inf_{n>N} x_n$ always exist (though they may be infinite), and can be easily defined in terms of infima and suprema. Because of this, it is often convenient to work with the lim sup and lim inf instead of a limit. For instance, to show that a limit $\lim_{n\to\infty} x_n$ exists, it suffices to show that

$$\limsup_{n\to\infty} x_n \leq \liminf_{n\to\infty} x_n + \varepsilon$$

for all $\varepsilon > 0$. In a similar spirit, to show that a sequence x_n of real numbers converges to zero, it suffices to show that

$$\limsup_{n\to\infty} |x_n| \leq \varepsilon$$

for all $\varepsilon > 0$. It can be more convenient to work with lim sups and lim infs instead of limits because one does not need to worry about the issue of whether the limit exists or not, and many tools (notably Fatou's lemma and its relatives) still work in this setting. One should however be cautious that lim sups and lim infs tend to have only one half of the linearity properties that limits do; for instance, lim sups are subadditive but not necessarily additive, while lim infs are superadditive but not necessarily additive.

The proof of the monotone differentiation theorem (Theorem 1.6.25) given in the text relies quite heavily on this strategy.

2.2. The Rademacher differentiation theorem

The Fubini-Tonelli theorem (Corollary 1.7.23) is often used in extending lower-dimensional results to higher-dimensional ones. We illustrate this by extending the one-dimensional Lipschitz differentiation theorem (Exercise 1.6.41) to higher dimensions, obtaining the *Rademacher differentiation theorem*.

We first recall some higher-dimensional definitions:

Definition 2.2.1 (Lipschitz continuity)**.** A function $f : X \to Y$ from one metric space (X, d_X) to another (Y, d_Y) is said to be *Lipschitz continuous* if there exists a constant $C > 0$ such that $d_Y(f(x), f(x')) \leq Cd_X(x, x')$ for all $x, x' \in X$. (In the applications of this section, X will be $\mathbf{R}^d$ and Y will be $\mathbf{R}$, with the usual metrics.)

Exercise 2.2.1. Show that Lipschitz continuous functions are uniformly continuous, and hence continuous. Then give an example of a uniformly continuous function $f : [0, 1] \to [0, 1]$ that is not Lipschitz continuous.

Definition 2.2.2 (Differentiability)**.** Let $f : \mathbf{R}^d \to \mathbf{R}$ be a function, and let $x_0 \in \mathbf{R}^d$. For any $v \in \mathbf{R}^d$, we say that f is *directionally differentiable* at x_0 in the direction v if the limit

$$D_v f(x_0) := \lim_{h \to 0; h \in \mathbf{R} \setminus \{0\}} \frac{f(x_0 + hv) - f(x_0)}{h}$$

exists, in which case we call $D_v f(x_0)$ the *directional derivative* of f at x_0 in this direction. If $v = e_i$ is one of the standard basis vectors $e_1, \ldots, e_d$ of $\mathbf{R}^d$, we write $D_v f(x_0)$ as $\frac{\partial f}{\partial x_i}(x_0)$, and refer to this as the *partial derivative* of f at x_0 in the e_i direction.

We say that f is *totally differentiable* at x_0 if there exists a vector $\nabla f(x_0) \in \mathbf{R}^d$ with the property that

$$\lim_{h \to 0; h \to \mathbf{R}^d \backslash \{0\}} \frac{f(x_0 + h) - f(x_0) - h \cdot \nabla f(x_0)}{|h|} = 0,$$

where $v \cdot w$ is the usual dot product on $\mathbf{R}^d$. We refer to $\nabla f(x_0)$ (if it exists) as the *gradient* of f at x_0.

Remark 2.2.3. From the viewpoint of differential geometry, it is better to work not with the gradient vector $\nabla f(x_0) \in \mathbf{R}^d$, but rather with the derivative covector $df(x_0) : \mathbf{R}^d \to \mathbf{R}$ given by $df(x_0) : v \mapsto \nabla f(x_0) \cdot v$. This is because one can then define the notion of total differentiability without any mention of the Euclidean dot product, which allows one to extend this notion to other manifolds in which there is no Euclidean (or more generally, Riemannian) structure. However, as we are working exclusively in Euclidean space for this application, this distinction will not be important for us.

Total differentiability implies directional and partial differentiability, but not conversely, as the following three exercises demonstrate.

Exercise 2.2.2 (Total differentiability implies directional and partial differentiability)**.** Show that if $f : \mathbf{R}^d \to \mathbf{R}$ is totally differentiable at x_0, then it is directionally differentiable at x_0 in each direction $v \in \mathbf{R}^d$, and one has the formula

$$D_v f(x_0) = v \cdot \nabla f(x_0). \tag{2.2}$$

In particular, the partial derivatives $\frac{\partial f}{\partial x_i} f(x_0)$ exist for $i = 1, \ldots, d$ and

$$\nabla f(x_0) = \left(\frac{\partial f}{\partial x_1}(x_0), \ldots, \frac{\partial f}{\partial x_d}(x_0) \right). \tag{2.3}$$

Exercise 2.2.3 (Continuous partial differentiability implies total differentiability)**.** Let $f : \mathbf{R}^d \to \mathbf{R}$ be such that the partial derivatives $\frac{\partial f}{\partial x_i} : \mathbf{R}^d \to \mathbf{R}$ exist everywhere and are continuous. Then show that f is totally differentiable everywhere, which in particular implies that the gradient is given by the formula (2.3) and the directional derivatives are given by (2.2).

Exercise 2.2.4 (Directional differentiability does not imply total differentiability)**.** Let $f : \mathbf{R}^2 \to \mathbf{R}$ be defined by setting $f(0,0) := 0$ and $f(x_1, x_2) := \frac{x_1 x_2^2}{x_1^2 + x_2^2}$ for $(x_1, x_2) \in \mathbf{R}^2 \backslash \{(0,0)\}$. Show that the directional derivatives $D_v f(x)$ exist for all $x, v \in \mathbf{R}^2$ (so in particular, the partial derivatives exist), but that f is not totally differentiable at the origin $(0,0)$.

Now we can state the *Rademacher differentiation theorem.*

Theorem 2.2.4 (Rademacher differentiation theorem)**.** *Let $f : \mathbf{R}^d \to \mathbf{R}$ be Lipschitz continuous. Then f is totally differentiable at x_0 for almost every $x_0 \in \mathbf{R}^d$.*

Note that the $d = 1$ case of this theorem is Exercise 1.6.41, and indeed we will use the one-dimensional theorem to imply the higher-dimensional one, though there will be some technical issues due to the gap between directional and total differentiability.

Proof. The strategy here is to first aim for the more modest goal of directional differentiability, and then find a way to link the directional derivatives together to get total differentiability.

Let $v, x_0 \in \mathbf{R}^d$. As f is continuous, we see that in order for the directional derivative

$$D_v f(x_0) := \lim_{h \to 0; h \in \mathbf{R} \backslash \{0\}} \frac{f(x_0 + hv) - f(x_0)}{h}$$

to exist, it suffices to let h range in the dense subset $\mathbf{Q} \backslash \{0\}$ of $\mathbf{R} \backslash \{0\}$ for the purposes of determing whether the limit exists. In particular, $D_v f(x_0)$ exists if and only if

$$\limsup_{h \to 0; h \in \mathbf{Q} \backslash \{0\}} \frac{f(x_0 + hv) - f(x_0)}{h} = \liminf_{h \to 0; h \in \mathbf{Q} \backslash \{0\}} \frac{f(x_0 + hv) - f(x_0)}{h}.$$

From this we easily conclude that for each direction $v \in \mathbf{R}^d$, the set

$$E_v := \{x_0 \in \mathbf{R}^d : D_v f(x_0) \text{ does not exist}\}$$

is Lebesgue measurable in $\mathbf{R}^d$ (indeed, it is even Borel measurable). A similar argument reveals that $D_v f$ is a measurable function outside of E_v. From the Lipschitz nature of f, we see that $D_v f$ is also a bounded function.

Now we claim that E_v is a null set for each v. For $v = 0$ E_v is clearly empty, so we may assume $v \neq 0$. Applying an invertible linear transformation to map v to e_1 (noting that such transformations will map Lipschitz functions to Lispchitz functions, and null sets to null sets) we may assume without loss of generality that v is the basis vector e_1. Thus our task is now to show that $\frac{\partial f}{\partial x_1}(x)$ exists for almost every $x \in \mathbf{R}^d$.

We now split $\mathbf{R}^d$ as $\mathbf{R} \times \mathbf{R}^{d-1}$. For each $x_0 \in \mathbf{R}$ and $y_0 \in \mathbf{R}^{d-1}$, we see from the definitions that $\frac{\partial f}{\partial x_1}(x_0, y_0)$ exists if and only if the one-dimensional function $x \mapsto f(x, y_0)$ is differentiable at x_0. But this function is Lipschitz continuous (this is inherited from the Lipschitz continuity of f), and so we see that for each fixed $y_0 \in \mathbf{R}^{d-1}$, the set $E^{y_0} := \{x_0 \in \mathbf{R} : (x_0, y_0) \in E\}$ is a null set in $\mathbf{R}$. Applying Tonelli's theorem for sets (Corollary 1.7.19), we conclude that E is a null set as required.

We would like to now conclude that $\bigcup_{v \in \mathbf{R}^d} E_v$ is a null set, but there are uncountably many v's, so this is not directly possible. However, as $\mathbf{Q}^d$ is rational, we can at least assert that $E := \bigcup_{v \in \mathbf{Q}^d} E_v$ is a null set. In particular, for almost every $x_0 \in \mathbf{R}^d$, f is directionally differentiable in every rational direction $v \in \mathbf{Q}^d$.

Now we perform an important trick, in which we interpret the directional derivative $D_v f$ as a *weak derivative*. We already know that $D_v f$ is almost everywhere defined, bounded and measurable. Now let $g : \mathbf{R}^d \to \mathbf{R}$ be any function that is compactly supported and Lipschitz continuous. We investigate the integral

$$\int_{\mathbf{R}^d} D_v f(x) g(x)\ dx.$$

This integral is absolutely convergent since $D_v f(x)$ is bounded and measurable, and $g(x)$ is continuous and compactly supported, hence bounded. We expand this out as

$$\int_{\mathbf{R}^d} \lim_{h \to 0; h \in \mathbf{R} \backslash \{0\}} \frac{f(x + hv) - f(x)}{h} g(x)\ dx.$$

Note (from the Lipschitz nature of f) that the expression $\frac{f(x+hv)-f(x)}{h} g(x)$ is bounded uniformly in h and x, and is also uniformly compactly supported in x for h in a bounded set. We may thus apply the dominated convergence theorem (Theorem 1.4.49) to pull the limit out of the integral to obtain

$$\lim_{h\to 0; h\in\mathbf{R}\backslash\{0\}} \int_{\mathbf{R}^d} \frac{f(x+hv)-f(x)}{h} g(x)\ dx.$$

Now, from translation invariance of the Lebesgue integral (Exercise 1.3.15) we have

$$\int_{\mathbf{R}^d} f(x+hv)g(x)\ dx = \int_{\mathbf{R}^d} f(x)g(x-hv)\ dx$$

and so (by the lienarity of the Lebesgue integral) we may rearrange the previous expression as

$$\lim_{h\to 0; h\in\mathbf{R}\backslash\{0\}} \int_{\mathbf{R}^d} f(x)\frac{g(x-hv)-g(x)}{h}\ dx.$$

Now, as g is Lipschitz, we know that $\frac{g(x-hv)-g(x)}{h}$ is uniformly bounded and converges pointwise almost everywhere to $D_{-v}g(x)$ as $h \to 0$. We may thus apply the dominated convergence theorem again and end up with the *integration by parts formula*

$$\int_{\mathbf{R}^d} D_v f(x)g(x)\ dx = \int_{\mathbf{R}^d} f(x)D_{-v}g(x)\ dx. \tag{2.4}$$

This formula moves the directional derivative operator D_v from f over to g. At present, this does not look like much of an advantage, because g is the same sort of function that f is. However, the key point is that we can choose g to be whatever we please, whereas f is fixed. In particular, we can choose g to be a compactly supported, *continuously differentiable* function (such functions are Lipschitz from the fundamental theorem of calculus, as their derivatives are bounded). By Exercise 2.2.3, one has $D_{-v}g = -v\cdot\nabla g$ for such functions, and so

$$\int_{\mathbf{R}^d} D_v f(x)g(x)\ dx = -\int_{\mathbf{R}^d} f(x)(v\cdot\nabla g)(x)\ dx.$$

The right-hand side is linear in v, and so the left-hand side must be linear in v also. In particular, if $v = (v_1, \ldots, v_d)$, then we have

$$\int_{\mathbf{R}^d} D_v f(x)g(x)\ dx = \sum_{j=1}^d v_j \int_{\mathbf{R}^d} D_{e_j} f(x)g(x)\ dx.$$

If we define the gradient candidate function

$$\nabla f(x) := (D_{e_1} f(x), \ldots, D_{e_d} f(x)) = (\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_d}(x))$$

(note that this function is well-defined almost everywhere, even though we don't know yet whether f is totally differentiable almost everywhere), we thus have

$$\int_{\mathbf{R}^d} (D_v f - v \cdot \nabla f)(x) g(x)\ dx = 0$$

for all compactly supported, continuously differentiable g. This implies (see Exercise 2.2.5 below) that $F_v := D_v f - v \cdot \nabla f$ vanishes almost everywhere, thus (by countable subadditivity) we have

$$D_v f(x_0) = v \cdot \nabla f(x_0) \tag{2.5}$$

for almost every $x_0 \in \mathbf{R}^d$ and every $v \in \mathbf{Q}^d$.

Let x_0 be such that (2.5) holds for all $v \in \mathbf{Q}^d$. We claim that this forces f to be totally differentiable at x_0, which would give the claim. Let $F : \mathbf{R}^d \to \mathbf{R}^d$ be the modified function

$$F(h) := f(x_0 + h) - f(x_0) - h \cdot \nabla f(x_0).$$

Our objective is to show that

$$\lim_{h \to 0; h \in \mathbf{R}^d \backslash \{0\}} |F(h)|/|h| = 0.$$

On the other hand, we have $F(0) = 0$, F is Lipschitz, and from (2.5) we see that $D_v F(0) = 0$ for every $v \in \mathbf{Q}^d$.

Let $\varepsilon > 0$, and suppose that $h \in \mathbf{R}^d \backslash \{0\}$. Then we can write $h = ru$ where $r := |h|$ and $u := h/|h|$ lies on the unit sphere. This u need not lie in $\mathbf{Q}^d$, but we can approximate it by some vector $v \in \mathbf{Q}^d$ with $|u - v| \leq \varepsilon$. Furthermore, by the total boundedness of the unit sphere, we can make v lie in a finite subset V_ε of $\mathbf{Q}^d$ that only depends on ε (and on d).

Since $D_v F(0) = 0$ for all $v \in V_\varepsilon$, we see (by making $|h|$ small enough depending on V_ε) that we have

$$|\frac{F(rv) - F(0)}{r}| \leq \varepsilon$$

for all $v \in V_\varepsilon$, and thus

$$|F(rv)| \leq \varepsilon r.$$

On the other hand, from the Lipschitz nature of F, we have

$$|F(ru) - F(rv)| \leq Cr|u - v| \leq Cr\varepsilon$$

where C is the Lipschitz constant of F. As $h = ru$, we conclude that

$$|F(h)| \leq (C + 1)r\varepsilon.$$

In other words, we have shown that

$$|F(h)|/|h| \leq (C + 1)\varepsilon$$

whenever $|h|$ is sufficiently small depending on ε. Letting $\varepsilon \to 0$, we obtain the claim. $\square$

Exercise 2.2.5. Let $F : \mathbf{R}^d \to \mathbf{R}$ be a locally integrable function with the property that $\int_{\mathbf{R}^d} F(x)g(x)\, dx = 0$ whenever g is a compactly supported, continuously differentiable function. Show that F is zero almost everywhere. (*Hint:* if not, use the Lebesgue differentiation theorem to find a Lebesgue point x_0 of F for which $F(x_0) \neq 0$, then pick a g which is supported in a sufficiently small neighbourhood of x_0.)

2.3. Probability spaces

In this section we isolate an important special type of measure space, namely a *probability space*. As the name suggests, these spaces are of fundamental importance in the foundations of probability, although it should be emphasised that probability theory should *not* be viewed as the study of probability spaces, as these are merely models for the true objects of study of that theory, namely the behaviour of random events and random variables. (See §??? of ??? for further discussion of this point. **Crossreference will be added once the remaining sections of the blog are converted to book form - T.**) This text will however not be focused on applications to probability theory

Definition 2.3.1 (Probability space)**.** A *probability space* is a measure space $(\Omega, \mathcal{F}, \mathbf{P})$ of total measure 1: $\mathbf{P}(\Omega) = 1$. The measure $\mathbf{P}$ is known as a *probability measure*.

Note the change of notation: whereas measure spaces are traditionally denoted by symbols such as $(X, \mathcal{B}, \mu)$, probability spaces are traditionally denoted by symbols such as $(\Omega, \mathcal{F}, \mathbf{P})$. Of course, such notational changes have no impact on the underlying mathematical formalism, but they reflect the different cultures of measure theory and probability theory. In particular, the various components Ω, $\mathcal{F}$, $\mathbf{P}$ carry the following interpretations in probability theory, that are absent in other applications of measure theory:

(i) The space Ω is known as the *sample space*, and is interpreted as the set of all possible *states* $\omega \in \Omega$ that a random system could be in.

(ii) The σ-algebra $\mathcal{F}$ is known as the *event space*, and is interpreted as the set of all possible *events* $E \in \mathcal{F}$ that one can measure.

(iii) The measure $\mathbf{P}(E)$ of an event is known as the *probability* of that event.

The various axioms of a probability space then formalise the *foundational axioms of probability*, as set out by Kolmogorov.

Example 2.3.2 (Normalised measure)**.** Given any measure space $(X, \mathcal{B}, \mu)$ with $0 < \mu(X) < +\infty$, the space $(X, \mathcal{B}, \frac{1}{\mu(X)}\mu)$ is a probability space. For instance, if Ω is a non-empty finite set with the discrete σ-algebra 2^Ω and the counting measure $\#$, then the *normalised counting measure* $\frac{1}{\#\Omega}\#$ is a probability measure (known as the *(discrete) uniform probability measure* on Ω), and $(\Omega, 2^\Omega, \frac{1}{\#\Omega}\#)$ is a probability space. In probability theory, this probability spaces models the act of drawing an element of the discrete set Ω uniformly at random.

Similarly, if $\Omega \subset \mathbf{R}^d$ is a Lebesgue measurable set of positive finite Lebesgue measure, $0 < m(\Omega) < \infty$, then $(\Omega, \mathcal{L}[\mathbf{R}^d]\downharpoonright_\Omega, \frac{1}{m(\Omega)} m\downharpoonright_\Omega)$ is a probability space. The probability measure $\frac{1}{m(\Omega)} m\downharpoonright_\Omega$ is known as the *(continuous) uniform probability measure* on Ω. In probability theory, this probability spaces models the act of drawing an element of the continuous set Ω uniformly at random.

Example 2.3.3 (Discrete and continuous probability measures)**.** If Ω is a (possibly infinite) non-empty set with the discrete σ-algebra 2^Ω, and if $(p_\omega)_{\omega\in\Omega}$ are a collection of real numbers in $[0,1]$ with $\sum_{\omega\in\Omega} p_\omega = 1$, then the probability measure $\mathbf{P}$ defined by $\mathbf{P} :=$

$\sum_{\omega\in\Omega} p_\omega \delta_\omega$, or in other words

$$\mathbf{P}(E) := \sum_{\omega\in E} p_\omega,$$

is indeed a probability measure, and $(\Omega, 2^\Omega, \mathbf{P})$ is a probability space. The function $\omega \mapsto p_\omega$ is known as the *(discrete) probability distribution* of the state variable ω.

Similarly, if Ω is a Lebesgue measurable subset of $\mathbf{R}^d$ of positive (and possibly infinite) measure, and $f : \Omega \to [0, +\infty]$ is a Lebesgue measurable function on Ω (where of course we restrict the Lebesgue measure space on $\mathbf{R}^d$ to Ω in the usual fashion) with $\int_\Omega f(x)\ dx = 1$, then $(\Omega, \mathcal{L}[\mathbf{R}^d]\ \lfloor_\Omega, \mathbf{P})$ is a probability space, where $\mathbf{P} := m_f$ is the measure

$$\mathbf{P}(E) := \int_\Omega 1_E(x) f(x)\ dx = \int_E f(x)\ dx.$$

The function f is known as the *(continuous) probability density* of the state variable ω. (This density is not quite unique, since one can modify it on a set of probability zero, but it is well-defined up to this ambiguity. See §1.2 of *An epsilon of room, Vol. I* for further discussion.)

Exercise 2.3.1 (No translation-invariant random integer)**.** Show that there is no probability measure $\mathbf{P}$ on the integers $\mathbf{Z}$ with the discrete σ-algebra $2^{\mathbf{Z}}$ with the translation-invariance property $\mathbf{P}(E+n) = \mathbf{P}(E)$ for every event $E \in 2^{\mathbf{Z}}$ and every integer n.

Exercise 2.3.2 (No translation-invariant random real)**.** Show that there is no probability measure $\mathbf{P}$ on the reals $\mathbf{R}$ with the Lebesgue σ-algebra $\mathcal{L}[\mathbf{R}]$ with the translation-invariance property $\mathbf{P}(E+x) = \mathbf{P}(E)$ for every event $E \in \mathcal{L}[\mathbf{R}]$ and every real x.

Many concepts in measure theory are of importance in probability theory, although the terminology is changed to reflect the different perspective on the subject. For instance, the notion of a property holding almost everywhere is now replaced with that of a property holding *almost surely*. A measurable function is now referred to as a *random variable* and is often denoted by symbols such as X, and the integral of that function on the probability space (if the random variable is unsigned or absolutely convergent) is known as the *expectation* of that random variable, and is denoted $\mathbf{E}(X)$. Thus, for instance, the Borel-Cantelli lemma (Exercise 1.4.44) now reads as follows: given any sequence $E_1, E_2, E_3, \ldots$ of events such that $\sum_{n=1}^\infty \mathbf{P}(E_n) < \infty$, it is almost surely true that at most finitely many of these events hold. In a similar spirit, Markov's inequality (Exercise 1.4.36(vi)) becomes the assertion that $\mathbf{P}(X \geq \lambda) \leq \frac{1}{\lambda}\mathbf{E}X$ for any non-negative random variable X and any $0 < \lambda < \infty$.

2.4. Infinite product spaces and the Kolmogorov extension theorem

In Section 1.7.4 we considered the product of two sets, measurable spaces, or (σ-finite) measure spaces. We now consider how to generalise this concept to products of more than two such spaces. The axioms of set theory allow us to form a Cartesian product $X_A := \prod_{\alpha \in A} X_\alpha$ of any family $(X_\alpha)_{\alpha \in A}$ of sets indexed by another set A, which consists of the space of all tuples $x_A = (x_\alpha)_{\alpha \in A}$ indexed by A, for which $x_\alpha \in X_\alpha$ for all $\alpha \in A$. This concept allows for a succinct formulation of the *axiom of choice* (Axiom 0.0.4), namely that an arbitrary Cartesian product of non-empty sets remains non-empty.

For any $\beta \in A$, we have the coordinate projection maps $\pi_\beta : X_A \to X_\beta$ defined by $\pi_\beta((x_\alpha)_{\alpha \in A}) := x_\beta$. More generally, given any $B \subset A$, we define the partial projections $\pi_B : X_A \to X_B$ to the partial product space $X_B := \prod_{\alpha \in B} X_\alpha$ by $\pi_B((x_\alpha)_{\alpha \in A}) := (x_\alpha)_{\alpha \in B}$. More generally still, given two subsets $C \subset B \subset A$, we have the partial subprojections $\pi_{C \leftarrow B} : X_B \to X_C$ defined by $\pi_{C \leftarrow B}((x_\alpha)_{\alpha \in B}) := (x_\alpha)_{\alpha \in C}$. These partial subprojections obey the composition law $\pi_{D \leftarrow C} \circ \pi_{C \leftarrow B} := \pi_{D \leftarrow B}$ for all $D \subset C \subset B \subset A$ (and thus form a very simple example of a *category*).

As before, given any σ-algebra $\mathcal{B}_\beta$ on X_β, we can pull it back by π_β to create a σ-algebra

$$\pi_\beta^*(\mathcal{B}_\beta) := \{\pi_\beta^{-1}(E_\beta) : E_\beta \in \mathcal{B}_\beta\}$$

on X_A. One easily verifies that this is indeed a σ-algebra. Informally, $\pi_\beta^*(\mathcal{B}_\beta)$ describes those sets (or "events", if one is thinking in probabilistic terms) that depend only on the x_β coordinate of the state $x_A = (x_\alpha)_{\alpha \in A}$, and whose dependence on x_β is $\mathcal{B}_\beta$-measurable. We can then define the product σ-algebra

$$\prod_{\beta \in A} \mathcal{B}_\beta := \langle \bigcup_{\beta \in A} \pi_\beta^*(\mathcal{B}_\beta) \rangle.$$

We have a generalisation of Exercise 1.7.18:

Exercise 2.4.1. Let $((X_\alpha, \mathcal{B}_\alpha))_{\alpha \in A}$ be a family of measurable spaces. For any $B \subset A$, write $\mathcal{B}_B := \prod_{\beta \in B} \mathcal{B}_\beta$.

(1) Show that $\mathcal{B}_A$ is the coarsest σ-algebra on X_A that makes the projection maps π_β measurable morphisms for all $\beta \in A$.

(2) Show that for each $B \subset A$, that π_B is a measurable morphism from $(X_A, \mathcal{B}_A)$ to $(X_B, \mathcal{B}_B)$.

(3) If E in $\mathcal{B}_A$, show that there exists an at most countable set $B \subset A$ and a set $E_B \in \mathcal{B}_B$ such that $E_A = \pi_B^{-1}(E_B)$. Informally, this asserts that a measurable event can only depend on at most countably many of the coefficients.

(4) If $f : X_A \to [0, +\infty]$ is $\mathcal{B}_A$-measurable, show that there exists an at most countable set $B \subset A$ and a $\mathcal{B}_B$-measurable function $f_B : X_B \to [0, +\infty]$ such that $f = f_B \circ \pi_B$.

(5) If A is at most countable, show that $\mathcal{B}_A$ is the σ-algebra generated by the sets $\prod_{\beta \in A} E_\beta$ with $E_\beta \in \mathcal{B}_\beta$ for all $\beta \in A$.

(6) On the other hand, show that if A is uncountable and the $\mathcal{B}_\alpha$ are all non-trivial, show that $\mathcal{B}_A$ is *not* the σ-algebra generated by sets $\prod_{\beta \in A} E_\beta$ with $E_\beta \in \mathcal{B}_\beta$ for all $\beta \in A$.

(7) If $B \subset A$, $E \in \mathcal{B}_A$, and $x_{A \backslash B} \in X_{A \backslash B}$, show that the set $E_{x_{A \backslash B}, B} := \{x_B \in X_B : (x_B, x_{A \backslash B}) \in E\}$ lies in $\mathcal{B}_B$, where we identify $X_B \times X_{A \backslash B}$ with X_A in the obvious manner.

(8) If $B \subset A$, $f : X_A \to [0, +\infty]$ is $\mathcal{B}_A$-measurable, and $x_{A \backslash B} \in X_{A \backslash B}$, show that the function $f_{x_{A \backslash B}, B} : x_B \to f(x_B, x_{A \backslash B})$ is $\mathcal{B}_B$-measurable.

Now we consider the problem of constructing a measure μ_A on the product space X_A. Any such measure μ_A will induce *pushforward measures* $\mu_B := (\pi_B)_* \mu_A$ on X_B (introduced in Exercise 1.4.38), thus

$$\mu_B(E_B) := \mu_A(\pi_B^{-1}(E_B))$$

for all $E_B \in \mathcal{B}_B$. These measures obey the compatibility relation

$$(\pi_{C \leftarrow B})_* \mu_B = \mu_C \tag{2.6}$$

whenever $C \subset B \subset A$, as can be easily seen by chasing the definitions.

One can then ask whether one can reconstruct μ_A from just from the projections μ_B to *finite* subsets B. This is possible in the important special case when the μ_B (and hence μ_A) are probability measures, provided one imposes an additional *inner regularity* hypothesis on the measures μ_B. More precisely:

Definition 2.4.1 (Inner regularity)**.** A (metrisable) inner regular measure space $(X, \mathcal{B}, \mu, d)$ is a measure space $(X, \mathcal{B}, \mu)$ equipped with a metric d such that

(1) Every compact set is measurable; and
(2) One has $\mu(E) = \sup_{K \subset E, K \text{ compact}} \mu(K)$ for all measurable E.

We say that μ is *inner regular* if it is associated to an inner regular measure space.

Thus for instance Lebesgue measure is inner regular, as are Dirac measures and counting measures. Indeed, most measures that one actually encounters in applications will be inner regular. For instance, any finite Borel measure on $\mathbf{R}^d$ (or more generally, on a locally compact, σ-compact space) is inner regular, as is any *Radon measure*; see §1.10 of *An epsilon of room, Vol. I.*

Remark 2.4.2. One can generalise the concept of an inner regular measure space to one which is given by a topology rather than a metric; Kolmogorov's extension theorem still holds in this more general setting, but requires *Tychonoff's theorem*, which is discussed in §1.8 of *An epsilon of room, Vol. I.* However, some minimal regularity hypotheses of a topological nature are needed to make the Kolmogorov extension theorem work, although this is usually not a severe restriction in practice.

Theorem 2.4.3 (Kolmogorov extension theorem)**.** *Let* $((X_\alpha, \mathcal{B}_\alpha), \mathcal{F}_\alpha)_{\alpha \in A}$ *be a family of measurable spaces* $(X_\alpha, \mathcal{B}_\alpha)$*, equipped with a topology*

$\mathcal{F}_\alpha$. For each finite $B \subset A$, let μ_B be an inner regular probability measure on $\mathcal{B}_B := \prod_{\alpha \in B} \mathcal{B}_\alpha$ with the product topology $\mathcal{F}_B := \prod_{\alpha \in B} \mathcal{F}_\alpha$, obeying the compatibility condition (2.6) whenever $C \subset B \subset A$ are two nested finite subsets of A. Then there exists a unique probability measure μ_A on $\mathcal{B}_A$ with the property that $(\pi_B)_* \mu_A = \mu_B$ for all finite $B \subset A$.

Proof. Our main tool here will be the Hahn-Kolmogorov extension theorem for pre-measures (Theorem 1.7.8), combined with the Heine-Borel theorem.

Let $\mathcal{B}_0$ be the set of all subsets of X_A that are of the form $\pi_B^{-1}(E_B)$ for some finite $B \subset A$ and some $E_B \in \mathcal{B}_B$. One easily verifies that this is a Boolean algebra that is contained in $\mathcal{B}_A$. We define a function $\mu_0 : \mathcal{B}_0 \to [0, +\infty]$ by setting

$$\mu_0(E) := \mu_B(E_B)$$

whenever E takes the form $\pi_B^{-1}(E_B)$ for some finite $B \subset A$ and $E_B \in \mathcal{B}_B$. Note that a set $E \in \mathcal{B}_0$ may have two different representations $E = \pi_B^{-1}(E_B) = \pi_{B'}^{-1}(E_{B'})$ for some finite $B, B' \subset A$, but then one must have $E_B = \pi_{B \leftarrow B \cup B'}(E_{B \cup B'})$ and $E_{B'} = \pi_{B' \leftarrow B \cup B'}(E_{B \cup B'})$, where $E_{B \cup B'} := \pi_{B \cup B'}(E)$. Applying (2.6), we see that

$$\mu_B(E_B) = \mu_{B \cup B'}(E_{B \cup B'})$$

and

$$\mu_{B'}(E_{B'}) = \mu_{B \cup B'}(E_{B \cup B'})$$

and thus $\mu_B(E_B) = \mu_{B'}(E_{B'})$. This shows that $\mu_0(E)$ is well defined. As the μ_B are probability measures, we see that $\mu_0(X_A) = 1$.

It is not difficult to see that μ_0 is finitely additive. We now claim that μ_0 is a pre-measure. In other words, we claim that if $E \in \mathcal{B}_0$ is the disjoint countable union $E = \bigcup_{n=1}^\infty E_n$ of sets $E_n \in \mathcal{B}_0$, then $\mu_0(E) = \sum_{n=1}^\infty \mu_0(E_n)$.

For each $N \geq 1$, let $F_N := E \backslash \bigcup_{n=1}^N E_N$. Then the F_N lie in $\mathcal{B}_0$, are decreasing, and are such that $\bigcap_{N=1}^\infty F_N = \emptyset$. By finite additivity (and the finiteness of μ_0), we see that it suffices to show that $\lim_{N \to \infty} \mu_0(F_N) = 0$.

Suppose this is not the case, then there exists $\varepsilon > 0$ such that $\mu_0(F_N) > \varepsilon$ for all N. As each F_N lies in $\mathcal{B}_0$, we have $F_N = \pi_{B_N}^{-1}(G_N)$ for some finite sets $B_N \subset A$ and some $\mathcal{B}_{B_N}$-measurable sets G_N. By enlarging each B_N as necessary we may assume that the B_N are increasing in N. The decreasing nature of the F_N then gives the inclusions

$$G_{N+1} \subset \pi_{B_N \leftarrow B_{N+1}}^{-1}(G_N).$$

By inner regularity, one can find a compact subset K_N of each G_N such that

$$\mu_{B_N}(K_N) \geq \mu_{B_N}(G_N) - \varepsilon/2^{N+1}.$$

If we then set

$$K'_N := \bigcup_{N'=1}^{N} \pi_{B_{N'} \leftarrow B_N}^{-1}(K_N)$$

then we see that each K'_N is compact and

$$\mu_{B_N}(K'_N) \geq \mu_{B_N}(G_N) - \varepsilon/2^N \geq \varepsilon - \varepsilon/2^N.$$

In particular, the sets K'_N are non-empty. By construction, we also have the inclusions

$$K'_{N+1} \subset \pi_{B_N \leftarrow B_{N+1}}^{-1}(K'_N)$$

and thus the sets $H_N := \pi_{B_N}^{-1}(K'_N)$ are decreasing in N. On the other hand, since these sets are contained in F_N, we have $\bigcap_{N=1}^\infty H_N = \emptyset$.

By the axiom of choice, we can select an element $x_N \in H_N$ from H_N for each N. Observe that for any N_0, that $\pi_{B_{N_0}}(x_N)$ will lie in the compact set K'_{N_0} whenever $N \geq N_0$. Applying the Heine-Borel theorem repeatedly, we may thus find a subsequence $x_{N_{1,m}}$ of the x_N for $m = 1, 2, \ldots$ such that $\pi_{B_1}(x_{N_{1,m}})$ converges; then we can find a further subsequence $x_{N_{2,m}}$ of that subsequence such that $\pi_{B_2}(x_{N_{2,m}})$, and more generally obtain nested subsequences $x_{N_{j,m}}$ for $m = 1, 2, \ldots$ and $j = 1, 2, \ldots$ such that for each $j = 1, 2, \ldots$, the sequence $m \mapsto \pi_{B_j}(x_{N_{j,m}})$ converges.

Now we use the diagonalisation trick. Consier the sequence $x_{N_{m,m}} =: (y_{m,\alpha})_{\alpha \in A}$ for $m = 1, 2, \ldots$. By construction, we see that for each j, $\pi_{B_j}(x_{N_{m,m}})$ converges to a limit as $m \to \infty$. This implies that for each $\alpha \in \bigcup_{j=1}^\infty B_j$, $y_{m,\alpha}$ converges to a limit y_α as $m \to \infty$. As K'_j is closed, we see that $(y_\alpha)_{\alpha \in B_j} \in K'_j$ for each j. If we then extend y_α

arbitrarily from $\alpha \in \bigcup_{j=1}^\infty B_j$ to $\alpha \in A$, then the point $y := (y_\alpha)_{\alpha \in A}$ lies in H_j for each j. But this contradicts the fact that $\bigcap_{N=1}^\infty H_N = \emptyset$. This contradiction completes the proof that μ_0 is a pre-measure.

If we then let μ be the Hahn-Kolmogorov extension of μ_0, one easily verifies that μ obeys all the required properties, and the uniqueness follows from Exercise 1.7.7. □

The Kolmogorov extension theorem is a fundamental tool in the foundations of probability theory, as it allows one to construct a probability space to hold a variety of random processes $(X_t)_{t \in T}$, both in the discrete case (when the set of times T is something like the integers $\mathbf{Z}$) and in the continuous case (when the set of times T is something like $\mathbf{R}$). In particular, it can be used to rigorously construct a process for *Brownian motion*, known as the *Wiener process*. We will however not focus on this topic, which can be found in many graduate probability texts. But we will give one common special case of the Kolmogorov extension theorem, which is to construct product probability measures:

Theorem 2.4.4 (Existence of product measures)**.** *Let A be an arbitrary set. For each $\alpha \in A$, let $(X_\alpha, \mathcal{B}_\alpha, \mu_\alpha)$ be a probability space in which X_α is a locally compact, σ-compact metric space, with $\mathcal{B}_\alpha$ being its Borel σ-algebra (i.e. the σ-algebra generated by the open sets). Then there exists a unique probability measure $\mu_A = \prod_{\alpha \in A} \mu_\alpha$ on $(X_A, \mathcal{B}_A) := (\prod_{\alpha \in A} X_\alpha, \prod_{\alpha \in A} \mathcal{B}_\alpha)$ with the property that*

$$\mu_A(\prod_{\alpha \in A} E_\alpha) = \prod_{\alpha \in A} \mu_\alpha(E_\alpha)$$

whenever $E_\alpha \in \mathcal{B}_\alpha$ for each $\alpha \in A$, and one has $E_\alpha = X_\alpha$ for all but finitely many of the α.

Proof. We apply the Kolmogorov extension theorem to the finite product measures $\mu_B := \prod_{\alpha \in B} \mu_\alpha$ for finite $B \subset A$, which can be constructed using the machinery in Section 1.7.4. These are Borel probability measures on a locally compact, σ-compact space and are thus inner regular (see §1.10 of *An epsilon of room, Vol. I*). The compatibility condition (2.6) can be verified from the uniqueness properties of finite product measures. □

Remark 2.4.5. This result can also be obtained from the *Riesz representation theorem*, which is covered in §1.10 of *An epsilon of room, Vol. I.*

Example 2.4.6 (Bernoulli cube)**.** Let $A := \mathbf{N}$, and for each $\alpha \in A$, let $(X_\alpha, \mathcal{B}_\alpha, \mu_\alpha)$ be the two-element set $X_\alpha = \{0,1\}$ with the discrete metric (and thus discrete σ-algebra) and the uniform probability measure μ_α. Then Theorem 2.4.4 gives a probability measure μ on the infinite discrete cube $X_A := \{0,1\}^{\mathbf{N}}$, known as the (uniform) *Bernoulli measure* on this cube. The coordinate functions $\pi_\alpha : X_A \to \{0,1\}$ can then be interpreted as a countable sequence of random variables taking values in $\{0,1\}$. From the properties of product measure one can easily check that these random variables are uniformly distributed on $\{0,1\}$ and are jointly independent[2]. Informally, Bernoulli measure allows one to model an infinite number of "coin flips". One can replace the natural numbers here by any other index set, and have a similar construction.

Example 2.4.7 (Continuous cube)**.** We repeat the previous example, but replace $\{0,1\}$ with the unit interval $[0,1]$ (with the usual metric, the Borel σ-algebra, and the uniform probability measure). This gives a probability measure on the infinite continuous cube $[0,1]^{\mathbf{N}}$, and the coordinate functions $\pi_\alpha : X_A \to [0,1]$ can now be interpreted as jointly independent random variables, each having the uniform distribution on $[0,1]$.

Example 2.4.8 (Independent gaussians)**.** We repeat the previous example, but now replace $[0,1]$ with $\mathbf{R}$ (with the usual metric, and the Borel σ-algebra), and the normal probability distribution $d\mu_\alpha = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$ (thus $\mu_\alpha(E) = \int_E \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$ for every Borel set E). This gives a probability space that supports a countable sequence of jointly independent gaussian random variables π_α.

[2] A family of random variables $(Y_\alpha)_{\alpha \in A}$ is said to be *jointly independent* if one has $\mathbf{P}(\bigwedge_{\alpha \in B} Y_\alpha \in E_\alpha) = \prod_{\alpha \in B} \mathbf{P}(Y_\alpha \in E_\alpha)$ for every finite subset B of A and every collection E_α of measurable sets in the range of Y_α.

Bibliography

[BrLi1983] H. Brezis, E. Lieb, *A relation between pointwise convergence of functions and convergence of functional*, Proc. Amer. Math. Soc. **88** (1983), 486–490.

[De1901] M. Dehn, *Über den Rauminhalt*, Mathematische Annalen **55** (1901), no. 3, pages 465-478.

[deG1981] M. de Guzmán, Real variable methods in Fourier analysis. North-Holland Mathematics Studies, 46. Notas de Matemática , 75. North-Holland Publishing Co., Amsterdam-New York, 1981.

[Go1938] K. Gödel, *Consistency of the axiom of choice and of the generalized continuum-hypothesis with the axioms of set theory*, Proc. Nat. Acad. Sci. **24** (1938), 556–557.

[Me2003] A. Melas, *The best constant for the centered Hardy-Littlewood maximal inequality*, Ann. of Math. **157** (2003), no. 2, 647-688.

[So1970] R. Solovay, *A model of set-theory in which every set of reals is Lebesgue measurable*, Annals of Mathematics **92** (1970), 1-56.

[StSk2005] E. Stein, R. Shakarchi, Real analysis. Measure theory, integration, and Hilbert spaces. Princeton Lectures in Analysis, III. Princeton University Press, Princeton, NJ, 2005.

[StSt1983] E. Stein, J.-O. Strömberg, *Behavior of maximal functions in R^n for large n*, Ark. Mat. **21** (1983), no. 2, 259-269.

[Ta2008] T. Tao, Structure and Randomness: pages from year one of a mathematical blog, American Mathematical Society, Providence RI, 2008.

[Ta2009] T. Tao, Poincaré's Legacies: pages from year two of a mathematical blog, Vol. I, American Mathematical Society, Providence RI, 2009.

[Ta2010] T. Tao, An epsilon of room, Vol. I, American Mathematical Society, Providence RI, 2010.

[Vi1908] G. Vitali, *Sui gruppi di punti e sulle funzioni di variabili reali*, Atti dell'Accademia delle Scienze di Torino **43** (1908), 75-92.

Index