Measure Theory

Jacob Denson

December 27, 2017

Contents

Chapter 1

The Lebesgue Measure

Area is one of the most primitive quantities in geometry. Every elementary school student knows that the area of a circle is πr^2, and that the area of a rectangle is equal to the product of the lengths of its two side lengths. But given a general shape in the plane, it suddenly becomes very difficult to reason precisely about a shape's area. In the work of the ancient Greeks, especially Archimedes, we find two methods of finding the area of more complex families of shapes:

- If a shape can be cut up into finitely many components, and then rearranged into the form of a different shape by a series of rigid motions, then the new shape has the same area as the old.

- We can obtain upper bounds on the area of the shape by enveloping the shape by another shape with an already known area, and lower bounds by finding the area of shapes enclosed by the shape.

Naive applications of these techniques sufficed for much of the classical era of mathematics. Nonetheless, with the advent of Cartesian coordinates and the subsequent modern introduction of set theory, shapes in the Euclidean plane are identified with subsets of $\mathbf{R}^2$. Set theory enables us to obtain rather strange subsets of $\mathbf{R}^2$, for which the methods of Archimedes begins to break down. For instance, what is the 'area' of the set of all points in the unit circle whose coordinates are rational numbers – every circle, or finite union of rectangles containing these points has area π, yet the set seems to have too many 'holes' to have this area. More generally, it isn't clear how to define the lengths of subsets of the real line $\mathbf{R}$, or volumes

of general subsets of $\mathbf{R}^3$. These are all *measures* of size in their relative dimensions, and we call the general study of these objects *measure theory*. In line with this unification, we will let $|S|$, or $\mu(S)$, denote the length, area, volume, and higher dimensional measures of a given shape S in $\mathbf{R}^d$, regardless of the dimension.

Remark. *Even simple shapes may require approximation techniques rather than just employing the first. The Bolyai-Gerwien theorem says that if P and Q are polytopes with the same area in the plane, then one can cut up P into finitely many polytopes, which can be individually rotated and translated to form Q. Dehn proved that this theorem fails in dimensions three or higher, which means we need approximation techniques to measure area.*

These set theoretic problems are essentially irrelevant in the classical theory of analysis before the 20th century, where functions and sets are specified by analytical formulas in terms of a few basic operations, which are almost certainly continuous, except at a mild set of points, and are essentially guaranteed to be infinitely differentiable. Nonetheless, near the end of the 19th century the mathematical problems of the era prompted the need to measure the volume of more pathological sets.

- Given a continuous curve in the plane, one often wants to measure it's length. In 1890, Peano constructed a continuous curve mapping the unit interval $[0,1]$ surjectively onto the unit rectangle $[0,1]^2$. It is suspect as to whether we can measure the length of this curve, because it is constructed as the limit of curves whose lengths tend to infinity. Thus a more particular analysis determining which curves have a well defined length was required, and this required the theory of measure. Curves without a well defined length are not purely artificial, for if we take a typical continuous curve obtained by following the random motion of a dust particle in air, then we find the path does not have a reasonable length, because there are too many 'bends' in the path. Similarly, very recently difficulties were found when measuring the length of the perimter of countries, which are too 'jagged' for us to obtain a length of a country.

- In 1924, Stefan Banach and Alfred Tarski showed that one can decompose a sphere into a finite number of components, which by a sequence of rigid motions can be rearranged into two copies of the

3

original sphere! We seem to have produced two equal things from one thing – a feat not far from biblical miracle (though we have to use oranges, not loaves and fishes). In duplicating the sphere, Banach and Tarski showed that even the old methods of are geometry do not stand up to the techniques of modern mathematics, and a reanalysis of the entire field is required.

- Fourier analysis cannot be underestimated in the degree to which it prompted the development of measure theory. In the analysis of convergence properties of Fourier series and Fourier transforms, one often works with a sequence of continuous functions $f_1, f_2, \ldots$, and one wants to analyze the pointwise limit $f(x)$, defined to be $\lim f_n(x)$. If the convergence of the f_n is uniform, then f itself is uniform, and one can easily determine that

$$\int_0^1 f(x) \, dx = \lim n \to \infty \int_0^1 f_n(x) \, dx$$

On the other hand, one can choose a monotone decreasing sequence f_i whose pointwise limit f is not even Riemann integrable. Nonetheless, the right hand limit certainly exists, as a bounded, monotone sequence of numbers, so a question arises as to how we may interpret the integral on the left so that the equation holds in a more general sense. This sense is provided, in an essentially complete manner, by measure theory.

One key idea of measure theory is that the old methods of geometry continue to work, provided that we only concentrate on certain 'nice' subsets of space which obey our intuitions about size. Stefan Banach and Alfred Tarski engineered one of the first partitions of space which do not obey geometric intuition. We call these **unmeasurable sets**, because trying to measure their size causes problems. The main theorems of measure theory only work when we work with **measurable sets**. Indeed, we can reinterpret the methods of ancient geometry into two principles of measure theory:

- If a *measurable* set S has a decomposition into disjoint, *measurable* subsets $S_1, \ldots, S_n$, then $|S| = \sum |S_i|$, and if S is transformed into T by a rigid motion, then $|S| = |T|$.

- If a set S is *measurable*, then there is a decreasing sequence of shapes $S_1, S_2, \ldots$, each containing S, and each decomposable into disjoint intervals/squares/boxes, which we can measure by taking the sum of the volumes of the boxes, and $|S_i| \to |S|$.

We will begin by extending the notion of sets to fairly general subsets of $\mathbf{R}^d$. We do this not only because this is classically how measure theory was introduced, but also because it brings to light the many intricate parts of the theory which we consider when we build measures on more general 'measure spaces', which is required for understanding the modern theory of probability and certain other 'size' metrics which occur in analytical mathematics, most importantly, the modern foundations of probability theory.

1.1 Measuring Elementary Sets

Let's begin by using basic ideas of Euclidean geometry to find a basic class of sets, which are suitably elementary that the paradoxes we have seen above cannot occur. It is unarguable that the length of an interval with start point a and end point b, whether closed, open, or half open, is $b - a$. We know from elementary geometry that the area of a rectangle is the product of the length of the intervals which define it, and if we are suitably relaxed, we can take this as the definition of the volume of a rectangle in higher dimensions. That is, if $I = I_1 \times \cdots \times I_d$ is a box in $\mathbf{R}^d$, in the sense that it is the product of intervals I_i between a_i and b_i, then we define

$$|I| = (b_1 - a_1) \ldots (b_d - a_d) = |I_1| \ldots |I_d|$$

The easiest sets to measure are those covered by finitely many *disjoint* boxes, and we will show this family has a well defined theory of area.

Lemma 1.1. *If we write a box R as the disjoint union of finitely many boxes R_i, then $|R| = \sum |R_i|$.*

Proof. We proceed by means of a grid decomposition, which refines the objects in the problem in such a way that the proof of the proposition is trivial. Write

$$R = I_1 \times \cdots \times I_d$$

Suppose first that the rectangular decomposition forms a grid, in the sense that we can index the decomposition as $R_{i_1 \ldots i_d}$, where

$$R_{i_1 \ldots i_d} = I^1_{i_1} \times \cdots \times I^d_{i_d}$$

and the endpoint of I^k_i is the startpoint of I^k_{i+1}. We can then just apply distributivity to conclude that

$$\sum_{i_1, \ldots, i_d} |R_{i_1 \ldots i_d}| = \sum_{i_1, \ldots, i_d} |I^1_{i_1}| \ldots |I^d_{i_d}| = \prod_{k=1}^d \left(\sum_j |I^k_j| \right)$$

and all that remains is to show $\sum_j |I^k_j| = |I_k|$. This follows because the sum is telescopic, with the highest indexed interval's endpoint the endpoint of I_k, and the lowest indexed intervals startpoint the startpoint of I_k. In general, it suffices to break a general decomposition into a further decomposition forming a grid, in such a way that the sum of the boxes in the first decomposition is equal to the sum of the boxes in the second. This is proven by slicing all boxes up by the startpoints and endpoints of the starting decomposition, and a telescopic sum argument in each dimension shows the area is maintained. $\qquad \square$

A similar grid decomposition like argument proves the following very similar statements.

Lemma 1.2. *If a family of boxes $R_1, \ldots, R_n$ covers R (perhaps not disjointly), then $|R| \leqslant \sum |R_i|$.*

Lemma 1.3. *If $R_1, \ldots, R_n$ and $S_1, \ldots, S_m$ are two families of disjoint boxes with $\bigcup R_i = \bigcup S_i$, then $\sum |R_i| = \sum |S_i|$.*

Remark. *These theorems can alternatively be proven by combinatorial arguments, using the fact that the disjoint sum properties of area are obvious in the case of measuring the cardinality of finite sets. First, one shows that*

$$|I| = \lim_{N \to \infty} \frac{|\mathbf{Z}/N \cap I|}{N}$$

where the 'measure' on the right hand size is just the cardinality of the finite set. From this, it is easy to argue that for any box R,

$$|R| = \lim_{N \to \infty} \frac{|\mathbf{Z}^d/N \cap R|}{N^d}$$

But now if we write $R = \bigcup R_i$ as the union of disjoint boxes, then

$$|R| = \lim_{N \to \infty} \frac{|\mathbf{Z}^n/N \cap R|}{N^n} = \sum_i \lim_{N \to \infty} \frac{|\mathbf{Z}^n/N \cap R_i|}{N^n} = \sum_i |R_i|$$

and this proves the theorem, not only in the finite case but even when the decomposition is infinite. One might be tempted to define the measure of an arbitrary subset of $\mathbf{R}^d$ by the formula

$$|E| = \lim_{N \to \infty} \frac{|\mathbf{Z}^n/N \cap E|}{N^n}$$

However, this definition runs into problems for more general sets than intervals. One can find sets where this limit doesn't exist, and even if the limit does exist, we might not even have translation invariance; $\mathbf{Q} \cap [0,1]$ has length 1 with respect to this metric, whereas $\mathbf{Q} \cap [0,1] + \sqrt{2}$ has length 0. For the class of Jordan measurable sets, which we will soon discuss, this definition works. A more suitable way to obtain a continuous uniform measure from some discrete quantity is by the theory of Monte Carlo distributions probability, but since the analysis of this method already involves deep results in measure theory, we won't cover it here.

The above lemmas guarantee that if $E \subset \mathbf{R}^d$ is the disjoint union of boxes $R_1, \ldots, R_n$, then the definition

$$|E| = \sum |R_i|$$

is well defined, and fairly unarguably *the* choice for the volume of this family of sets. We call such a set E an **elementary set**. We will use the unarguable measure of elementary sets to approximate the measure of arbitrary subsets of the plane.

The measure of size on elementary sets have the following fundamental properties, which are *very important* going forward. One can summarize the development of measure theory as trying to measure the most general sets possible, while still preserving the properties we describe below. First, because of the corresponding property for the measure of boxes, if $E = E_1 \cup \cdots \cup E_N$ is the union of disjoint elementary sets, then

$$|E| = |E_1| + \cdots + |E_N|$$

This is the property of *finite additivity*. If the union is not disjoint, then we have *subadditivity*,

$$|E| \leqslant |E_1| + \cdots + |E_N|$$

(In the next paragraph we will show that E is an elementary set if $E_1, \ldots, E_N$ are). The measure is also *translation invariant*, in the sense that $|E+x| = |E|$ for any elementary set E. If $E \subset F$, then $F - E$ is an elementary set, and so

$$|F| = |E| + |F - E| \geqslant |E|$$

so the measure is *monotone*. We also have the *inclusion-exclusion principle*

$$|E \cup F| = |E| + |F| - |E \cap F|$$

which holds for any not necessarily disjoint E and F, because

$$|E| = |E - F| + |F| = |E - (E \cap F)| + |F| = |E| - |E \cap F| + |F|$$

Note that this equation holds with any 'measure' on any family of sets which has finite additivity. These properties uniquely define the measure of a set up to a scalar factor. Since all the properties are intuitive to us, this tells us that our construction is going in the right direction!

In math, we often transform sets by taking set theoretic operations, like the union and intersection. The fact that a family of sets with a well defined area is closed under these operation hints at the feasibility for this family to be used in mathematic developments. Our eventual description of the 'Lebesgue measurable sets' will be a family which are closed under essentially every set theoretic operation. In fact, without applying one arcane principle of mathematics, the axiom of choice, it is impossible to begin to find a set which we *cannot* measure the area of, so this family is almost perfect. We refer to these techniques as measuring the *algebraic power* of the family of sets.

Theorem 1.4. *Let R and S denote boxes, and E and F denote elementary sets.*

(a) *If E is the finite union of boxes, then E is the finite union of disjoint boxes. Thus the union of two elementary sets is also an elementary set.*

(b) *The intersection of two boxes is a box, and the intersection of two elementary sets is an elementary set.*

(c) $R - S$ is an elementary set, and so the difference of two elementary sets is an elementary set. The family of elementary sets is not closed under the complement operation, but this proposition says the family is 'almost' closed under this operation.

(d) $x + E$ is an elementary set, so the family of elementary sets is translation invariant.

(e) The symmetric difference $E \triangle F = (E - F) \cup (F - E)$ is an elementary set.

(f) If $E \subset \mathbf{R}^d$ and $F \subset \mathbf{R}^s$, then $E \times F$ is an elementary subset of $\mathbf{R}^{d+s}$, and $|E \times F| = |E||F|$.

Proof. Proposition (a) follows from a simple grid decomposition argument, left as an exercise. To prove (b), note first that if I is an interval with start point a and endpoint b, and J is an interval with start point c and endpoint d, then $I \cap J$ is either empty, or an interval with startpoint $\max(a,c)$, and endpoint $\min(b,d)$. But then if $R = I_1 \times \cdots \times I_n$ and $S = J_1 \times \cdots \times J_n$, then

$$R \cap S = (I_1 \cap J_1) \times \cdots \times (I_n \cap J_n)$$

which is a box. If $E = \bigcup R_i$, and $F = \bigcup S_i$ are two elementary sets, then $E \cap F = \bigcup (R_i \cap S_j)$ is also an elementary set, since each $R_i \cap S_j$ is a box, which are disjoint over all the indices. To prove (c), we leave the messy details of breaking $R - S$ into boxes to the reader. Once this is done, it is easy to see the difference of two elementary sets is elementary, because if $E = \bigcup R_i$ and $F = \bigcup S_i$, then

$$E - F = \bigcup \left(R_i \cap \bigcap S_i^c \right)$$

which is the union of elementary sets, hence $E - F$ is elementary. (d) is trivial, and is left to the reader. If $E = \bigcup E_i$ and $F = \bigcup S_i$, then $E \times F = \bigcup E_i \times S_j$, and each $E_i \times S_j$ is a disjoint box, proving (e), once we note that $|R \times S| = |R||S|$ for any two boxes R and S. $\square$

A family of sets over some set X is an **algebra** if it is closed under the union and complement operations, and contains $\varnothing$. The reason for the name is that, if we view the operation $\triangle$ as a commutative addition operation on the family, and the intersection as a commutative multiplication operation, then this turns the family into an algebra over $\mathbf{F}_2$, with $\varnothing$ as

the zero element. If we consider a box, and consider the family of elementary sets contained within the box, then our arguments above result in showing this family forms an algebra. Nonetheless, the family of all elementary subsets in $\mathbf{R}^d$ is not an algebra, and this is one of the reasons for us to enlarge the family of sets we can measure.

1.2 Jordan Measurable Sets

In the last section, we constructed a consistant measure on the family of elementary sets. However, this family is obviously limited. We cannot even use this quantity to measure the area of a circle, or the volume of a sphere. However, we have really only applied the first ancient technique of measuring area, forming disjoint unions of simple shapes. It was the development of the basic tools of analysis, and the invention of Newton and Leibnitz's calculus which made the most headway into these techniques, using the Riemann integral. Recall that if f is a bounded function on a box R, then we lower and upper bound the integral by partitions P of the box into subboxes, setting

$$I_*(P, f) = \sum_{S \in P} |S| \inf_{x \in S} f(x) \quad I^*(P, f) = \sum_{S \in P} \sup_{x \in S} f(x)$$

and we say f is Riemann integrable if, as P becomes suitably fine, $I_*(P, f)$ and $I^*(P, f)$ converge to a common value, which we call the Riemann integral of f, and denote by $\int_R f(x)\, dx$. Applying this technique to our scenario, if E is a bounded set, contained in some box R, then the area of E can be defined by the formula

$$|E| = \int_R \chi_E(x)\, dx$$

provided the integral on the right exists, in which case we say E is **Jordan measurable**, and we call $|E|$ the Jordan measure of the set. No unbounded set is Jordan measurable.

χ_E is only zero-one valued, and so the full definition of the Riemann integral is overkill to define the Jordan measure of E. Given some partition P of R into boxes, we can write

$$I_*(P, \chi_E) = \sum_{\substack{S \in P \\ S \subset E}} |S| \qquad I^*(P, \chi_E) = \sum_{\substack{S \in P \\ E \cap S \neq \varnothing}} |S|$$

10

From this formula, we see that the Jordan measure is really just obtained by approximating sets from above and below by elementary sets. Thus we can define the inner and outer Jordan measures

$$J_*(E) = \sup\{|F| : F \subset E, F \text{ elementary}\}$$

$$J^*(E) = \inf\{|F| : E \subset F, F \text{ elementary}\}$$

a bounded set E is Jordan measurable precisely when $J_*(E) = J^*(E)$, and this common value is then the Jordan measure of E. This is equivalent to the existence of elementary sets $E_1 \subset E \subset E_2$ with $|E_2 - E_1| < \varepsilon$, for any $\varepsilon > 0$. It is easy to see that J_* and J^* are both monotone, and therefore the Jordan measure is monotone when it exists. Furthermore, J^* is subadditive. For any two sets E and F, regardless of whether they are Jordan measurable or not, we conclude $J^*(E \cup F) \leqslant J^*(E) + J^*(F)$. For disjoint Jordan measurable sets, J^* is additive, but J^* is *not* additive for general disjoint sets. This reason is that Jordan measure depends very much on the topology of the underlying space, and is therefore unable to distinguish between sets which are 'too entangled' with one another.

Example. *Let $E_1 = \mathbf{Q} \cap [0,1]$. Then any elementary set containing E_1 also contains $[0,1]$, and so $J^*(E_1) = 1$. Similarily, if $E_2 = (\mathbf{Q} + \sqrt{2}) \cap [0,1]$, then $J^*(E_2) = 1$. E_1 and E_2 are disjoint, but their union is contained in $[0,1]$, and it is again easy to see that $J^*(E_1 \cup E_2) = 1$.*

The Jordan measurable sets are precisely those which are efficiently covered by elementary sets. Noe that $E = \mathbf{Q} \cap [0,1]$ is not efficiently covered, since $J^*(E \triangle F) = 1$ for every elementary set F.

Theorem 1.5. *A bounded set E is Jordan measurable if and only if for every $\varepsilon > 0$, there is an elementary set F with $J^*(E \triangle F) \leqslant \varepsilon$.*

Proof. Suppose E is Jordan measurable. Then there are elementary sets E_1 and E_2, with $E_1 \subset E \subset E_2$, with $|E_2 - E_1| \leqslant \varepsilon$. But $E_2 - E \subset E_2 - E_1$, and so the monotonicity of J^* allows us to conclude that

$$J^*(E \triangle E_2) = J^*(E_2 - E) \leqslant |E_2 - E_1| \leqslant \varepsilon$$

Conversely, suppose E is a bounded set, and for any ε, we can find an elementary set F with $J^*(E \triangle F) \leqslant \varepsilon$. This means that we can find an elementary set L covering $E \triangle F$ with $|L| \leqslant 2\varepsilon$, and then $F - L \subset E \subset F \cup L$, and

$$|(F \cup L) - (F - L)| = |F \cup L| - |F - L| \leqslant |F| + 2\varepsilon - |F| \leqslant 2\varepsilon$$

11

taking ε to zero, we conclude E is Jordan measurable. □

Theorem 1.6. *If E and F are Jordan measurable subsets of $\mathbf{R}^d$, then*

(a) *$E \cup F$ is Jordan measurable.*

(b) *$E \cap F$ is Jordan measurable.*

(c) *$F - E$ is Jordan measurable. If $E \subset F$, then $|F - E| = |F| - |E|$.*

(d) *$E \triangle F$ is Jordan measurable.*

(e) *Jordan measure is finitely additive, and if T is an invertible linear transformation, then $|E| = |\det(T)||T^{-1}(E)|$. In particular, if $|\det(T)| = 1$ (if T is a rotation, for instance), then $|E| = |T^{-1}(E)|$.*

Proof. To prove (a), note that if there are elementary sets E_1 and F_1 with $J^*(E_1 \triangle E), J^*(F_1 \triangle F) \leqslant \varepsilon$, then

$$(E \cup F) \triangle (E_1 \cup F_1) \subset (E \triangle E_1) \cup (F \triangle F_1)$$

so $J^*((E \triangle E_1) \cup (F \triangle F_1)) \leqslant J^*(E \triangle E_1) + J^*(F \triangle F_1) \leqslant 2\varepsilon$. To prove (b), if $E_1 \subset E \subset E_2$ and $F_1 \subset F \subset F_2$ are given with $|E_2 - E_1|, |F_2 - F_1| \leqslant \varepsilon$, then

$$E_1 \cap F_1 \subset E \cap F \subset E_2 \cap F_2$$

and

$$(E_2 \cap F_2) - (E_1 \cap F_1) \subset (E_2 - E_1) \cup (F_2 - F_1)$$

so $|E_2 \cap F_2 - E_1 \cap F_1| \leqslant 2\varepsilon$. If $E \subset F$ then $\chi_{F-E} + \chi_E = \chi_F$, and the additivity of the Riemann integral gives that $F - E$ is Jordan measurable and $|F| = |E| + |F - E|$. In general, $F - E = F - (E \cap F)$, so $F - E$ is Jordan measurable in general. (d) is then obvious. Finite additivity can be obtained by a similar Riemann integration trick, from the fact that if E and F are disjoint, then $\chi_E + \chi_F = \chi_{E \cup F}$. The fact that the area of Jordan measurable sets are invariant follows from the change of variables formula from calculus, because if T is a linear transformation, and E is Jordan measurable, then

$$|E| = \int \chi_E(x)\,dx = \int \chi_E(Tx)|\det(T)|\,dx = \chi_{T^{-1}(E)}(x)|\det(T)|\,dx$$
$$= |\det(T)||T^{-1}(E)|$$

which proves the theorem. □

Thus Jordan measurability has all the properties we liked about the measure of elementary sets, like that the family of Jordan measurable sets form an algebra over every bounded box. Before we move on, we note a theorem of the Riemann integral which makes the identification of Jordan measurable sets elementary. We say a subset E has *measure zero* if, for every $\varepsilon > 0$, there exists a *countable* family of boxes $R_1, R_2, \ldots$ covering E with

$$\sum_{i=1}^{\infty} |R_i| < \varepsilon$$

The next theorem shows that determining if a set is Jordan measurable reduces to showing that the boundary of the set has measure zero.

Theorem 1.7. *A bounded function $f : R \to \mathbf{R}$ on a box R is Riemann integrable if and only if the set of discontinuity points of f has measure zero.*

Proof. Define the **oscillation** of f at a point x to be

$$\mathrm{osc}(f, x) = \lim_{r \to 0} \sup_{y \in B_r(x)} |f(x) - f(y)|$$

For each r, the function $\mathrm{osc}(f, x, r) = \sup_{y \in B_r(x)} |f(x) - f(y)|$ is a continuous function of x, so $\mathrm{osc}(f, x) = \inf_{r \to 0} \mathrm{osc}(f, x, r)$ is an *upper semicontinuous* function, in the sense that $\limsup_{y \to x} \mathrm{osc}(f, y) \leqslant \mathrm{osc}(f, x)$ for each point $x \in R$. It follows that for each $\varepsilon > 0$ the sets

$$A_\varepsilon = \{x : \mathrm{osc}(f, x) \geqslant \varepsilon\}$$

are closed and bounded, and therefore compact, sets. Any covering of A_ε has a finite subcovering, and since the discontinuities of f are a set of measure zero, there exist finitely many open boxes $R_1, \ldots, R_N$ covering A_ε with $|R_1| + \cdots + |R_N| \leqslant \varepsilon$. Now $K = (R_1 \cup \cdots \cup R_n)^c$ is compact, and every point $x \in K$ has $\mathrm{osc}(f, x) < \varepsilon$, so there is an open rectangle around x such that for any y in the set, $|f(y) - f(x)| \leqslant \varepsilon$. These rectangles cover K, so we can select finitely many $R_{N+1}, \ldots, R_M$ covering K. By dividing these rectangles suitably, these rectangles correspond to some partition P of R,

and we have

$$U(f,P) - L(f,P) = \sum_{k=1}^{M} |R_i| \left[\sup_{x \in R_i} f(x) - \inf_{x \in R_i} f(x) \right]$$

$$\leqslant \sum_{k=1}^{N} |R_i| 2\|f\|_\infty + \sum_{k=N+1}^{M} |R_i| \varepsilon$$

$$\leqslant (2\|f\|_\infty + |R|)\varepsilon$$

and by taking $\varepsilon \to 0$, we obtain suitably fine partitions showing that f is Riemann integrable. Conversely, note that if $E_1, E_2, \dots$ all have measure zero, then $\bigcup E_i$ has measure zero, because if we find covers U_{ij} of E_j with $\sum_i |U_{ij}| \leqslant \varepsilon/2^n$, then $\sum_{ij} |U_{ij}| \leqslant \varepsilon$ and covers $\bigcup E_i$. Thus, given a Riemann integrable function f, it suffices to prove that A_ε has measure zero for each $\varepsilon > 0$, for the set of discontinuities of f is $\bigcup A_{1/n}$. But if we find a partition P with $U(f,P) - L(f,P) < \varepsilon$. P corresponds to a sequence of boxes $R_1, \dots, R_N$ with

$$\sum |R_i| \left[\sup_{x \in R_i} f(x) - \inf_{x \in R_i} f(x) \right] \leqslant \varepsilon$$

But then using a Markov type inequality, we conclude that if S is the set of indices i with $\sup_{x \in R_i} f(x) - \inf_{x \in R_i} f(x) \geqslant \delta$, then

$$\sum_{i \in S} |R_i| \leqslant \varepsilon/\delta$$

But the R_i, for $i \in S$, are a cover of A_δ, so taking $\varepsilon \to 0$, we conclude that A_δ is a set of measure zero. $\qquad\square$

Example. *Let R be a closed box in $\mathbf{R}^d$, and $f : R \to \mathbf{R}$ a continuous function. Then the graph $\Gamma(f)$ of f in $\mathbf{R}^{n+1}$ is Jordan measurable, and has Jordan measure zero. This follows because, since R is compact, f is uniformly continuous, and therefore for any ε we can partition R into finitely many almost disjoint boxes $S_1, \dots, S_N$ covering R such that if two points x and y lie in the same box, then $|f(x) - f(y)| < \varepsilon$. But this implies that if we fix x_i in S_1, then the sets $S_i \times [x_i - \varepsilon, x_i + \varepsilon]$ cover the graph, and so*

$$J^*(\Gamma(f)) \leqslant \sum |S_i \times [x_i - \varepsilon, x_i + \varepsilon]| = 2\varepsilon \sum |S_i| = 2\varepsilon |R|$$

14

If we let $\varepsilon \to 0$, we obtain that $\Gamma(f)$ has upper measure zero, and thus lower measure zero. The set $X = \{(x,t) \in R \times \mathbf{R} : 0 \leqslant f(x) \leqslant t\}$ is also Jordan measurable. Given the same S_i, the sets $S_i \times [0, x_i - \varepsilon)$ are contained in X, and $S_i \times [0, x_i + \varepsilon]$ contain X, and the difference between these two sets is exactly the sets we used to show the measure of the boundary is zero, hence X is Jordan measurable because it can be approximated from above and below, and

$$|X| = \int f^+(x)\, dx$$

where $f^+(x) = \max(f(x), 0)$.

Example. *In the theory of elementary sets, we couldn't even measure the area of a triangle in the plane. Let P be a parallelogram in the plane. By translation invariance, we may assume one of the vertices lies at the origin. We can then describe the vertices by $\{0, x, y, x + y\}$. The linear transformation*

$$T = \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix}$$

maps the unit square to P, and therefore

$$|P| = |\det(T)| = |x_1 y_2 - x_2 y_1|$$

If T is a triangle defined by three vertices $\{0, x, y\}$, then one verifies that T is exactly half of the parallelogram $\{0, x, y, x + y\}$, and therefore

$$|T| = \frac{|x \times y|}{2}$$

Alternatively, if we assume that one of the edges of the triangle is vertical, then one can directly estimate the edges of the triangle by taking squares overapproximating and underapproximating the triangle, just as if we were directly calculating the Riemann integral by sums. We can then split a triangle up into two vertical triangles.

Example. *A closed convex polytope is an intersection of finitely many closed half spaces, of the form*

$$\mathbf{H}(y, a) = \{x : x \cdot y \leqslant a\}$$

and is compact if it is bounded. Every compact closed convex polytope is Jordan measurable. One way to see this directly is to note that by continuity,

$$\partial\mathbf{H}(y,a) = \{x : x \cdot y = a\}$$

and then

$$\partial\left(\bigcap\mathbf{H}(y,a)\right) \subset \bigcup\partial\mathbf{H}(y,a)$$

and the right hand side is a finite union of planes, which are a set of measure zero. Alternatively, if a compact polytope P is non-empty, pick some $x \in P$. If $P = \bigcap\mathbf{H}(y_n,a_n)$, then for each n, consider the convex hull P_n obtained from connecting the convex set $S_n = P \cap \partial\mathbf{H}(y_n,a_n)$ to the point x. Then $P_n \subset P$, and in fact, $P = \bigcup P_n$, because if $y \in P$ is any point, then the line generated from x to y must eventually hit the boundary of some half plane, because the set is compact, and then this point lies in some $P \cap \partial\mathbf{H}(y_n,a_n)$. It suffices to show that P_n is Jordan measurable. By a rotation and translation, we may assume that $y_n = e_1$ and $a_n = 0$, so S_n lies on the half plane $\{x : x_1 = 0\}$. Consider the barycentric coordinate change

$$g(\lambda,\gamma) = \gamma(\lambda,0) + (1-\gamma)x$$

for $\gamma \in \mathbf{R}$, $\lambda \in \mathbf{R}^{n-1}$. Then

$$Dg(\lambda,\gamma) = \begin{pmatrix} \gamma I_{n-1} & 0 \\ \lambda - (x_1,\ldots,x_{n-1}) & -x_n \end{pmatrix}$$

where we assumed by induction that since S_n is a compact, convex polytope, hence it was Jordan measurable.

Example. *If B_r is a ball of radius r, then $|B_r| = r^n|B_1|$. This follows because the linear transformation $T(x) = rx$ has $|\det T| = r^n$ and $T^{-1}(B_r) = B_1$. If $x \in B_1$, then $|x_i| \leqslant |x| \leqslant 1$, so B_1 is contained in the square $[-1,1]^d$. Similarily, if $|x_i| \leqslant 1/\sqrt{d}$ for each i, then $|x| \leqslant 1$, so $[-1/\sqrt{d}, 1/\sqrt{d}] \subset B_1$. This yields the bounds*

$$\left(\frac{2}{\sqrt{d}}\right)^d \leqslant |B_1| \leqslant 2^d$$

These bounds are pretty crude, however, and are on the order of $2\pi^{d/2}/d\Gamma(d/2)$, which at least for even n is equal to $2\pi^{d/2}/d(d/2)!$.

The Jordan measure is too connected to the topology of $\mathbf{R}^d$, and thus is unable to distinguish untangled sets, or measure the area of sets with 'too many holes'. In particular, one can easily verify that for any set E, $J^*(E) = J^*(\overline{E})$, and $J_*(E) = J_*(E^\circ)$, and this be easily used to show $E = \mathbf{Q} \cap [0,1]$ is not Jordan measurable. Another consequence is that even if a sequence $E_1, E_2, \ldots$ of sets are all Jordan measurable, their union $\bigcup E_i$ need not be (Let $E = \{x_1, x_2, \ldots\}$, and then let $E_i = \{x_i\}$). This has consequences in Riemann integrability, where the pointwise limit of Riemann integrable functions may not be Riemann integrable. The Lebesgue measurable sets do not have this problem, and the space of Lebesgue integrable functions is closed under pointwise convergence.

1.3   Lebesgue Measure

To obtain a family of sets with a well defined measure theory which satisfies countable additivity, we must tinker with the definition of Jordan measure. Recall that to obtain Jordan measure, we took outer and inner estimates of arbitrary sets by covers of *finitely many* rectangles. We obtain Lebesgue integrability when we replace the finiteness with countably many rectangles. Analogous to the Jordan exterior measure J^*, we consider the values

$$\mu^*(E) = \inf\left\{ \sum_{k=1}^{\infty} |R_i| : E \subset \bigcup R_i \right\}$$

So the limitations of covers by finitely many rectangles are removed by allowing covers by countably many rectangles, which should provide a much better upper bound on the volume of E.

Example. *Consider the set $E = \mathbf{Q} \cap [0,1]$, for which $J^*(E) = 1$, since elementary sets could not 'penetrate' deep enough to break the density of E in $[0,1]$. If we now consider an enumeration $E = \{x_1, x_2, \ldots\}$, and for each n, consider a box R_n containing x_n with volume $\varepsilon/2^n$, then we obtain a cover $\{R_n\}$ of E which shows that*

$$\mu^*(E) \leqslant \sum_{n=1}^{\infty} |R_n| = \varepsilon \sum_{n=1}^{\infty} 2^{-n} = \varepsilon$$

Letting $\varepsilon \to 0$, we conclude $\mu^*(E) = 0$. More generally, if F is any countable set, then $\mu^*(F) = 0$. Note this holds even if F is unbounded.

The next computation was first introduced by Borel, who invented the theorem we now call the Heine-Borel theorem, and the modern definition of compactness, to show that $\mu^*(R) = |R|$ for any box R.

Example. *Let R be a rectangle in $\mathbf{R}^d$. It is obvious that $\mu^*(R) \leqslant J^*(R) = |R|$. On the other hand, if $S_1, S_2, \ldots$ is a cover of R, then we may consider a slightly larger family of open rectangles S_n' with $|S_n'| = |S_n| + \varepsilon 2^{-n}$, and then since R is compact, we may replace S_n' with a finite subcover $S_{i_1}, \ldots, S_{i_N}$, so*

$$|R| = J^*(R) \leqslant \sum_{n=1}^{N} |S_{i_n}| \leqslant \sum_{n=1}^{\infty} |S_n'| \leqslant \sum_{n=1}^{\infty} |S_n| + \varepsilon$$

Since S_n was arbitrary, we conclude that $|R| \leqslant \mu^*(R) + \varepsilon$, and we can then let $\varepsilon \to 0$ to conclude $|R| = \mu^*(R)$.

Remark. *Note that as a consequence of this theorem, we conclude that $\mathbf{R}^d$ is uncountable for each n, because, if $\mathbf{R}^d$ was countable, every subset E would be countable, and therefore we would conclude $\mu^*(E) = 0$. Thus the last example must use a fact about $\mathbf{R}^d$ which distinguishes it from other countable sets containing 'box like sets', such as $\mathbf{Q}$. This fact was provided by the Heine-Borel theorem, which exploits the completeness of Euclidean space.*

The next theorem, and the following, which is it's generalization in higher dimensions, shows that $\mu^*(E)$ is essentially obtained by approximating E by an open set. It is easy to show that μ^* is monotone, and finitely subadditive. With only slightly more work, we may show that μ^* is countably subadditive, in the sense that if $E \subset E_1 \cup E_2 \cup \ldots$, then

$$\mu^*(E) \leqslant \sum \mu^*(E_n)$$

Which is obtained by applying covers of E_n to the sets E.

Theorem 1.8. *Every open subset of $\mathbf{R}$ can be uniquely written as the countable union of disjoint open intervals.*

Proof. If $x \in U$, consider the sets

$$A = \{y \in U : (y, x] \in U\} \quad B = \{y \in U : [x, y) \in U\}$$

Then if $a \in \inf A$ and $b = \sup B$, then $(a, b) \in U$. Because U is open, $a, b \notin U$, and therefore no open interval which is contained in U contains (a, b), so

the interval is maximal. This argument shows that if we let $\mathcal{U}$ denote the family of maximal intervals contained in U, then $\mathcal{U}$ covers U. If I and J are maximal open intervals, then I must be disjoint from J, and therefore U is the disjoint union of the intervals in $\mathcal{U}$. Since each interval contains a rational number, and distinct maximal intervals are disjoint, we can index maximal intervals by rational numbers, and so the set $\mathcal{U}$ is countable. This proves the theorem. $\qquad\square$

Given an open set $U \subset \mathbf{R}$, we can uniquely decompose U as $\bigcup I_i$, and then define the length

$$|U| = \sum |I_i|$$

Then if U and V are disjoint open sets, then $|U| + |V| = |U \cup V|$, so we have finite additivity. However, it is not clear how we can generalize this to non open sets, except by means of upper bounding the measure of a set from above, and in the higher dimensions $\mathbf{R}^d$, we do not have a unique decomposition into boxes, so the definition is not immediately well defined.

Nonetheless, we do have a generalization of this theorem into higher dimensions. We say two boxes R and S are **almost disjoint** if they only intersect at their boundary. The technique of this proof uses an important fact about intervals with *dyadic endpoints*, which are those with rational endpoints of the form $n/2^m$; these intervals are either nested or almost disjoint.

Theorem 1.9. *Every open set in $\mathbf{R}^d$ can be written as the almost disjoint union of countably many boxes.*

Proof. Let U be an open set. For each m, we will successively define $U_m \subset U$ by adding dyadic boxes such that $\bigcup U_m = U$. Set $U_0 = \varnothing$. For each $m > 0$, consider the dyadic intervals at the resolution m. For each box containing points not already in U_{m-1}, but fully contained within U, we add the box to U_m. Because of the properties of dyadic intervals, these boxes are almost disjoint from the boxes in U_{m-1}. If $x \in U$, then x is contained in a dyadic box in U of a suitably small resolution, and so $U_m \to U$ as $m \to \infty$. $\qquad\square$

The volume of the boundary of a box should be equal to zero, because it is sufficiently small. Thus, if we have written a set U as the almost disjoint

union of boxes R_n, it makes sense to define

$$|U| = \sum_{n=1}^{\infty} |R_n|$$

Though this is intuitive, it is not immediately obvious that this definition is independent of the decomposition of U into the R_n, though this identity will hold for the definition we construct.

Corollary 1.10. *For any set E, $\mu^*(E) = \inf\{\mu^*(U) : U \text{ open}\}$.*

Proof. Let E be covered by rectangles $S_1, S_2, \dots$. We consider slight enlargements S_n' of the S_n, possibly increasing their total area by some ε, so that the S_n are open. Then $\bigcup S_n'$ is open, and

$$\inf\{\mu^*(U) : U \text{ open}\} \leqslant \sum |S_n'| = \sum |S_n| + \varepsilon$$

Since the S_n are arbitrary, we conclude that

$$\inf\{\mu^*(U) : U \text{ open}\} \leqslant \mu^*(E) + \varepsilon$$

and we can take $\varepsilon \to 0$. But $\mu^*(E) \leqslant \mu^*(U)$ for any open U containing E, because μ^* is monotone, and so we obtain the required equality by taking infima on the right. $\qquad\square$

Though the exterior measure is not necessarily additive, it is additive if sets are not entangled from one another. If E and F are subsets of a metric space, then define the **Hausdorff distance**

$$d(E, F) = \inf\{d(x, y) : x \in E, y \in F\}$$

which measures the entanglement of two sets.

Theorem 1.11. *If $d(E, F) > 0$, then $\mu^*(E \cup F) = \mu^*(E) + \mu^*(F)$.*

Proof. By subadditivity, it suffices to prove that $\mu^*(E \cup F) \geqslant \mu^*(E) + \mu^*(F)$. Let $E \cup F$ be covered by rectangles $R_1, R_2, \dots$. By subdividing each rectangle, we may assume that each is contained within circles of radius $d(E, F)/2$. Then each R_i either only contains points in E, only contains points in F, or contains points in neither set, and so we can partition the rectangles into a cover $S_1, S_2, \dots$ of E, and a cover $S_1', S_2', \dots$ of F, hence

$$\mu^*(E) + \mu^*(F) \leqslant \sum |S_i| + \sum |S_i'| = \sum |R_i|$$

Since the R_n were arbitrary, we conclude that $\mu^*(E) + \mu^*(F) \leqslant \mu^*(E \cup F)$. $\quad\square$

Example. *If E is the countable union of almost disjoint rectangles $R_1, R_2, \ldots$, then $\mu^*(E) = \sum |R_n|$. If we consider slightly smaller rectangles R'_n with total area only ε less then the area of the R_n, we may assume the R'_n are closed and completely disjoint from one another. Compactness shows $d(R'_n, R'_m) > 0$ if $n \neq m$, and so applications of the last theorem yield that*

$$\mu^* \left(\bigcup_{n=1}^{N} R'_n \right) = \sum_{n=1}^{N} |R'_n|$$

Letting $N \to \infty$, we conclude by monotonicity that

$$\mu^*(E) \geqslant \sum_{n=1}^{\infty} |R'_n| \geqslant \sum_{n=1}^{\infty} |R_n| - \varepsilon$$

Letting $\varepsilon \to 0$, we conclude that $\mu^*(E) \geqslant \sum |R_n|$. But we already know that $\mu^*(E) \leqslant \sum |R_n|$ by countable subadditivty, so $\mu^*(E) = \sum |R_n|$.

As with the theory of Jordan measure, it is not in general true that for any two sets E and F, $\mu^*(E \cup F) = \mu^*(E) + \mu^*(F)$, as Banach and Tarski warned us. This means we cannot define the Lebesgue measure $|E|$ for all sets. For the class of Lebesgue measurable sets, we will show that we can obtain additivity. There isn't a great definition of the Lebesgue inner measure, because fractal sets may not even contain rectangles, let alone countably many. But we saw that a good alternate definition for Jordan measurability was the fact that these sets were well approximated by Jordan measurable sets, and we use this for the Lebesgue measure definition. We say a set E is **Lebesgue measurable** if, for every $\varepsilon > 0$, there is an open set U wth $\mu^*(U - E) \leqslant \varepsilon$. Note that for any set E with $\mu^*(E) < \infty$, there is an open set $U \supset E$ with $\mu^*(U) - \mu^*(E) \leqslant \varepsilon$, but this doesn't necessarily imply $\mu^*(U - E) \leqslant \varepsilon$.

Theorem 1.12. *These properties are used to show Lebesgue measurability:*

(a) *Every open set is measurable.*

(b) *If $\mu^*(E) = 0$, then E is Lebesgue measurable.*

(c) *Countable unions of measurable sets are measurable.*

(d) *Closed sets are measurable.*

21

(e) The complement of a measurable set is measurable.

(f) Countable intersections of measurable sets are measurable.

(g) Every Jordan measurable set is Lebesgue measurable.

Proof. Every open set approximates itself, so it is obvious that every open set is measurable. Now suppose $\mu^*(E) = 0$. Then there is $R_1, R_2, \ldots$ covering E with $\sum |R_i| \leqslant \varepsilon$. Without loss of generality, we may assume that R_i are open. Then

$$\mu^*(\bigcup R_i \,\triangle\, E) = \mu^*(\bigcup R_i - E) \leqslant \mu^*(\bigcup R_i) \leqslant \varepsilon$$

If $E_1, E_2, \ldots$ are measurable, then we can find $U_1, U_2, \ldots$ with $\mu^*(U_n - E_n) \leqslant \varepsilon/2^n$, and then

$$\mu^* \left(\bigcup E_n - \bigcup U_n \right) \leqslant \bigcup (E_n - U_n)$$

so

$$\mu^* \left(\bigcup E_n - \bigcup U_n \right) \leqslant \sum_{n=1}^{\infty} \varepsilon 2^{-n} \leqslant \varepsilon$$

which proves (c). To prove (d), it suffices to show that a compact set K is measurable, because every closed set is the countable union of compact sets. We know $\mu^*(K) < \infty$, because K is contained in some open box R, and monotonicity implies $\mu^*(K) \leqslant |R| < \infty$. Consider some open set U containing K with $\mu^*(U) - \mu^*(K) \leqslant \varepsilon$. Then $U - K$ is open, and therefore the union of almost disjoint closed rectangles $R_1, R_2, \ldots$. But this means that $\mu^*(U - K) = \sum |R_i|$, and by subadditivity,

$$\mu^*(U) = \mu^*(K) + \sum |R_i|$$

and our discussion entails that $\sum |R_i| \leqslant \varepsilon$, hence $\mu^*(U - K) \leqslant \varepsilon$. To prove (e), we note that if E is measurable, and there is an open set U with $\mu^*(U \triangle E) \leqslant \varepsilon$, then $E^c \triangle U^c = E \triangle U$, and since U^c is closed, there is another open set V with $\mu^*(U^c \triangle V) \leqslant \varepsilon$, and so

$$\mu^*(E^c \triangle V) \leqslant \mu^*(E^c \triangle U^c) + \mu^*(U^c \triangle V) = \mu^*(E \triangle U) + \mu^(U^c \triangle V) \leqslant 2\varepsilon$$

and we can let $\varepsilon \to 0$. (f) is then a trivial consequence of the fact that $(\bigcup E_n)^c = \bigcap E_n^c$. If E is Jordan measurable, this means that there are elementary sets E_1, E_2 with $E_1 \subset E \subset E_2$ and $|E_2 - E_1| = \mu^*(E_2 - E_1) \leqslant \varepsilon$. By choosing E_2 slightly larger, we can assume it is open, and then $\mu^*(E_2 - E) \leqslant \mu^*(E_2 - E_1) \leqslant \varepsilon$, so E is Jordan measurable. $\qquad\square$

Here are some alternate characterizations of measurability.

Theorem 1.13. *The following are equivalent for any set E:*

(a) *For any $\varepsilon > 0$, there is an open set $U \supset E$ with $\mu^*(U - E) \leqslant \varepsilon$. This is our definition of Lebesgue measurability.*

(b) *For every $\varepsilon > 0$, there is an open set U with $\mu^*(E \bigtriangleup U) \leqslant \varepsilon$.*

(c) *For all $\varepsilon > 0$, there is a closed set $F \subset E$ with $\mu^*(E - F) \leqslant \varepsilon$.*

(d) *For all $\varepsilon > 0$, there is a closed set F with $\mu^*(E \bigtriangleup F) \leqslant \varepsilon$.*

(e) *For every $\varepsilon > 0$, there is a Lebesgue measurable set F with $\mu^*(E \bigtriangleup F) \leqslant \varepsilon$.*

Thus these conditions provide an alternate definition of measurability.

Proof. (a) implies (b) is trivial, as is (c) to (d) and (d) to (e). The converse, (b) to (a) is given by the following strategy. If U is given with $\mu^*(E \bigtriangleup U) \leqslant \varepsilon$, then there is an open set V covering $E \bigtriangleup U$ with $\mu^*(V) \leqslant \varepsilon$, and then $U \cup V$ is an open set covering E with

$$\mu^*(U \cap V - E) \leqslant \mu^*(U - E) + \mu^*(V - E) \leqslant \mu^*(U - E) + \mu^*(V) \leqslant 2\varepsilon$$

To show that (a) implies (c), we use the fact that if E is Lebesgue measurable, then E^c is Lebesgue measurable. Thus if there is an open set U covering E^c with $\mu^*(U - E^c) \leqslant \varepsilon$, then $U - E^c = E - U^c$, and U^c is closed, hence we have proved (c) holds. All that remains is to prove (e) implies (a). But this follows because if F is Lebesgue measurable, then there is an open set U with $\mu^*(U - F) \leqslant \varepsilon$, and an open set V covering $E \bigtriangleup F$ with $\mu^*(V) \leqslant \varepsilon$. Then

$$U \cup V \bigtriangleup E = U \cup V \bigtriangleup F \bigtriangleup F \bigtriangleup E$$

and so

$$\mu^*(U \cup V \bigtriangleup E) \leqslant \mu^*(U \bigtriangleup F) + \mu^*(V \bigtriangleup F) + \mu^*(F \bigtriangleup E) \leqslant 3\varepsilon$$

Hence we have shown (e) implies (b). $\qquad\square$

If E is Lebesgue measurable, we define $|E| = \mu^*(E)$ to be the **Lebesgue measure** of the set E. Unlike the value $\mu^*(E)$, which we view as a potential candidate for measuring the size of a set, which may fail for sets which

aren't well approximated, the Lebesgue measure has proven itself to be *the* correct definition of length, area, and volume in mathematical analysis. We now show that that Lebesgue measurable sets cannot be entangled enough for countable additivity to break (they reach all depths of the set, so that features of the sets not reached by the measure cannot be added up contably to obtain a significant feature of the set).

Remark. *If E is Lebesgue measurable, we can take U to be an open set well covering E, with $\mu^*(E \triangle U) \leqslant \varepsilon$. If $|E| < \infty$, then $|U| < \infty$, and if we write U as the countable union of almost disjoint boxes $R_1, R_2, \ldots$, then*

$$\sum_{n=1}^{\infty} |R_n| = |U|$$

we can choose N large enough that

$$\sum_{n=1}^{N} |R_n| \geqslant |U| - \varepsilon$$

If we now consider the elementary set $R = R_1 \cup \cdots \cup R_N$, then $\mu^*(U - R) \leqslant \varepsilon$, and so $\mu^*(E \triangle R) \leqslant 2\varepsilon$. Thus, with respect to area, the Lebesgue measurable sets are those which can be 'well pixelated' by elementary sets. This is different from Jordan measurable sets, which can be both well encapsulated by elementary sets, and cover an elementary set well – here we allow R to be contain new points, and to avoid certain points of E.

Note that the theorem above shows that the family of Lebesgue measurable sets is closed under unions and complements, and since $\mu^*(\varnothing) = 0$, $\varnothing$ is Lebesgue measurable. Thus the family is an algebra even without restricting the family to a bounded region, and what's more, not only can we take finite unions, we can take countable unions of Lebesgue measurable sets in the family as well. We call an algebra where we can take countable unions a σ **algebra**. Another important σ algebra is the **Borel algebra**, which is the smallest σ algebra containing all open and closed subsets of $\mathbf{R}^d$ (this is well defined since the intersection of an arbitrary family of σ algebras is also a σ algebras). The Borel algebra is slightly smaller than the Lebesgue algebra, but has slightly nicer properties of the Lebesgue algebra which make it more suitable for certain applications, such as in the theory of random variables in probability. Almost every result for Borel set is true for Lebesgue measurable sets, as is proved by the next theorem.

24

Theorem 1.14. *A set E is Lebesgue measurable if and only if there is a Borel measurable set F with $\mu^*(E \triangle F) = 0$.*

Proof. If E is Lebesgue measurable, then for every ε then for every n is an open set U_n containing E with $\mu^*(U_n - E) \leqslant \varepsilon$, and we can define compact sets K_n with $\mu^*(E - K_n) \leqslant \varepsilon$, then setting F to be $\bigcap U_n$ or $\bigcup K_n$ leads to the conclusion of the theorem. □

We define the families G_δ, and F_σ, which consist of countable intersections of open sets, and countable unions of closed sets, respectively. Our proof above shows not only that Lebesgue measurable sets differ from Borel sets by a set of measure zero, but that they differ from a G_δ or an F_σ set by a set of measure zero.

Remark. *We can define a metric on the family of Lebesgue measurable sets, where the distance between two sets E and F is $|E \triangle F|$. It is fairly easy to prove the triangle inequality in this scenario, and it has been used freely in the theorem above. The metric also makes sense on the family of elementary sets. The theorem above shows that, though the metric space of elementary sets is not complete, the space of Lebesgue measurable sets is precisely a completion of this metric space. The use of G_δ and F_σ sets above to approximate Lebesgue measurable sets respect the fact that elements of the completion can be characterized by cauchy sequences in the original space. If $E_1, E_2, \ldots$ are a Cauchy sequence of Lebesgue measurable sets, then we may choose a subsequence $E_{n_1}, E_{n_2}, \ldots$ with $|E_{n_k} \triangle E_{n_{k+1}}| \leqslant 1/2^{n+1}$. Then the sets*

$$F_N = \bigcup_{n=N}^{\infty} E_{n_k}$$

are decreasing and Cauchy, since $|F_N| \leqslant \sum_{n=N}^{\infty} E_{n_k} \leqslant 2^{-N}$, and $|F_N \triangle E_N| \leqslant 2^{-N}$, so the E_n converge to $\lim F_N$, since the F_n also converge to this limit.

Theorem 1.15. *If $E_1, E_2, \ldots$ are disjoint Lebesgue measurable sets, and $E = \bigcup E_n$, then*

$$|E| = \sum |E_n|$$

So the Lebesgue measure has the property of countable additivity.

Proof. Without loss of generality, assume $|E_n| < \infty$ for all n, for the theorem is trivial otherwise. It suffices by subadditivity to show that $|E| \geqslant$

$\sum |E_n|$, for otherwise the theorem is trivial. We can find compact $K_n \subset E_n$ with $\mu^*(E_n - K_n) \leqslant \varepsilon 2^{-n}$. Then $d(K_n, K_m) > 0$, and so we find

$$|E| \geqslant \left| \bigcup K_n \right| = \sum |K_n| \geqslant \sum |E_n| - \varepsilon$$

and we can then let $\varepsilon \to 0$. □

The countable additivity of the Lebesgue measure has the following important consequence, which is equivalent, and equally important. If we have a nested family of sets $E_1 \subset E_2 \subset \ldots$, or $F_1 \supset F_2 \supset \ldots$, then we define the limits

$$\lim_{n \to \infty} E_n = \bigcup_{n=1}^{\infty} E_n \qquad \lim_{n \to \infty} F_n = \bigcap_{n=1}^{\infty} F_n$$

which makes sense, since the functions χ_{E_n} and χ_{F_n} converge pointwise to the functions $\chi_{\lim E_n}$ and $\chi_{\lim F_n}$.

Theorem 1.16. *If each E_n is a Lebesgue measurable set, then*

$$\left| \lim_{n \to \infty} E_n \right| = \lim_{n \to \infty} |E_n|$$

and if the F_n are Lebesgue measurable, with $|F_N| < \infty$ for some n, then

$$\left| \lim_{n \to \infty} F_n \right| = \lim_{n \to \infty} |F_n|$$

Proof. Define the sets $E_n' = E_n - E_{n-1}$, with $E_1' = E_1$. Then the E_n' are disjoint and measurable, and $\lim E_n = \bigcup E_n'$, and $E_N = \bigcup_{n=1}^{N} E_n'$, so that

$$\lim_{N \to \infty} |E_N| = \lim_{N \to \infty} \sum_{n=1}^{N} |E_n'| = \sum_{n=1}^{\infty} |E_n'| = |\lim E_n|$$

countable additivity proves the theorem. On the other hand, if $|F_N| < \infty$, and we define $F_{N+n}' = F_N - F_{N+n}$, then the F_n' are increasing, with $\lim F_n' = F_N - \lim F_n$, hence the last part of this theorem implies

$$|F_N| - |\lim F_n| = |F_N - \lim F_n| = |\lim F_n'| = \lim |F_n'| = |F_N| - \lim |F_n|$$

and we can rearrange this equation to get the equation required. □

Remark. *If $S_1, S_2, \ldots$ are boxes defined by the equation*

$$S_n = \{x : |x_i| \leqslant n\}$$

Then S_n are an increasing family of sets whose union is $\mathbf{R}^d$, and so this theorem guarantees that for any set E, $|E| = \lim |E \cap S_n|$. Now $E \cap S_n$ is measurable, and therefore can be approximated from inside by closed sets. Since $E \cap S_n$ is bounded, these closed sets are actually compact, and so we have argued that for any measurable set E,

$$|E| = \sup\{|K| : K \subset E, K \text{ compact}\}$$

This is known as the inner regularity of the Lebesgue measure. There is other ways to define the Lebesgue measure leading us to define $|K|$ in addition to $|U|$ before defining the measure of general sets. If we the right hand side of the equation as an inner measure $\mu_*(E)$, defined for any set E, then a set is Lebesgue measurable precisely when $\mu_*(E) = \mu^*(E)$.

This completes our construction of the Lebesgue measure, bringing our intuitive notions of length and volume forth into a rigourous framework by means of approximation by open and closed sets.

1.4 Symmetries of the Lebesgue Measure

A crucial property of the Lebesgue integral, especially in contexts such as the Fourier transform, is that it is translation invariant. In particular, if E is a subset of $\mathbf{R}^d$, and for $x \in \mathbf{R}^d$ we define

$$x + E = \{x + y : y \in E\}$$

then $\mu^*(x + E) = \mu^*(E)$, and E is measurable if and only if $x + E$ is measurable. In fact, the Lebesgue measure is essentially uniquely characterized by translation invariance, up to a constant, as a countably additive function on the σ algebra of Borel sets.

Another important symmetry is obtained from the fact that $\mathbf{R}^d$ is a vector space, and so we can consider invertible linear operators $T : \mathbf{R}^d \to \mathbf{R}^d$, which have a nonzero determinant. The chain rule in basic multivariate calculus then says that, if f is Riemann integrable, then

$$\int f(Tx) \, dx = |\det T| \int f(x) \, dx$$

In particular, this implies that if R is any box, then $|TR| = |\det T||R|$. This then implies that $|TU| = |\det T||U|$ for any open set U, which is the countable union of boxes. But now we find that $\mu^*(TE) = |\det T||E|$, because μ^* can be defined in terms of an outer approximation by open sets, and so $|TE| = |\det T||E|$ for any open set. In particular, if we define

$$\delta E = \{\delta x : x \in E\}$$

for $\delta > 0$, which is the set E scaled down or up by some constant amount, then $|\delta E| = \delta^d \mu^*(E)$. Again, this symmetry is also reflected in the properties of the Fourier transform.

1.5   Banach Tarski and Nonmeasurable Sets

We now prove the Banach Tarski theorem, assuming only some basic group theory. A nice property of the sphere in three dimensions is that it is invariant under rotation about the origin. Mathematically, we say that the orthogonal group $O(3)$ of orthogonal linear transformations acts on the sphere S^2. The core technique of the Banach Tarski theorem can be executed in a simpler form n free groups. Consider the free group $F(a,b)$ on two characters a and b. Let $S(x)$ be the set of all sequences in the group whose simplest form begins with an x. Then we have a decomposition

$$F(a,b) = \{e\} \cup S(a) \cup S(b) \cup S(a^{-1}) \cup S(b^{-1})$$

and $F(a,b) = S(a) \cup aS(a^{-1})$, $F(a,b) = S(b) \cup bS(b^{-1})$. This means we have a partition of $F(a,b)$ into four sets, and by 'rotating' two of the four of the sets the partition, we obtain two copies of the group. We will say a subset E of $\mathbf{R}^3$ can be *paradoxically decomposed* if it can be expressed as the disjoint union of subsets $E_1, \ldots, E_{n+1}$, and, under rotations $T_1, \ldots, T_n \in O(3)$, we can write $E = T_1 E_1 \cup \cdots \cup T_m E_m$ and $E = T_{m+1} E_{m+1} \cup \cdots \cup T_n E_n$, for some m (we can do the Banach Tarski paradox 'without leftovers', i.e., without E_{n+1}, but there's some more technicality that isn't really that interesting to discuss).

Lemma 1.17. *There is a subgroup of $O(3)$ isomorphic to $F(a,b)$.*

Proof. Map a to a horizontal rotation by $\arccos(1/3)$, which in quaternion notation is given by

$$\frac{\sqrt{2}+k}{\sqrt{3}}$$

28

and map b to a vertical rotation by $\arccos(1/3)$, which is expressed by the quaternion

$$\frac{\sqrt{2}+j}{\sqrt{3}}$$

To show that this maps $F(a,b)$ injectively into $O(3)$, it suffices to see that $(1,0,0)$ is only fixed by the identity transformation, and cannot be returned to its original place by applications of a and b. $\qquad\square$

Theorem 1.18 (Banach Tarski). *The sphere may be paradoxically decomposed.*

Proof. Partition S^2 into orbits for the faithful group action of $F(a,b)$ we constructed, and let M be a set obtained by selecting a single point from every orbit. Also define

$$B = a^{-1}M \cup a^{-2}M \cup \ldots$$

Then consider the partition

$$S(a)M \cup M \cup B \quad S(a^{-1})M - B \quad S(b)M \quad S(b^{-1})M$$

Then

$$a[S(a^{-1}M - B] = S(a^{-1})M - B \cup S(b)M \cup S(b^{-1})M$$
$$b[S(b^{-1})M] = S(a)M \cup M \cup B \cup S(a^{-1})M - B \cup S(b^{-1})M$$

$$\qquad\square$$

This set, as a subset of $\mathbf{R}^3$, is actually measurable, despite this decomposition, but this is only because of the fact that it is contained within S^2, which has measure in $\mathbf{R}^3$. But if we now consider rays through the origin, and extend the set M by projection into the unit ball, then we obtain a true non measurable subset of $\mathbf{R}^3$.

An alternate construction, using the same techniques, takes the group $\mathbf{T}$, which can be identified with the interval $[0,1]$, with the operation of addition modulo 1. Then there is a subgroup $\mathbf{Q} \cap [0,1]$, and so $\mathbf{Q} \cap [0,1]$ acts on $[0,1]$. Consider a set E obtained by picking one element from each of the orbits of this action. Then the sets $q + E$ are disjoint, for distinct elements of $\mathbf{Q} \cap [0,1]$, and their union (which is countable), is $[0,1]$. If E was measurable, then translation invariance would guarantee $q + E$ was measurable for all q, and so by translation invariance

$$1 = |[0,1]| = \sum_q |q + E| = \sum_q |E| = \infty \cdot |E|$$

29

Thuse we cannot have $|E| = 0$, nor $|E| > 0$, so E must be a nonmeasurable set. The set E also provides a way to construct disjoint sets E and F which do not satisfy the equation $\mu^*(E \cup F) = \mu^*(E) + \mu^*(F)$. μ^* is translation invariant, so $\mu^*(q + E) = \mu^*(E)$, and if $\mu^*(E) = 0$, then we would conclude by subadditivity that

$$1 = \mu^*([0,1]) \leqslant \sum_q \mu^*(E) = 0$$

So $\mu^*(E) > 0$. If we enumerate the rationals in $[0,1]$ by $q_1, q_2, \ldots$, and consider $q_1 + E, q_2 + E, \ldots$, and then define $F_{n+1} = F_n \cup q_n + E$, with $F_1 = E$ then eventually $\mu^*(F_n) \neq n\mu^*(E)$, for otherwise eventually we would find that $\mu^*(F_n) > 1$.

Chapter 2

Lebesgue Measurable Functions

Chapter 3

Differentiation and Averages

If $f : [a,b] \to \mathbf{R}$ is a continuous function, then the fundamental theorem of calculus says that the function

$$F(x) = \int_a^x f(t)\, dt$$

is differentiable on $[a,b]$, and $F'(x) = f(x)$. In this chapter, we attempt to generalize this problem to a more general class of functions. The formula above allows us to define a function F for any function $f \in L^1[a,b]$, and we ask under what conditions F is differentiable, and $F' = f$. In analyzing this problem, we shall use the fundamental tool of maximals functions introduced by Hardy and Littlewood, which is a general strategy for establishing regularity results for various operators on function spaces encountered in analysis. This tool is used in *density arguments*, which consist of two parts

- Establish a result for a suitably nice family of functions, which are dense in the class of all functions we wish to verify.

- A quantitative estimate bounding the maximal variation of some quantity in terms of some suitable norm for f. One often applies tools such as Fatou's lemma here, since we only need an upper bound.

With these two tools, one can normally establish a result for all functions, rather than just the nice class. This idea has been used in our notes to proof that for any $f \in L^1(\mathbf{R}^d)$, the functions f_h converge to f in the L^1 norm as $h \to 0$, where we proved the result for continuous functions, then bounded

$\|f_h - f\|_1$ in terms of $\|f\|_1$. The theory of maximal functions explores the second aspect of the technique above.

3.1   Maximal Functions and Differentiation

Let us start with analyzing the problem of differentiating the integral of a measurable function. Given $f \in L^1[a,b]$, we consider the function

$$F(x) = \int_a^x f(t)\, dt$$

The function F is uniformly continuous, because, as f is integrable, for every $\varepsilon > 0$, there is $\delta > 0$ such that if $|E| \leqslant \delta$, $\int_E |f| \leqslant \varepsilon$, and this means that for every $x \in [a, b - \delta]$,

$$|F(x + \delta) - F(x)| = \left| \int_x^{x+\delta} f(t)\, dt \right| \leqslant \int_x^{x+\delta} |f(t)|\, dt < \varepsilon$$

To understand the derivative of F, we now consider the limit of the quotients

$$\frac{F(x + h) - F(x)}{h} = \frac{1}{h} \int_x^{x+h} f(t)\, dt$$

We can interpret the quotient on the right as the average value of f on the interval $[x, x + h]$, and we begin to see how the fundamental theorem of calculus is implicitly tied to the problem of understanding 'averages' of f around small sets containing x. Because of this realization, we switch to studying the general question of, for an integrable $f : \mathbf{R}^d \to \mathbf{R}$,

$$\lim_{\substack{|E| \to 0 \\ x \in E}} \frac{1}{|E|} \int_E f(y)\, dy = f(x)$$

for a family of measurable sets E. It turns out that the validity of this result depends on the geometry of the sets E we consider. In this section we consider the simplest case, where E ranges over all open balls B.

Lemma 3.1. *If f is continuous at x, then*

$$\lim_{\substack{|B| \to 0 \\ x \in B}} \frac{1}{|B|} \int_B f(y)\, dy = f(x)$$

33

Proof. If ε is fixed, and δ is chosen such that $|f(y) - f(x)| < \varepsilon$ for $|y - x| < \delta$, and if B is a ball of radius $\delta/2$ containing x, every $y \in B$ has $|y - x| < \delta$. This implies

$$\left\| \frac{1}{|B|} \int_B [f(y) - f(x)] \, dy \right\| \leqslant \frac{1}{|B|} \int_B |f(y) - f(x)| \, dy \leqslant \frac{1}{B} \int_B \varepsilon = \varepsilon$$

hence the limit exists, and is equal to $f(x)$. Note this argument wouldn't work if we took the limit over general measurable sets E containing x, because the diameter of the sets E is not bounded by $|E|$ in any reasonable fashion. $\square$

In order to obtain results for non-continuous functions $f \in L^1[a,b]$, it would be nice to have an upper bound on the averages of f which allow us to apply the various convergence results of measure theory. This can be obtained by studying the **Hardy-Littlewood maximal function**

$$Mf(x) = f^*(x) = \sup_{x \in B} \frac{1}{|B|} \int_B |f(y)| \, dy$$

This function provides an upper bound on the function f. We would like to conclude that f^* is an integrable function, so we can apply dominated convergence-type results, but this is not always the case. Instead, we can only obtain a *weak L^1 inequality* establishing that $f^* \in L^{1,w}[a,b]$, which is the space of measurable functions g for which the **weak L^1 norm**

$$\|g\|_{1,w} = \sup_{t \geqslant 0} t|\{g > t\}|$$

is finite. In particular, this means $\{f^* > t\} \lesssim 1/t$, which will be sufficient to obtain results about the averaging problem. Our argument establishing the weak inequality rests on a combinatorial result about coverings of sets.

Lemma 3.2 (Vitali Covering Lemma)**.** *If $B_1, \ldots, B_n$ is a finite collection of balls in $\mathbf{R}^d$, then there is a disjoint subcollection $B_{i_1}, \ldots, B_{i_m}$ such that*

$$\left| \bigcup B_i \right| \leqslant 3^d \sum |B_{i_k}|$$

Proof. We begin with the elementary observation that if B and B' are two intersecting balls with the radius of B greater than the radius of B', then B'

34

is a subset of $3B$. Now, if we choose the balls greedily, picking a ball with maximal radius that doesn't intersect any of our previously chosen balls. It then follows that if $B_{i_1}, \dots B_{i_m}$ are the balls chosen then

$$\bigcup B_i \subset \bigcup 3B_{i_k}$$

and so

$$\left| \bigcup B_i \right| \leqslant \left| \bigcup 3B_{i_k} \right| \leqslant \sum |3B_{i_k}| = 3^d \sum |B_{i_k}|$$

and we therefore obtain the required inequality. $\qquad \square$

Theorem 3.3. *The function f^* is measurable, and for any $\alpha > 0$,*

$$|\{x \in \mathbf{R}^d : f^*(x) > \alpha\}| \leqslant \frac{3^d}{\alpha} \|f\|_1$$

Thus $\|f^*\|_{1,w} \leqslant 3^d \|f\|_1$, and $f^*(x) < \infty$ almost surely.

Proof. The measurability of f^* is easy to see, because

$$\{x \in \mathbf{R}^d : f^*(x) > \alpha\}$$

is open. This is because if $f^*(x) > \alpha$, we can choose a ball B containing a point x such that the average of $|f|$ on B is greater than α, and then for any $y \in B$, $f^*(y) > \alpha$ (Remark: this means f^* is lower semicontinuous). Set $E_\alpha = \{x \in \mathbf{R}^d : f^*(x) > \alpha\}$. Then for each x, we can find a ball B_x such that the average of $|f|$ on B_x is greater than α. This implies that

$$|B_x| < \frac{1}{\alpha} \int_{B_x} |f(y)| \, dy$$

Let $K \subset E_\alpha$ be compact. Then K is covered by finitely many of the balls, and applying the Vitali covering lemma, we can find a disjoint collection $B_{x_1}, \dots, B_{x_m}$ of these balls such that

$$|K| \leqslant 3^d \sum |B_{x_m}| < \frac{3^d}{\alpha} \int_{\bigcup B_{x_i}} |f(y)| \, dy \leqslant \frac{3^d}{\alpha} \|f\|_1$$

and since K was arbitrary, we obtain that $|E_\alpha| \leqslant 3^d \|f\|_1 / \alpha$. $\qquad \square$

Theorem 3.4 (Lebesgue Differentiation Theorem). *If f is integrable on $\mathbf{R}^d$, then for almost every x,*

$$\lim_{\substack{|B| \to 0 \\ x \in B}} \frac{1}{|B|} \int f(y) \, dy = f(x)$$

Proof. It suffices to show that for all $\alpha > 0$,

$$E_\alpha = \left\{ x \in \mathbf{R}^d : \limsup_{\substack{|B| \to 0 \\ x \in B}} \left| \frac{1}{|B|} \int_B [f(y) - f(x)] \, dy \right| > \alpha \right\}$$

is a set of measure zero. If we fix ε, and find a function $g \in C_c(\mathbf{R}^d)$ such that $\|f - g\|_1 < \varepsilon$, then

$$\left| \frac{1}{|B|} \int_B [f(y) - f(x)] \right| = \left| \frac{1}{|B|} \int_B [f(y) - g(y)] + [g(y) - g(x)] + [g(x) - f(x)] \, dy \right|$$

since g is continuous, we know the averages of g around a point convergence, and therefore

$$\limsup_{\substack{|B| \to 0 \\ x \in B}} \left| \frac{1}{|B|} \int_B [f(y) - f(x)] \right| \leqslant (f - g)^*(x) + |g(x) - f(x)|$$

If F_α is the set of values where $(f - g)^*(x) > \alpha/2$, and G_α is the set of values where $|g(x) - f(x)| > \alpha_2$, then $E_\alpha \subset F_\alpha \cup G_\alpha$, and

$$|F_\alpha| \leqslant \frac{2}{\alpha} \|(f - g)^*\|_{1,w} \leqslant \frac{2 \cdot 3^d \varepsilon}{\alpha} \qquad |E_\alpha| \leqslant \frac{\|f - g\|_1}{\alpha} \leqslant \varepsilon$$

But ε is arbitrary, so F_α, G_α, and therefore E_α, are sets of measure zero. $\quad\square$

The integrability of f over all of $\mathbf{R}^d$ isn't applied anywhere in the proof. Indeed, we only needed that the integral of f is bounded over suitably small balls. If we define the space of **locally integrable functions** $L^1_{\text{loc}}(\mathbf{R}^d)$ to be all functions f such that $f \chi_B$ is integrable for any ball B, then the Lebesgue differentiation theorem continues to hold for f.

36

3.2   Lebesgue Density Theorem

If E is a measurable subset of $\mathbf{R}^d$, and $x \in \mathbf{R}^d$, we say x is a point of **Lebesgue density** of E, or has **metric denstity 1** if

$$\lim_{\substack{|B| \to 0 \\ x \in B}} \frac{|B \cap E|}{|B|} = 1$$

This means that for every $\alpha < 1$, for suitably small balls, we conclude that $|B \cap E| \geqslant \alpha |B|$, so E asymptotically contains as large a fraction of the points in B. Since $\chi_E \in L^1_{\text{loc}}(\mathbf{R}^d)$, we can apply the Lebesgue differentiation theorem to immediately obtain an interesting result.

Theorem 3.5 (Lebesgue Density Theorem). *If E is a measurable subset, then almost every point in E is a point of Lebesgue density, and almost every point in E is not a point of Lebesgue density.*

The fact that a point is a point of Lebesgue density implies the existence of large sets of rigid patterns in E. First, note that if

$$|B \cap E|, |B \cap F| \geqslant \alpha |B|$$

then a union bound gives

$$|B \cap E \cap F| \geqslant |B| - |B \cap E^c| - |B \cap F^c| \geqslant (2\alpha - 1)|B|$$

as $\alpha \to 1$, $2\alpha - 1 \to 1$, so if x is a point of Lebesgue density for E and F, then x is a point of Lebesgue density of $E \cap F$. If 0 is a point of Lebesgue density for E, then 0 is a point of Lebesgue density for αE for any $\alpha \neq 0$, and so for any nonzero $\alpha_1, \dots, \alpha_n$, there exists nonzero $x_i \in E$ converging to zero, with $\alpha_1 x_i, \dots, \alpha_n x_i \in E$. This is very difficult to prove without the existence of the Lebesgue density theorem, because of the discreteness of the equations. Applying these results gives the following result, which means that the problem of quantifying 'how large' sets avoiding arithmetic progressions can be must use finer statistics then just the Lebesgue measure.

Theorem 3.6. *Every subset of $\mathbf{R}$ of positive measure contains infinitely many arbitrarily long arithmetic progressions.*

If f is locally integrable, the **Lebesgue set** of f consists of all points $x \in \mathbf{R}^d$ such that $f(x)$ is finite and

$$\lim_{\substack{|B| \to 0 \\ x \in B}} \frac{1}{|B|} \int |f(y) - f(x)|\, dy = 0$$

If f is continuous at x, it is obvious to see that x is in the Lebesgue set of f, and if x is in the Lebesgue set of f, then the averages of f on balls around x coverge to $f(x)$.

Theorem 3.7. *If $f \in L^1_{loc}(\mathbf{R}^d)$, almost every point is in the Lebesgue set of f.*

Proof. For each rational number p, the function $|f - p|$ is measurable, so that there is a set E_p of measure zero such that for $x \in E_p^c$,

$$\lim_{\substack{|B| \to 0 \\ x \in B}} \frac{1}{|B|} \int_B |f(y) - p|\, dy \to |f(x) - p|$$

Taking unions, we conclude that $E = \bigcup E_p$ is a set of measure zero. Suppose $x \in E^c$, and $f(x)$ is finite. For any ε, there is a rational p such that $|f(x) - p| < \varepsilon$, and we know the equation above holds, so

$$\lim_{\substack{|B| \to 0 \\ x \in B}} \frac{1}{|B|} \int_B |f(y) - f(x)|\, dy$$

$$\leqslant \limsup_{\substack{|B| \to 0 \\ x \in B}} \frac{1}{|B|} \int_B |f(y) - p| + |p - f(x)|\, dy \leqslant 2\varepsilon$$

we can then let $\varepsilon \to 0$. Since $f(x)$ is finite for almost all x when f is locally integrable, this completes the proof. $\square$

It is interesting to note that if $f = g$ almost everywhere, then the set of points x where the averages of f on balls around x converges is the same as the set of points x where the averages of f on balls around x converges, and so we can in some sense define a 'universal' function h from the equivalence class of these functions such that the averages of h on balls around x always converge to $h(x)$ when the limit exists. However, this isn't often done, because it doesn't really help in the analysis of integrable functions. We note, however, that the Lebesgue set of a function does depend on the function chosen from the equivalence class.

3.3 Generalizing The Differentiation Theorem

A family of sets U_α universally containing a point x is said to **shrink regularly** to x, or has **bounded eccentricity** at x, if $\inf |U_\alpha| = 0$, and there is a constant $c > 0$ such that for each U_α, there is a ball B with $x \in B$, $U_\alpha \subset B$, and $|U_\alpha| \geqslant c|B|$. Thus U_α contains a large percentage of ball B around x.

Example. *The set of all open cubes in $\mathbf{R}^d$ containing x shrinks regularly to x, because if a cube U centered at y with side lengths r contains x, then using the existence of a constant C such that for all $x, y \in \mathbf{R}^d$,*

$$\|x - y\|_\infty \leqslant C\|x - y\|_2$$

we conclude that the cube is contained within a ball B of radius $2Cr$, and since $|U| = r^d$, and $|B|$ is proportional to $(2Cr)^d$ up to a constant, so that U has bounded eccentricity.

Example. *The set of all rectangles in $\mathbf{R}^d$ containing x does not shrink regularly, because we can let the rectangle have one large side length while keeping all other side lengths relatively small, and then a ball containing this rectangle must be incredibly large.*

Theorem 3.8. *If f is locally integrable on $\mathbf{R}^d$, and $\{U_\alpha\}$ shrinks regularly to x, then for every point x in the Lebesgue set of f,*

$$\lim_{|U_\alpha| \to 0} \frac{1}{|U_\alpha|} \int_{U_\alpha} f(y) \, dy = f(x)$$

Proof. We just calculate that for every x in the Lebesgue set of f,

$$\lim_{|U_\alpha| \to 0} \frac{1}{|U_\alpha|} \int_{U_\alpha} |f(y) - f(x)| \, dy = 0$$

This follows because if $U_\alpha \subset B_\alpha$, with $|U_\alpha| \geqslant C|B_\alpha|$, then

$$\frac{1}{|U_\alpha|} \int_{U_\alpha} |f(y) - f(x)| \, dy \leqslant \frac{1}{C|B_\alpha|} \int_{B_\alpha} |f(y) - f(x)| \, dy$$

and since $|U_\alpha| \to 0$, $|B_\alpha| \to 0$, giving us the result. $\qquad\square$

3.4 Approximations to the Identity

We now switch to the study of how we can approximate functions by convolutions of concentrated functions around the origin. In this section we define the various classes of such functions which give convergence results, to various degrees of strength. We say a family $K_\alpha \in L^1(\mathbf{R}^d)$ is a **good kernel** if it is bounded in the L^1 norm, for every α,

$$\int K_\alpha(x)\, dx = 1$$

and if for every $\delta > 0$,

$$\int_{|x| \geqslant \delta} |K_\alpha(x)|\, dx \to 0$$

as $\alpha \to \infty$. It requires only basic analysis to verify convergence results for good kernels.

Theorem 3.9. *If K_α is a good kernel, then for any absolutely integrable function f, $f * K_\alpha \to f$ in the L^1 norm, and $(f * K_\alpha)(x) \to f(x)$ for every x which is a point of continuity of f.*

Proof. Note that

$$\|(f * K_\alpha) - f\|_1 = \int |(f * K_\alpha)(x) - f(x)|\, dx$$

$$= \int \left| \int K_\alpha(y)[f(x-y) - f(x)]\, dy \right|\, dx$$

$$\leqslant \int |K_\alpha(y)| \|T_y f - f\|_1\, dy$$

where $(T_y f)(x) = f(x - y)$. We know that $\|T_y f - f\|_1 \to 0$ as $y \to 0$. Thus, for each ε, we can pick δ such that if $|y| < \delta$, $\|T_y f - f\|_1 \leqslant \varepsilon$, and if we pick α large enough that $\int_{|y| \geqslant \delta} |K_\alpha(y)|\, dy \leqslant \varepsilon$, and then

$$\|(f * K_\alpha) - f\|_1 \leqslant \varepsilon \int_{|y| < \delta} |K_\alpha(y)|\, dy + 2\|f\|_1 \int_{|y| \geqslant \delta} |K_\alpha(y)|\, dy \leqslant \varepsilon[\|K_\alpha\|_1 + 2\|f\|_1]$$

Since $\|K_\alpha\|_1$ is universally bounded over α, we can let $\varepsilon \to 0$ to obtain convergence. If x is a fixed point of continuity, and for a given $\varepsilon > 0$, we

pick $\delta > 0$ with $|f(y) - f(x)| \leqslant \varepsilon$ for $|y - x| < \delta$, then

$$
\begin{aligned}
|(f * K_\alpha)(x) - f(x)| &= \left| \int_{-\infty}^{\infty} f(y) K_\alpha(x - y) \, dy - f(x) \right| \\
&= \left| \int_{-\infty}^{\infty} [f(y) - f(x)] K_\alpha(x - y) \, dy \right| \\
&= \left| \int_{-\delta}^{\delta} [f(y) - f(x)] K_\alpha(x - y) \, dy \right| \\
&\quad + \left| \int_{|y| \geqslant \delta} [f(y) - f(x)] K_\alpha(x - y) \, dy \right| \\
&\leqslant \varepsilon \|K_\alpha\|_1 + [\|f\|_1 + f(x)] \int_{|y| \geqslant \delta} |K_\alpha(y)| \, dy
\end{aligned}
$$

If $\|K_\alpha\|_1 \leqslant M$ for all α, and we choose α large enough that $\int_{|y| \geqslant \delta} |K_\alpha(y)| \leqslant \varepsilon$, then we conclude the value about is bounded by $\varepsilon[M + \|f\|_1 + f(x)]$, and we can then let $\varepsilon \to 0$. $\qquad \square$

To obtain almost sure pointwise convergence of $f * K_\alpha$ to f, we must place stronger conditions on our family. We say a family $K_\delta \in L^1(\mathbf{R}^d)$, is an **approximation to the identity** if $\int K_\delta = 1$, and

$$
|K_\delta(x)| \lesssim \frac{\delta}{|x|^{d+1}} \quad |K_\delta(x)| \lesssim \frac{1}{\delta^d}
$$

where the constant bound is independent of x and δ. These assumptions are stronger than being a good kernel, because if K_δ is an approximation to the identity, then

$$
\int_{|x| \geqslant \varepsilon} |K_\delta(x)| \leqslant \int_\varepsilon^\infty \int_{S^{d-1}} \frac{C\delta}{r} \, d\sigma dr = C\delta |S^{n-1}| \int_\varepsilon^\infty \frac{dr}{r} \leqslant \frac{C\delta |S^{n-1}|}{\varepsilon}
$$

which converges to zero as $\delta \to 0$. Combined with

$$
\int_{|x| < \varepsilon} |K_\delta(x)| \leqslant C \int_0^\varepsilon \int_{S^{d-1}} \frac{r^{d-1}}{\delta^d} d\sigma dr = \frac{C\varepsilon^d |S^{n-1}|}{d\delta^d}
$$

This calculation also implies

$$
\|K_\delta\|_1 \leqslant C |S^{n-1}| \left[\frac{\delta}{\varepsilon} + \frac{\varepsilon^d}{\delta^d} \right]
$$

41

Setting $\varepsilon = \delta$ optimizes this value, and gives a bound

$$\|K_\delta\|_1 \leqslant 2C|S^{n-1}|$$

So an approximation to the identity is a stronger version of a good kernel.

Example. *If φ is a bounded function in $\mathbf{R}^d$ supported on the closed ball of radius one with $\int \varphi(x)\, dx = 1$, then $K_\delta(x) = \delta^{-d}\varphi(\delta^{-1}x)$ is an approximation to the identity, because by a change of variables, we calculate*

$$\int_{\mathbf{R}^d} \frac{\varphi(\delta^{-1}x)}{\delta^d} = \int_{\mathbf{R}^d} \varphi(x) = 1$$

Because φ is bounded, we find

$$|K_\delta(x)| \leqslant \frac{\|\varphi\|_\infty}{\delta^d}$$

Now K_δ is supported on a disk of radius δ, this bound also shows

$$|K_\delta(x)| \leqslant \frac{\delta\|\varphi\|_\infty}{|x|^{d+1}}$$

and so K_δ is an approximation to the identity. If φ is an arbitrary integrable function, then K_δ will only be a good kernel.

Example. *The Poisson kernel in the upper half plane is given by*

$$P_y(x) = \frac{1}{\pi}\frac{y}{x^2 + y^2}$$

where $x \in \mathbf{R}$, and $y > 0$. It is easy to see that

$$P_y(x) = y^{-1}P_1(xy^{-1})$$

And

$$\int \frac{1}{1 + x^2} = \arctan(\infty) - \arctan(-\infty) = \pi$$

We easily obtain the bounds

$$|P_y(x)| \leqslant \frac{\|P_1\|_\infty}{y} \qquad |P_y(x)| \leqslant \frac{y}{\pi|x|^2}$$

so the Poisson kernel is an approximation to the identity.

Example. *The heat kernel in $\mathbf{R}^d$ is defined by*

$$H_t(x) = \frac{e^{-|x|^2/4t}}{(4\pi t)^{d/2}}$$

where $\delta = t^{1/2} > 0$. Then $H_t(x) = \delta^{-d} H_1(x\delta^{-1})$, and

$$\int e^{-|x|^2/4} = \frac{1}{2^d} \int e^{-|x|^2} = \frac{|S^{n-1}|}{2^d} \int_0^\infty r^{d-1} e^{-r^2} dr$$

Example. *The Poisson kernel for the disk is*

$$\frac{P_r(x)}{2\pi} = \begin{cases} \frac{1}{2\pi} \frac{1-r^2}{1-2r\cos x + r^2} & : |x| \leqslant \pi \\ 0 & : |x| > \pi \end{cases}$$

where $0 < r < 1$, and $\delta = 1 - r$.

Example. *The Féjer kernel is*

$$\frac{F_N(x)}{2\pi} = \left\{ \frac{1}{2\pi N} \frac{\sin^2(Nx/2)}{\sin^2(x/2)} \right.$$

where $\delta = 1/N$.

As $\delta \to 0$, we may think of the K_δ as 'tending to the unit mass' Dirac delta function δ at the origin. δ may given a precise term, either in the theory of Lebesgue-Stieltjes measures or as a 'generalized function', in which it plays the role of the identity for convolution, but we don't need it to discuss the actual convergence results of the functiosn K_δ.

Theorem 3.10. *If $\{K_\delta\}$ is an approximation to the identity, and f is integrable on $L^1(\mathbf{R}^d)$, then $(f * K_\delta)(x) \to f(x)$ for every x in the Lebesgue set of f, and $f * K_\delta$ converges to f in the L^1 norm.*

Proof. We rely on the fact that if x is in the Lebesgue set, then the function

$$A(r) = \frac{1}{r^d} \int_{|y| \leqslant r} |f(x-y) - f(x)| \, dy$$

is a bounded continuous function of $r > 0$, converging to 0 as $r \to 0$. This means that if $\Delta(y) = |f(x-y) - f(x)||K_\delta(y)|$, then

$$\int \Delta(y) \, dy = \int_{|y| \leqslant \delta} \Delta(y) + \sum_{k=0}^\infty \int_{2^k\delta \leqslant |y| \leqslant 2^{k+1}\delta} \Delta(y)$$

43

The first term is easily upper bounded by $CA(\delta)$, and the k'th term of the sum by $C' 2^{-k} A(2^{k+1} \delta) \leqslant C'' 2^{-k}$ for constants C', C'' that do not depend on δ. Letting $\delta \to 0$ gives us the convergence result. $\qquad\square$

3.5 Differentiability of Measurable Functions

We now switch our object of study to finding a condition on a measurable function f which guarantees differentiability almost everywhere, such that the derivative is absolutely integrable, and

$$f(b) - f(a) = \int_a^b f'(t) \, dt$$

holds almost everywhere. One way we can solve our problem is to fix our attention to functions f obtained by indefinite integrals. The results we have established guarantee that this theorem holds. But this leads to the extended problem of considering ways to characterize the properties of functions that arise from these indefinite integrals. We shall find that if f has *bounded variation*, then most of these problems are answered.

This problem is very connected to the problem of the *rectifiability of curves*. If $x : [a, b] \to \mathbf{R}^d$ parameterizes a continuous curve in the plane, then, for a given partition $P = a \leqslant t_0 \leqslant \cdots \leqslant t_n$, we can consider an approximate length

$$L_P(x) = \sum_{k=1}^n |x(t_i) - x(t_{i-1})|$$

If x has a reasonable notion of length, then the straight lines between $x(t_{i-1})$ and $x(t_i)$ should be shorter than the length of x between t_{i-1} and t_i. It therefore makes sense to define the **length** of x as

$$L(x) = \sup L_P(x)$$

The triangle inequality implies that the map $P \mapsto L_P(x)$ is an increasing net, so L is also the limit of the meshes as they become finer and finer. If $L(x) < \infty$, we say x is a **rectifiable curve**. One problem is to determine what analytic conditions one must place on x in order to guarantee regularity, and what further conditions guarantee that, if x_i is differentiable

almost everywhere,

$$L(x) = \int_a^b \sqrt{x_1'(t)^2 + \cdots + x_n'(t)^2} \, dt$$

Considering rectifiable curves leads directly to the notion of a function with bounded variation.

If f is a complex valued function on $[a,b]$, and P is a partition, we can consider it's variation on a partition $P = a \leqslant t_0 < \cdots < t_n \leqslant b$ to be

$$V_P(f) = \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})|$$

we say f has **bounded variation** if there is a constant M such that for any partition P, $V_P(f) \leqslant M$. This implies that, since the net $P \mapsto V_P(f)$ is increasing, the net converges to a value $V(f)$, the **total variation** of f on $[a,b]$.

Theorem 3.11. *A curve x is rectifiable iff each x_i has bounded variation.*

Proof. We can find a universal constants $A, B > 0$ such that for any $x, y \in \mathbf{R}^d$,

$$A \sum |x_i - y_i| \leqslant |x - y| \leqslant B \sum |x_i - y_i|$$

This means that if P is a partition of $[a,b]$, then

$$A \sum_{ij} |x_j(t_i) - x_j(t_{i-1})| \leqslant \sum |x(t_i) - x(t_{i-1})| \leqslant B \sum_{ij} |x_j(t_i) - x_j(t_{i-1})|$$

So $A \sum V_P(x_i) \leqslant L_P(x) \leqslant B \sum V_P(x_i)$ gives the required result. $\qquad\square$

Example. *If f is a real-valued, monotonic and bounded function on $[a,b]$, then f has bounded variation, and one can verify that $V(f) = f(b) - f(a)$.*

Example. *If f is differentiable at every point, and f' is bounded, then f has bounded variation. The mean value theorem implies that if $|f'| \leqslant M$, then for all $x, y \in [a,b]$,*

$$|f(x) - f(y)| \leqslant M|x - y|$$

This implies that $V_P(f) \leqslant M(b - a)$ for all partitions P.

Example. *Set*

$$f(x) = \begin{cases} x^a \sin(x^{-b}) & : 0 < x \leqslant 1 \\ 0 & : x = 0 \end{cases}$$

Then f has bounded variation on $[0,1]$ if and only if $a > b$.

The next result says that a function with bounded variation is 'essentially' increasing. We define the **positive variation** of a real valued function f on $[a,b]$ to be

$$P_{[a,b]}(f) = \sup_P \sum_{f(t_i) \geqslant f(t_{i-1})} f(t_i) - f(t_{i-1})$$

The **negative variation** is

$$N_{[a,b]}(f) = \sup_P \sum_{f(t_i) \leqslant f(t_{i-1})} -[f(t_i) - f(t_{i-1})]$$

Note that for each partition P, the sums of the two values above add up to the variation with respect to the partition.

Lemma 3.12. *If f is real-valued and has bounded variation on $[a,b]$, then for all $a \leqslant x \leqslant b$,*

$$f(x) - f(a) = P_{[a,x]}(f) - N_{[a,x]}(f)$$
$$V(f) = P_{[a,b]}(f) + N_{[a,b]}(f)$$

Proof. Given ε, there exists a partition $a = t_0 < \cdots < t_n = x$ such that

$$\left| P_{[a,x]}(f) - \sum_{f(t_i) \geqslant f(t_{i-1})} f(t_i) - f(t_{i-1}) \right| < \varepsilon$$

$$\left| N_{[a,x]}(f) + \sum_{f(t_i) \leqslant f(t_{i-1})} f(t_i) - f(t_{i-1}) \right| < \varepsilon$$

It follows that

$$|f(x) - f(a) - [P_{[a,x]}(f) - N_{[a,x]}(f)]| < 2\varepsilon$$

and we can then take $\varepsilon \to 0$. The second identity follows the same way. $\qquad \square$

A real function f on $[a,b]$ has bounded variation if and only if f is the difference of two increasing bounded functions, because if f has bounded variation, then

$$f(x) = [f(a) + P_{[a,x]}(f)] - N_{[a,x]}(f)$$

is the difference of two bounded increasing functions. On the other hand, the difference of two bounded increasing functions is clearly of bounded variation. A complex function has bounded variation if and only if it is the linear combination of four increasing functions in each direction.

Theorem 3.13. *If f is a continuous function of bounded variation, then*

$$x \mapsto V_{[a,x]}(f) \quad x \mapsto V_{[x,b]}$$

are continuous functions.

Proof. $V_{[a,x]}(f)$ is an increasing functin of x, so for continuity on the left it suffices to prove that for each x and ε, there is $x_1 < x$ such that $V_{[a,x_1]}(f) \geqslant V_{[a,x]}(f) - \varepsilon$. If we consider a partition

$$P = \{a = t_0 < \cdots < t_n = x\}$$

where $|V_P(f) - V_{[a,x]}(f)| \leqslant \varepsilon$, then by continuity of f at x, there is $t_{n-1} < x_1 < x$ with $|f(x) - f(x_1)| < \varepsilon$, and then if we modify P to obtain Q by swapping t_n with x_1, we find

$$V_{[a,x_1]}(f) \geqslant V_Q(f) = V_P(f) - |f(x) - f(t_{n-1})| + |f(x_1) - f(t_{n-1})|$$
$$\geqslant V_P(f) - \varepsilon \geqslant V_{[a,x]}(f) - \varepsilon$$

A similar argument gives continuity on the right, and the continuity as the left bound of the interval changes. □

To obtain the differentiation theorem for functions of bounded variation, we require a lemma of F. Riesz.

Lemma 3.14 (Rising Sun lemma). *If f is real-valued and continuous on $\mathbf{R}$, and E is the set of points x where there exists $h > 0$ such that $f(x + h) > f(x)$, then, provided E is non-empty, it must be open, and can be written as a union of disjoint intervals (a_i, b_i), where $f(b_i) = f(a_i)$. If f is continuous on $[a,b]$, then E is still an open subset of $[a,b]$, and can be written as the disjoint union of countably many intervals, with $f(b_i) = f(a_i)$ except if $a_i = a$, where we can only conclude $f(a_i) \leqslant f(b_i)$.*

47

Theorem 3.15. *If f is an increasing, continuous function, then f is differentiable almost everywhere. That is,*

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

exists for almost every $x \in [a, b]$, f' is measurable, and

$$\int_a^b f'(x) \leqslant f(b) - f(a)$$

In particular, if f is bounded on $\mathbf{R}$, then f' is integrable on $\mathbf{R}$.

Proof. It suffices to assume that f is increasing, and we shall start by proving the theorem in the case where f is continuous. We define

$$\Delta_h(x) = \frac{f(x+h) - f(x)}{h}$$

and the four *Dini derivatives*

$$D_+(x) = \liminf_{h \downarrow 0} \Delta_h(x) \quad D^+(x) = \limsup_{h \downarrow 0} \Delta_h(x)$$

$$D_-(x) = \liminf_{h \uparrow 0} \Delta_h(x) \quad D^-(x) = \limsup_{h \uparrow 0} \Delta_h(x)$$

Clearly, we have $D_+ \leqslant D^+$ and $D_- \leqslant D^-$, It suffices to show $D^+(x) < \infty$ for almost every x, and $D^+(x) \leqslant D_-(x)$ for almost every x, because if we consider the function $g(x) = -g(-x)$, then we obtain $D^-(x) \leqslant D_+(x)$ for almost every x, so

$$D^+(x) \leqslant D_-(x) \leqslant D^-(x) \leqslant D_+(x) \leqslant D^+(x) < \infty$$

for almost every x, implying the derivative exists at x. For a fixed $\gamma > 0$, consider $E_\gamma = \{x : D^+(x) > \gamma\}$. TODO: FINISH THIS PROOF. $\qquad \square$

Even for increasing functions, we cannot extend the result above to an equality. This follows from the construction of the Cantor-Lebesgue function, which is a continuous function $f : [0, 1] \to [0, 1]$ with $f(0) = 0$, $f(1) = 1$, but with $f'(x) = 0$ almost everywhere. f has bounded variation, but

$$\int_a^b f'(x) = 0 \neq f(b) - f(a)$$

48

Consider the Cantor set $C = \bigcap C_k$, where C_k is the disjoint union of 2^k closed intervals. Set $f_0 = 0$, and $f_1(0) = 0$, $f_1(x) = 1/2$ on $[1/3, 2/3]$, $f_1(1) = 1$, and f linear between $[0, 1/3]$ and $[2/3, 1]$. Similarily, set $f_2(0) = 0$, $f_2(x) = 1/4$ on $[1/9, 2/9]$, $f_2(x) = 1/2$ on $[1/3, 2/3]$, $f_2(x) = 3/4$ on $[7/9, 8/9]$, and $f_2(1) = 1$. The functions f_i are increasing and cauchy in the uniform norm, so they converge to a continuous function f called the **Cantor function**. f is constant on each interval in the complement of the cantor set, so $f'(x) = 0$ almost everywhere. To obtain equality in the integral formula, we require additional conditions on our increasing functions.

3.6 Absolute Continuity

A function $f : [a, b] \to \mathbf{R}$ is **absolutely continuous** if for every $\varepsilon > 0$, there is $\delta > 0$ such that whenever $(a_1, b_1), \ldots, (a_n, b_n)$ are disjoint intervals with $\sum(b_i - a_i) < \delta$, $\sum |f(b_i) - f(a_i)| < \varepsilon$. It is easy to see from this that absolutely continuous functions must be uniformly continuous, and have bounded variation. Thus f has a decomposition into the difference of two continuous increasing functions, and one can see quite easily that these functions are also absolutely continuous. Most promising to us, if f is a function defined by $f(x) = \int_a^x g(t) \, dt$, where $g \in L^1[a, b]$, then f is absolutely continuous. This shows that absolute continuity is necessary in order to hope for the integral formula

$$\int_a^b f'(x) \, dx = f(b) - f(a)$$

One can verify that the Cantor function is *not* absolutely continuous, and therefore we do not have the integration equation above.

To prove the differentiation theorem, we require a covering estimate not unlike that used to prove the Lebesgue differentiation theorem. We say a collection of balls is a **Vitali covering** of a set E if for every $x \in E$ and every $\eta > 0$, there is a ball B in the cover containing x with $|B| < \eta$. Thus every point is covered by an arbitrary small ball.

Lemma 3.16. *If E is a set of finite measure, and $\{B_\alpha\}$ is a Vitali covering of E, then for any $\delta > 0$, we can find finitely many disjoint balls $B_1, \ldots, B_n$ in the covering such that*

$$\left| \bigcup B_i \right| = \sum |B_i| \geqslant |E| - \delta$$

Proof. Without loss of generality, assume $\delta \leqslant |E|$. By inner regularity, pick a compact set $K \subset E$ with $|K| \geqslant |E| - \delta/2$. Then K is covered by finitely many balls of radius less than η in the covering $\{B_\alpha\}$, and the elementary Vitali covering lemma gives a disjoint subcollection of balls $B_1, \ldots, B_{n_0}$ with

$$|K| \leqslant \left| \bigcup B_\alpha \right| \leqslant 3^d \sum |B_k|$$

so $\sum |B_k| \geqslant 3^{-d} |K|$. If $\sum |B_k| \geqslant |K| - \delta/2$, we're done. Otherwise, define $E_1 = K - \bigcup \overline{B_k}$. Then

$$|E_1| \geqslant |K| - \sum |\overline{B_k}| = |K| - \sum |B_k| > \delta/2$$

If we pick a compact set $K_1 \subset E_1$ with $|K_1| \geqslant \delta/2$, then if we remove all sets in the Vitali covering which intersect $B_1, \ldots, B_{n_0}$, then we still obtain a Vitali covering for K_1, and we can repeat the argument above to find a disjoint collection of open sets $B_1^1, \ldots, B_{n_1}^1$ with $\sum |B_k^1| \geqslant 3^{-d} |K_1|$. Then $\sum |B_k| + \sum |B_k^1| \geqslant 2(3^{-d} \delta)$. If $\sum |B_k| + \sum |B_k^1| < |K| - \delta/2$, we repeat the same process, finding a disjoint family for $K_2 \subset E_2$, where $E_2 = K_1 - \bigcup \overline{B_k^1}$. If this process repeats itself k times, then we obtain a family of open sets with total measure greater than or equal to $k(3^{-d} \delta)$. But then if we eventually have $k \geqslant (|E| - \delta) 3^d / \delta$, then the family of open sets satisfies the requirements of the theorem. $\square$

Corollary 3.17. *We can arrange the choice of balls such that*

$$\left| E - \bigcup B_i \right| < 2\delta$$

Proof. Let $E \subset U$, where U is an open set with $|U - E| < \delta$. In the algorithm above, we may consider all balls in the Vitali covering as contained within U. But then

$$\left| E - \bigcup B_i \right| \leqslant |U| - \sum |B_i| = |U| - \bigcup E_i \leqslant \delta + |E| - \sum |B_i| \leqslant 2\delta$$

and this gives the required bound. $\square$

Theorem 3.18. *If $f : [a, b] \to \mathbf{R}$ is absolutely continuous, then f' exists almost everywhere, and if $f'(x) = 0$ almost surely, then f is constant.*

Proof. It suffices to prove that $f(a) = f(b)$, since we can then apply the theorem on any subinterval. Let $E = \{x \in (a,b) : f'(x) = 0\}$. Then $|E| = b - a$. Fix $\varepsilon > 0$. Since for each $x \in E$, we have

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = 0$$

This implies that the family of intervals (x,y) such that the inequality $|f(y) - f(x)| \leqslant \varepsilon(y - x)$ holds forms a Vitali covering of E, and we may therefore select a family of disjoint intervals $I_i = (x_i, y_i)$ with

$$\sum |I_i| \geqslant |E| - \delta = (b - a) - \delta$$

But $|f(y_i) - f(x_i)| \leqslant \varepsilon(y_i - x_i)$, so we conclude

$$\sum |f(y_i) - f(x_i)| \leqslant \varepsilon(b - a)$$

The complement of I_i is a union of intervals $J_i = (x'_i, y'_i)$ of total length $\leqslant \delta$. Applying the absolute continuity of f, we conclude

$$\sum |f(y'_i) - f(x'_i)| \leqslant \varepsilon$$

so applying the triangle inequality,

$$|f(b) - f(a)| \leqslant \sum |f(y'_i) - f(x'_i)| + \sum |f(y_i) - f(x_i)| \leqslant \varepsilon(b - a + 1)$$

We can then let $\varepsilon \to 0$ to obtain equality. $\square$

Theorem 3.19. *Suppose f is absolutely continuous on $[a,b]$. Then f' exists almost every and is integrable, and*

$$f(b) - f(a) = \int_a^b f'(y) \, dy$$

so the fundamental theorem of calculus holds everywhere. Conversely, if $f \in L^1[a,b]$, then there is an absolutely continuous function g with $g' = f$ almost everywhere.

Proof. Since f is absolutely continuous, we can write f as the difference of two continuous increasing functions on $[a,b]$, and this easily implies f is differentiable almost everywhere and is integrable on $[a,b]$. If $g(x) = \int_a^x f'(x)$, then g is absolutely continuous, hence $g - f$ is also absolutely continuous. But we know that $(g - f)' = g' - f' = 0$ almost everywhere, so the last theorem implies that g differs from f by a constant. Since $g(a) = 0$, $g(x) = f(x) - f(a)$. The converse was proved exactly in our understanding of differentiating integrals. $\square$

3.7 Differentiability of Jump Functions

We now consider the differentiability of not necessarily continuous mono-
tonic functions. Set f to be an increasing function on $[a,b]$, which we
may assume to be bounded. Then the left and right limits of f exist at
every point, which we will denote by $f(x-)$ and $f(x+)$. Of course, we
have $f(x-) \leqslant f(x) \leqslant f(x+)$. If there is a discontinuity, this means we are
forced to have a 'jump discontinuity' where f skips an interval. This im-
plies that f can only have countably many such discontinuities, because
a family of disjoint intervals on $\mathbf{R}$ is at most countable. Now define the
jump function $\Delta f(x) = f(x^+) - f(x-)$, with $\theta(x) \in [0,1]$ defined such that
$f(x_n) = f(x_n-) + \theta(x)\Delta f(x)$. If we define the functions

$$j_y(x) = \begin{cases} 0 & : x < y \\ \theta(y) & : x = y \\ 1 & x > y \end{cases}$$

then we can define the **jump function** associated with f by

$$J(x) = \sum_x \Delta f(x) j_n(x)$$

Since f is bounded on $[a,b]$, we make the final observation that

$$\sum_{x \in [a,b]} \Delta f(x) \leqslant f(b) - f(a) < \infty$$

so the series defining J converges absolutely and uniformly.

Lemma 3.20. *If f is increasing and bounded on $[a,b]$, then J is discontinuous
precisely at the values x with $\Delta f(x) \neq 0$ with $\Delta J(x) = \Delta f(x)$. The function
$f - J$ is continuous and increasing.*

Proof. If x is a continuity point of f, then j_y is continuous at x, and hence,
because $\sum_y \Delta f(y) j_y(x) \to J(x)$ uniformly, so we conclude that J is continu-
ous at x. On the other hand, for each y, $j_y(y-) = 0$ and $j_y(y+) = 1$, and if
we label the points of discontinuity of f by $x_1, x_2, \ldots$, then

$$J(x) = \sum_{i=1}^{k} \Delta f(x_i) j_{x_i} + \sum_{i=k+1}^{\infty} \Delta f(x_i) j_{x_i}$$

The right hand partial sums are continuous at x_k, whereas the left hand sum has a jump discontinuity of the same order as f at x_k, we conclude J also has this discontinuity. But this means that

$$(f - J)(x_k+) - (f - J)(x_k-) = 0$$

so $f - J$ is continuous at every point. $f - J$ is increasing because of the inequality

$$J(y) - J(x) \leqslant \sum_{x < x_n \leqslant y} \alpha_n \leqslant f(y) - f(x)$$

which follows because J is just the sum of jump discontinuities, and the right hand side because f can decrease and increase outside of the jump discontinuities. $\qquad \square$

Since $f - J$ is continuous and increasing, it is differentiable almost everywhere. It therefore remains to analyze the differentiability of the jump function J.

Theorem 3.21. *J' exists and vanishes almost everywhere.*

Proof. Fix $\varepsilon > 0$, and consider

$$E = \left\{ x \in [a,b] : \limsup_{h \to 0} \frac{J(x+h) - J(x)}{h} > \varepsilon \right\}$$

Then E is measurable, because we can take the $\limsup$ over rational numbers because J is increasing. We want to show it has measure zero. Suppose $\delta = |E|$. Consider $\eta > 0$ to be chosen later, and find n such that $\sum_{k=n}^{\infty} \alpha_k < \eta$. Write

$$J_0(x) = \sum_{n > N} \alpha_n j_n$$

Then $J_0(b) - J_0(a) < \eta$. Now E differs from the set

$$E' = \left\{ x \in [a,b] : \limsup_{h \to 0} \frac{J_0(x+h) - J_0(x)}{h} > \varepsilon \right\}$$

by finitely many points. Using inner regularity, find a compact set $K \subset E'$ with $|K| \geqslant \delta/2$. For each $x \in K$, we can find intervals (α_x, β_x) upon which $J_0(\beta_x) - J_0(\alpha_x) \geqslant \varepsilon |\beta_x - \alpha_x|$. But applying the elementary Vitali covering

53

lemma, we can find a disjoint family of such intervals with $\sum (\beta_{x_i} - \alpha_{x_i}) \geqslant |K|/3 \geqslant \delta/6$. But now we find

$$J_0(b) - J_0(a) \geqslant \sum J_0(\beta_{x_i}) - J_0(\alpha_{x_i}) \geqslant \varepsilon \delta/6$$

This means $\delta \leqslant 6\eta/\varepsilon$, and by letting $\eta \to 0$, we can conclude $\delta = 0$. $\qquad \square$

3.8 Application: Rectifiable Curves

We now consider the validity of the length formula

$$L = \int_a^b (x'(t)^2 + y'(t)^2)^{1/2} \, dt$$

where L is the length of the curve parameterized by (x, y) on $[a, b]$. We cannot always expect this formula to hold, because if x and y are both the Cantor devil staircase function, then the formula above gives a length of zero, whereas we know the curve traces a line between 0 and 1, hence has length at least $\sqrt{2}$ (this value is actually exact).

Theorem 3.22. *If a curve is parameterized by absolutely continuous functions x and y on $[a, b]$, then the curve is rectifiable, and has length*

$$\int_a^b (x'(t) + y'(t))^{1/2} \, dt$$

Chapter 4

Measure Extension

To construct the Lebesgue measure, we began with an intuitive notion of area over the set of intervals on the real line, and then extended the notion of length to a bigger class of sets, the measurable sets. This chapter attempts to generalize this strategy to work on arbitrary measure spaces. The idea is to take a countably additive measure defined on a family of sets satisfying weaker conditions than that of a σ algebra, and then to extend this measure to the σ algebra this family generates. If X is a set, a family Σ_0 of subsets of X is known as an **algebra** if it is closed under complements and finite unions (it is an algebra under $\mathbf{F}_2$, if we consider intersection as multiplication, and the *symmetric difference $A \triangle B = (A - B) \cup (B - A)$* as addition). A σ algebra is just an algebra where we can take countable unions instead of finite unions, and this is the natural place to study the analytical theory of measure, because we often want to apply limiting estimates.

    The way we constructed the Lebesgue measure was to take a family of elementary sets (disjoint unions of boxes), which we intuitively know the measure of, and then define the measure of arbitrary sets by approximating them from the outside by these elementary sets. The family of elementary sets formed an algebra, but not a σ algebra, and thus the extension was necessary to construct the satisfying theory of the Lebesgue measure. To generalize this construction, we take a algebra Σ_0, on which a sense of measure is defined, and try to extend this to a σ algebra containing Σ_0. The intuitive notion of measure is an additive function $\mu_0 : \Sigma_0 \to [0, \infty]$ such that $\mu_0(\varnothing) = 0$, which is countably additive when countable additivity is well defined in Σ_0. We call such functions **pre-measurable**. Alternatively,

a premeasure is a monotone, countably subadditive function where defined, or a monotone, upward continuous function. Being a premeasure is the weakest condition we could put on a function which is a candidate for extension to a full measure, but Caratheodory proved that the condition is sufficient for extension.

To obtain an extension, we employ the same ideas used in the construction of the Lebesgue measure. In that situation, we started with a simple algebra of sets (unions of boxes), and constructed the upper approximations

$$\mu^*(S) = \inf\left\{\sum_{k=0}^{\infty} \mu(A_k) : A_k \in \Sigma_0, S \subset \bigcup A_k\right\}$$

which are defined for any subset of $\mathbf{R}^d$. This function is monotone, countably σ subadditive, and $\mu^*(\varnothing) = 0$. On any set X, we call a function μ^* on X of this form an **exterior measure**. One of Caratheodory's major contribution to the theory of measure is to notice that a set E is Lebesgue measurable if for any subset A of $\mathbf{R}^d$,

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

For a general exterior measure, we say a set E is **Caratheadory measurable** if it satisfies every equation of the form above. Since μ^* is countably subadditive, it suffices to verify that $\mu^*(A)$ upper bounds the sum of the other two sets. This implies every set of exterior measure zero is Caratheadory measurable.

Lemma 4.1. *The set of Caratheadory measurable sets forms a σ algebra, and μ^* is a measure on this family.*

Proof. Clearly $\varnothing$ and X are Caratheadory measurable, and the symmetry of the statement implies the class of Caratheadory measurable sets is closed under complements. Next, we show measurable sets are closed under finite unions of disjoint sets, and that μ^* is finitely additive. If E_1 and E_2 are disjoint and measurable, and A is arbitrary, then

$$\begin{aligned}
\mu^*(A) &= \mu^*(E_1 \cap A) + \mu^*(E_1^c \cap A) \\
&= \mu^*(E_1 \cap E_2 \cap A) + \mu^*(E_1 \cap E_2^c \cap A) \\
&\quad + \mu^*(E_1^c \cap E_2 \cap A) + \mu^*(E_1^c \cap E_2^c \cap A) \\
&\geqslant \mu^*((E_1 \cup E_2) \cap A) + \mu^*((E_1 \cup E_2)^c \cap A)
\end{aligned}$$

This shows measurability, and

$$\mu^*(E_1 \cup E_2) = \mu^*((E_1 \cup E_2) \cap E_2) + \mu^*((E_1 \cup E_2) \cap E_2^c) = \mu^*(E_2) + \mu^*(E_1)$$

Finally, we need to show that the class of Caratheadory measurable sets is closed under countable unions of disjoint sets, and μ^* is countably additive on this family. Consider measurable sets $E_1, E_2, \ldots$, and let

$$K_n = \bigcup_{k \leqslant n} E_k \quad K = \bigcup E_k$$

Each K_n is measurable, and $\mu^*(K_n) = \sum_{k \leqslant n} \mu^*(E_k)$. If A is arbitrary, by induction, we find

$$\mu^*(K_n \cap A) = \sum_{k=1}^n \mu^*(E_n + A)$$

and thus

$$\mu^*(A) = \mu^*(K_n \cap A) + \mu^*(K_n^c \cap A) \geqslant \sum_{k=1}^n \mu^*(E_n \cap A) + \mu^*(K^c \cap A)$$

Letting $n \to \infty$ gives that

$$\mu^*(A) \geqslant \sum_{n=1}^\infty \mu^*(E_n \cap A) + \mu^*(K^c \cap A) \geqslant \mu^*(K \cap A) + \mu^*(K^c \cap A)$$

This shows K is measurable, and taking $A = K$ in the calculation with an equality gives countable additivity. $\qquad\square$

Caratheodory's theorem generates measure spaces that are **complete**, in the sense that any set contained within a set of measure zero also has measure zero. The lemma leads to a quick proof of the extension theorem.

Theorem 4.2 (Caratheodory's Extension Theorem)**.** *Every premeasure on an algebra Σ_0 can be extended uniquely to a measure on the σ algebra Σ that Σ_0 generates.*

Proof. Given μ_0, define the exterior measure

$$\mu^*(S) = \inf \left\{ \sum_{k=1}^\infty \mu(A_k) : A_k \in \Sigma_0, S \subset \bigcup A_k \right\}$$

57

It is easy to verify this is an exterior measure. Monotonicity is easy, and to prove countable subadditivity, note that if $S_1, S_2, \dots$ are a family of sets with union S and $\sum \mu^*(S_i) < \infty$, then we can find sets $A_1^i, A_2^i, \dots, A_{k_i}^i$ covering S_i with

$$\mu^*(S_i) + \varepsilon/2^n \geq \sum \mu(A_k^i)$$

Then the union over all A_j^i is a cover of S, and then

$$\mu^*(S) \leq \sum \mu(A_j^i) = \sum \mu^*(S_i) + \varepsilon/2^n \leq \sum \mu^*(S_i) + \varepsilon$$

We can then let $\varepsilon \to 0$. For any $S \in \Sigma_0$, $\mu^*(S) = \mu * (S)$, because if $S \subset \bigcup A_k$, where $A_k \in \Sigma_0$, then by monotonicity, $\mu(A_k \cap S) \leq \mu(A_k)$, and because

$$S = \lim_{n \to \infty} S \cap \bigcup_{k \leq n} A_k$$

we conclude

$$\mu(S) = \lim_{n \to \infty} \mu \left(S \cap \bigcup_{k \leq n} A_k \right) \leq \lim_{n \to \infty} \mu \left(\bigcup_{k \leq n} A_k \right) \leq \sum_{k=0}^{\infty} \mu(A_k)$$

This shows in particular that $\mu^*(\varnothing) = \mu(\varnothing) = 0$, so μ^* is an exterior measure. Because of Caratheadory's lemma, all that remains is to show that all elements E of Σ_0 are Caratheadory measurable. Given A, first assume $\mu^*(A) < \infty$. Then we can find a family of sets $A_1, A_2, \dots \in \Sigma_0$ with $\sum \mu(A_k) \leq \mu^*(A) + \varepsilon$. But then

$$\sum \mu(A_k) = \sum \mu(A_k \cap E) + \mu(A_k \cap E^c)$$

The sets $A_k \cap E$ cover $A \cap E$, and the sets $A_k \cap E^c$ cover $A \cap E^c$, so

$$\mu^*(A \cap E) + \mu^*(A \cap E^c) \leq \sum \mu(A_k \cap E) + \mu(A_k \cap E^c) = \sum \mu(A_k) \leq \mu^*(A) + \varepsilon$$

We then let $\varepsilon \to 0$. If $\mu^*(A) = \infty$, $\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A \cap E^c)$ is obvious. This shows that the σ algebra of Caratheadory measurable sets contains Σ_0, and the measure μ^* defined on it extends μ. Note, however, that there may be many more Caratheadory measurable sets than the subsets of Σ, but we can solve the problem by restriction. $\square$

Corollary 4.3. *If X is σ-finite, the extension is unique on Σ.*

Proof. Let μ be another extension of μ_0 to Σ rather than μ^*. Choose some $E \in \Sigma$, and fix $\varepsilon > 0$. For any countable covering of E by elements $A_1, \cdots \in \Sigma_0$,

$$\mu(A) \leqslant \sum \mu(A_k) = \sum \mu_0(A_k)$$

It follows that $\mu \leqslant \mu^*$. By symmetry, $\mu(A^c) \leqslant \mu^*(A^c)$. But then $\mu(A) + \mu(A^c) = \mu(X) = \mu^*(X)$, and assuming $\mu(X) < \infty$, we conclude

$$\mu(A) = \mu(X) - \mu(A^c) \geqslant \mu^*(X) - \mu^*(A^c) = \mu^*(A)$$

In general, if $X = \lim X_n$, where X_n are finite measure sets, we can apply the theorem to $E \cap X_n$, and then take limits. $\qquad \square$

For the purposes of using the Caratheodory extension theorem, is is often useful to use the following method of generating a premeasure. A **semialgebra** on a set X is a family subsets containing the empty set, closed under intersections, and such that if A is in the family, then A^c can be broken into finitely many disjoint sets in the family. The algebra generated by a semialgebra is then the family of all disjoint unions of elements of the semialgebra. A content is a $[0, \infty]$ valued function μ defined on a semialgebra with $\mu(\varnothing) = 0$ and $\mu(A \cup B) = \mu(A) + \mu(B)$ for any two disjoint A, B in the semialgebra. Any content can then be extended uniquely to a premeasure on the algebra generated by the semialgebra by taking sums of disjoint unions.

Example. *Let us consider the construction of the Lebesgue measure in the form of the Caratheodory extension theorem. Consider the semialgebra of intervals of the form $[a, b)$ and $(-\infty, b)$, where we allow open brackets to take the value ∞. Define*

$$\mu[a, b) = b - a \quad \mu(-\infty, b) = \infty$$

This extends to the algebra of disjoint unions of intervals. It is nontrivial to verify that μ is countably additive, which we partook in the first chapter, but once this is done, the Carathedory extension theorem provides a black box to extend μ to the Lebesgue measure on $\mathbf{R}$. More generally, we can consider the semialgebra of boxes with open and closed boundaries in $\mathbf{R}^d$, and the resulting measure we construct from this family will be the Lebesgue measure in $\mathbf{R}^d$.

Chapter 5

Product Spaces

Let X and Y be measure spaces, with respective σ algebras Σ and Π. The idea of product spaces is to define a natural measure space structure on $X \times Y$. We define an **elementary rectangle** in $X \times Y$ to be a set of the form $E \times F$, where $E \in \Sigma$ and $F \in \Pi$. We shall find the natural σ algebra on $X \times Y$ is the one generated by elementary rectangles, and we denote this σ algebra by $\Sigma \otimes \Pi$ (it isn't *exactly* the tensor product if we view the σ algebras as algebras over $\mathbf{F}_2$, but it has analogous properties to the tensor product). Since

$$(E_1 \times F_1) \cap (E_2 \cap F_2) = (E_1 \cap E_2) \times (F_1 \cap F_2)$$

we conclude the family of elementary rectangles is a π system. Given a set $E \subset X \times Y$, we define the *sections*

$$E_x = \{y : (x, y) \in E\} \quad E^y = \{x : (x, y) \in E\}$$

These sections preserve measurability.

Lemma 5.1. *If E is a measurable subset of $X \times Y$, then E_x and E^y are measurable subsets of X and Y.*

Proof. Fix $y \in Y$. If $E \times F$ is an elementary rectangle, then if $y \in F$, $(E \times F)^y = E$, and if $y \notin F$, $(E \times F)^y = \varnothing$, and these sets are both measurable in X. If E^y and F^y are measurable subsets of Y, then $(F - E)^y = F^y - E^y$ is measurable. If $E = \bigcup E_i$ is the countable union of sets with E_i^y measurable, then $E^y = \bigcup E_i^y$ is also measurable. We conclude the class of sets E with E^y measurable is a λ system containing the π system of elementary rectangles, so

the π-λ theorem guarantees the theorem is true for all measurable subsets of $X \times Y$. The symmetry of the situation shows that E_x is measurable for all measurable sets. $\square$

It follows that if $f : X \times Y \to Z$ is a measurable function, then the functions $f^y : x \mapsto f(x,y)$ and $f_x : y \mapsto f(x,y)$ are also measurable, because they are the composition of f with the maps $x \mapsto (x,y)$, $y \mapsto (x,y)$, and these project backward onto the sections of measurable sets in $X \times Y$.

Remark: The $\pi - \lambda$ theorem isn't often used in measure theory, where it is often substituted for the monotone class theorem. For completeness, we also discuss this method. We say a family of sets Σ is a **monotone class** if it is closed under upward limits: If $E_1 \subset E_2 \subset \cdots \in \Sigma$, then $\lim E_i \in \Sigma$, and if $E_1 \supset E_2 \supset \cdots \in \Sigma$, then $\lim E_i \in \Sigma$.

Theorem 5.2. *The smallest monotone class containing an algebra is the σ algebra generated by the algebra.*

Proof. If Σ is an algebra, then it is a π system. The $\pi - \lambda$ theorem says that the smallest λ system containing Σ is the σ algebra containing the π system. But if Σ_* is a monotone class containing Σ, then it is almost a λ system, except that we may not have $B - A \in \Sigma_*$ if $A \subset B$ are both members of Σ_*. We show that this does hold if Σ_* contains Σ. For any set E, let

$$X_E = \{F : E - (E \cap F) \in \Sigma_*\} \quad Y_E = \{F : F - (E \cap F) \in \Sigma_*\}$$

For any set E, X_E and Y_E are monotone sets, which essentially follows because if $F_i \to F$ (upward or downward), then $E - (E \cap F_i) \to E - (E \cap F)$ and $F_i - (E \cap F_i) \to F - (E \cap F)$. Now if $E \in \Sigma$, then X_E and Y_E both contain Σ, so $\Sigma_* \subset X_E, Y_E$. But this means if $E \in \Sigma_*$, $F \in \Sigma$, then $E \in X_F$, so $F - (E \cap F) \in \Sigma_*$. This implies that $F \in Y_E$. We conclude if $E \in \Sigma_*$, Y_E is a monotone set containing Σ, so $\Sigma_* \subset Y_E$. This shows that Σ_* is a λ system, and therefore that Σ_* contains the σ algebra generated by Σ. $\square$

We can use the monotone class theorem to understand the structure of the product σ algebra $\Sigma \otimes \Pi$ in more detail. We know that the set of all elementary rectangles $E \times F$ is closed under intersection, but isn't closed under the other algebraic operations which would allow us to use the monotone class theorem. Nonetheless, if we consider the class of all **elementary sets**, which are subsets of $X \times Y$ formed from *disjoint unions*

61

of elementary rectangles, then this does form an algebra. The essential reason for this is that

$$(E_1 \times F_1) - (E_2 \times F_2) = (E_1 - E_2) \times F_1 \cup (E_1 \cap E_2) \times (F_1 - F_2)$$

$$(E \times F)^c = E^c \times F \cup E \times F^c \cup E^c \times F^c$$

$$(E_1 \times F_1) \cup (E_2 \times F_2) = (E_1 \times F_1) \cup [(E_2 \times F_2) - (E_1 \times F_1)]$$

and the fact that algebraic operations distribute themselves over unions. Since the family of elementary sets *does* form an algebra, we conclude that $\Sigma \otimes \Pi$ is the smallest monotone class which contains the family of elementary sets. It is of course also useful to note that $\Sigma \otimes \Pi$ is the smalles the λ system containing the π system of elementary rectangles.

Now suppose we have a positive measure μ defined on Σ, and a positive measure λ on Π. It makes sense to define a measure $\mu \otimes \lambda$ on $\Sigma \otimes \Pi$, such that $(\mu \otimes \lambda)(E \times F) = \mu(E)\lambda(F)$, just like we would define the area of a rectangle based on the length of its sides. If $E_1 \times F_1$ and $E_2 \times F_2$ are two disjoint elementary rectangles whose union is also an elementary rectangle $E_3 \times F_3$, then one finds that either E_1 is disjoint from E_2, and their union is E_3, and $F_1 = F_2 = F_3$, or $E_1 = E_2 = E_3$ and F_1 is disjoint from F_2. In the first case, one find

$$(\mu \otimes \lambda)(E_1 \times F_1) + (\mu \otimes \lambda)(E_2 \times F_2) = [\mu(E_1) + \mu(E_2)]\lambda(F_3) = \mu(E_3)\lambda(F_3)$$

and in the second

$$(\mu \otimes \lambda)(E_1 \times F_1) + (\mu \otimes \lambda)(E_2 \times F_2) = \mu(E_3)[\lambda(F_1) + \lambda(F_2)] = \mu(E_3)\lambda(F_3)$$

so $(\lambda \otimes \mu)$ is a content, and thus extends to an additive measure on the family of elementary sets in $\Sigma \otimes \Pi$. Now if a rectangle $E \times F$ is a disjoint, countable union of disjoint rectangles $E_i \times F_i$, we claim that

$$(\mu \otimes \lambda)(E \times F) = \sum (\mu \otimes \lambda)(E_i \times F_i) = \sum \mu(E_i)\lambda(F_i)$$

Fixing $x \in E$, for each $y \in F$ there is exactly one index i with $x \in E_i$ and $y \in F_i$. The countable additivity of λ therefore guarantee that

$$\chi_E(x)\lambda(F) = \sum_{j=1}^{\infty} \chi_{E_j}(x)\lambda(F_j)$$

Applying the monotone convergence theorem over Y, we conclude that

$$\mu(E)\lambda(F) = \int \chi_E(x)\lambda(F)d\mu(x) = \sum_{j=1}^{\infty} \int \chi_{E_j}(x)\lambda(F_j) = \sum_{j=1}^{\infty} \mu(E_j)\lambda(F_j)$$

and this is exactly countable additivity. This shows that $(\lambda \otimes \mu)$ is a *pre-measure* on the class of elementary sets, and the Caratheadory extension theorem guarantees that $\lambda \otimes \mu$ extends to a measure on $\Sigma \otimes \Pi$. Provided that we are working over a σ finite space, this extension is unique, a fact that will become increasingly more important when we analyze the integration theory of product spaces, where we often have to assume we are working over σ finite spaces.

Theorem 5.3. *If X and Y are σ finite spaces, and $E \in \Sigma \otimes \Pi$, then*

$$\int \lambda(E_x)d\mu(x) = (\lambda \otimes \mu)(E) = \int \mu(E^y)d\lambda(y)$$

Proof. The theorem is trivially true if E is an elementary rectangle. The class of sets for which this theorem holds is also closed monotonely upward, because if $E_1 \subset E_2 \subset \cdots \to E$, then $\lambda((E_i)_x) \to \lambda(E_x)$ monotonely upward, and $\mu(E_i^y) \to \mu(E^y)$ monotonely upward, so the monotone convergence theorem guarantees that if the theorem holds for the E_i, it also holds for E. The hard part of the theorem is showing that if $E_1 \supset E_2 \supset \cdots \to E$, and E_i satisfies the properties of the theorem, then E satisfies the properties of the theorem. Assume first that X and Y are finite measure spaces. Then we can apply the dominated convergence theorem downward, because constant functions are in L^1, and $\lambda(E_x) \leqslant \lambda(Y)$, $\mu(E^y) \leqslant \mu(X)$, and this shows that the class of sets for which the theorem holds is monotone, and contains all elementary rectangles, hence containing all elements of $\Sigma \otimes \Pi$. To obtain the theorem for σ finite measure spaces, we apply the theorem for any set $E \in \Sigma \otimes \Pi$ to conclude that if $X_i \subset X$ and $Y_i \subset Y$ have finite measure, then

$$\int_{X_i} \lambda(E_x \times Y_i) = (\lambda \otimes \mu)(E \cap (X_i \times Y_i)) = \int_{Y_i} \mu(E^y \times X_i)d\lambda(y)$$

Now we let $X_i \to X$, $Y_i \to Y$, and apply monotone convergence. $\qquad\square$

Theorem 5.4 (Tonelli). *If X and Y are σ finite, and f is a non-negative measurable function, then*

$$\int\int f^y(x)d\mu(x)d\lambda(y) = \int f(x,y)d(\lambda\otimes\mu)(x,y) = \int\int f_x(y)d\lambda(y)d\mu(x)$$

Proof. We have already proven this theorem in the case that f is the indicator function of some measurable subset of $X \times Y$. If f and g satisfy the theorem, then $\alpha f + \beta g$ also satisfy the theorem, for $\alpha, \beta \geqslant 0$. The monotone convergence theorem guarantees the theorem is true for all monotone upward limits of functions for which the theorem is true. But this shows that the theorem is true for all non-negative measurable functions, which are the monotone pointwise limit of simple functions. $\qquad\square$

Corollary 5.5. *If f is a complex measurable function on X × Y, where X and Y are both σ finite, then f is in $L^1(X \times Y)$ if and only if the common value of*

$$\int_Y\int_X |f(x,y)|d\mu(x)d\lambda(y) = \int_X\int_Y |f(x,y)|d\lambda(y)d\mu(x)$$

is finite, and this value is equal to $\|f\|_1$.

Theorem 5.6 (Fubini). *If $f \in L^1(X \times Y)$, then $f_x \in L^1(Y)$ for almost all x, $f^y \in L^1(X)$ for almost all y, and*

$$\int\int f_x(y)d\lambda(y)d\mu(x) = \int f(x,y)d(\mu\otimes\lambda)(x,y) = \int\int f^y(x)d\mu(x)d\lambda(y)$$

Proof. Without loss of generality, assume f is real valued. Then

$$\int_X\int_Y |f(x,y)|d\lambda(y)d\mu(x) < \infty$$

Hence we conclude that for almost all y,

$$\int_Y |f(x,y)|d\lambda(y) = \int_Y |f_x(y)|d\lambda(y) < \infty$$

The same process given in the other direction shows that $|f^y| \in L^1$ for almost all y. To obtain Fubini's integral formula, we just write $f = f^+ - f^-$, and then apply Tonelli's theorem independently to f^+ and f^-. $\qquad\square$

64

5.1 Completion of Product Measures

Just because Σ and Π are complete σ algebras with respect to μ and λ, does not imply that $\Sigma \otimes \Pi$ is complete with respect to $\mu \otimes \lambda$. For instance, in $\mathbf{R}^2$, if E is a non-measurable subset of $\mathbf{R}$, then $E \times [0,1]$ is not measurable with respect to the twofold product of the Lebesgue measure on $\mathbf{R}$, yet it is measurable with respect to the products completion, because it has exterior measure zero. On the other hand, $E \times [0,1]$ is *Lebesgue* measurable.

Theorem 5.7. *The Lebesgue measure on $\mathbf{R}^{n+m}$ is the completion of the product of the Lebesgue measures on $\mathbf{R}^d$ and $\mathbf{R}^m$.*

Proof. It is easy to see that the Lebesgue measure on $\mathbf{R}^{n+m}$ agrees with the product of the Lebesgue measure $\lambda^n \otimes \lambda^m$ on any set $E \times F$, where E and F are disjoint unions of rectangles in $\mathbf{R}^d$ and $\mathbf{R}^m$. But this is a π system generating the Borel σ algebra of $\mathbf{R}^{n+m}$, so the Lebesgue measure on $\mathbf{R}^{n+m}$ agrees with $\lambda^n \otimes \lambda^m$ on any Borel measurable subset of $\mathbf{R}^{n+m}$. But the Lebesgue measurable subsets of $\mathbf{R}^{n+m}$ are exactly those obtained from the completion of the corresponding measure on the Borel measurable subsets, so this shows every Lebesgue measurable subset is measurable with respect to the completion of the product algebra and vice versa, and the measures agree here. □

Fubini's theorem takes essentially the same form in the complete measure spaces, except for a slight variation.

Theorem 5.8. *If X and Y are complete σ finite measure spaces with σ algebras Σ and Π, and $(\Sigma \otimes \Pi)^*$ is the completion of the product measure on $X \times Y$. If f is $(\Sigma \otimes \Pi)^*$ measurable, then f_x is Σ measurable for almost all x, f^y is Π measurable for almost all y, but the integrals*

$$\int \int f(x,y) d\mu(x) d\lambda(y) \quad \int \int f(x,y) d\lambda(y) d\mu(x)$$

are still well defined, since $\int f(x,y) d\mu(x)$ and $\int f(x,y) d\lambda(y)$ are well defined almost everywhere. Fubini's theorem continues to hold in this circumstance.

We prove this using the two lemmas below.

Lemma 5.9. *If μ is a positive measure on Σ, and f is Σ^* measurable, then there is a Σ measurable function g with $f = g$ almost everywhere.*

65

Proof. Suppose $f \geqslant 0$. There there are non-negative Σ^* measurable simple functions $s_0 \leqslant s_1 \leqslant \ldots$ converging to f. Then $f = \sum(s_{n+1} - s_n)$. Since $s_{n+1} - s_n$ is a finite combination of characteristic functions, it follows that

$$f(x) = \sum c_i \chi_{E_i}(x)$$

with $c_i > 0$. We know that there are Σ measurable subsets $A_i \subset E_i \subset B_i$ with $\mu(B_i - A_i) = 0$. The function

$$g(x) = \sum c_i \chi_{A_i}(x)$$

is Σ measurable, and $f = g$ except on

$$\bigcup \{E_i - A_i\}$$

which is a countable union of subsets of measure zero. The general case is now immediate. $\square$

Lemma 5.10. *If f is $(\Sigma \otimes \Pi)^*$ measurable functions on $X \times Y$ such that $f = 0$ almost everywhere with respect to $\mu \otimes \pi$, then for almost all $x \in X$, $f(x,y) = 0$ for almost all y. In particular, f_x is Σ measurable for almost all $x \in X$. Similar results hold for f^y.*

Proof. Let $E = \{(x,y) : f(x,y) \neq 0\}$. Then we know $(\mu \otimes \lambda)(E) = 0$, so there is $F \in \Sigma \otimes \Pi$ with $E \subset F$, and $(\mu \otimes \lambda)(F) = 0$. By Fubini's theorem,

$$\int_X \int_Y \chi_F(x,y) d\lambda(y) d\mu(x) = \int_X \lambda(F_x) d\mu(x) = 0$$

This means that for almost all x, $\lambda(F_x) = 0$, and since $E_x \subset F_x$, this implies $E_x \in \Pi$, because Π is complete, and $\lambda(E_x) = 0$. If $y \notin E_x$, then $f(x,y) = 0$, so $f(x,y) = 0$ for almost all y. $\square$

It is useful sometimes to note that $(\Sigma \otimes \Pi)^* = (\Sigma^* \otimes \Pi^*)^*$. To see this, it suffices to show that if $E \in \Sigma^*$ and $F \in \Pi^*$, then $E \times F \in (\Sigma \otimes \Pi)^*$. We note there are $E_1, E_2 \in \Sigma$, $F_1, F_2 \in \Pi$ with $E_1 \subset E \subset E_2$, $F_1 \subset F \subset F_2$, and $\mu(E_2 - E_1) = \lambda(F_2 - F_1) = 0$, and since

$$E_2 \times F_2 - E_1 \times F_1 = (E_2 - E_1) \times (F_2 \cap F_1)$$
$$\cup (E_2 \cap E_1) \times (F_2 - F_1)$$
$$\cup (E_2 - E_1) \times (F_2 - F_1)$$

is the union of three sets with negligable measure in $\mu \otimes \lambda$.

5.2 Integration in Polar Coordinates

It is well known that the space $\mathbf{R}^d - \{0\}$ can be uniquely expanded into polar coordinates in $\mathbf{R}^+ \times S^{d-1}$, by writing $x = r\hat{x}$, where $r = |x|$ and $\hat{x} = x/|x|$. In this section we argue that

$$\int_{\mathbf{R}^d} f(x)\, dx = \int_{S^{d-1}} \int_0^\infty f(r\hat{x}) r^{d-1} dr d\hat{x}$$

when S^{d-1} is made into a measure space with an appropriate measure, and f is not especially eccentric. The first step is to define the appropriate measure theory on S^{d-1}, and then to apply the theory of product spaces we have developed.

Given $E \subset S^{d-1}$, we say E is **surface measurable** precisely when it's associated unit sector

$$S_E = \{x \in B^d : \hat{x} \in E\}$$

is Lebesgue measurable. We then define the surface measure σ on the measurable subsets of S^{d-1} by setting $\sigma(E) = d \cdot |S_E|$. This is a scaled version of the *pushforward measure* induced from the projection $x \mapsto x/|x|$ from $B^d - \{0\}$ to S^{d-1}, where $B^d - \{0\}$ is given the Lebesgue measure on the Lebesgue measurable subsets. Now define a measure μ on the Lebesgue measurable subsets $(0, \infty)$ by the equation $d\mu = r^{d-1} dr$.

Lemma 5.11. *If $\pi : (0, \infty) \times S^{d-1} \to \mathbf{R}^{d-1}$ is defined by $\pi(r, x) = rx$, then π is Borel measurable, and $\pi_*(\mu \otimes \sigma)$ agrees with the Lebesgue measure on all Borel measurable sets.*

Proof. We note that

$$\mu((x, y]) = \int_{(x,y]} d\mu = \int_x^y r^{d-1} dr = \frac{y^d - x^d}{d}$$

It follows that if E is a Borel subset of S^{d-1}, then

$$\pi_*(\mu \otimes \sigma)(\pi((x, y] \times E)) = \mu((x, y])\sigma(E) = (y^d - x^d)|S_E|$$

Now $\pi((x, y] \times E) = yS_E - xS_E$ is Borel measurable, because S_E is Borel, and by scaling properties of subsets of $\mathbf{R}^d$, we conclude that

$$|\pi((x, y] \times E)| = |yS_E| - |xS_E| = (y^d - x^d)|S_E|$$

so $\pi_*(\mu \otimes \sigma)$ agrees with the Lebesgue measure on the π system of sets of the form $\pi((x,y] \times E)$, where E is Borel. But this family contains a countable basis of $\mathbf{R}^d - \{0\}$, so the π system generates all Borel subsets of $\mathbf{R}^d$, and therefore $\pi_*(\mu \otimes \sigma)$ agrees with the Lebesgue measure on all Borel subsets. $\qquad\square$

If E is a Lebesgue measurable subset of $\mathbf{R}^d$, there are Borel subsets $A \subset E \subset B$ with $|B - A| = 0$, which implies by the above theorem that

$$(\mu \otimes \sigma)(\pi^{-1}(B - A)) = (\mu \otimes \sigma)(\pi^{-1}(B) - \pi^{-1}(A)) = 0$$

and $\pi^{-1}(A) \subset \pi^{-1}(E) \subset \pi^{-1}(B)$, so $\pi^{-1}(E)$ is measurable with respect to the completion of the product of the Borel algebra on $(0, \infty)$ with the Borel algebra on S^{d-1}, which is similarily the completion of the product algebra formed from Lebesgue measurable subsets of $(0, \infty)$ and surface measurable subsets of S^{d-1}. The extension of the measure $\pi_*(\mu \otimes \sigma)$ then agrees with all Lebesgue measurable subsets of $\mathbf{R}^d$.

Theorem 5.12. *If $f : \mathbf{R}^d \to \mathbf{R}$ is Borel measurable, then the slice $f^x(r) = f(rx)$ is Borel measurable for all $x \in S^{n-1}$, the function*

$$x \mapsto \int_0^\infty f(rx) r^{d-1} dr$$

is measurable on S^{d-1}, and

$$\int_{\mathbf{R}^d} f(x) dx = \int_{S^{d-1}} \int_0^\infty f(rx) r^{d-1} dr$$

If f is Lebesgue measurable, then f^x is Lebesgue measurable for almost all x, and the integral identity still holds.

Proof. The map $f : \mathbf{R}^d - \{0\} \to \mathbf{R}$ is also Borel measurable, and the fact that f^x is measurable for all $x \in S^{n-1}$ follows because $f^x = f \circ \pi^x$ is the composition of measurable functions, and π is Borel measurable, and we

now calculate that if f is integrable, then Fubini's theorem gives

$$\int_{\mathbf{R}^d} f(x)dx = \int_{\mathbf{R}^d - \{0\}} f(x)dx = \int_{\mathbf{R}^d - \{0\}} f(x)d\pi_*(\mu \otimes \sigma)(x)$$

$$= \int_{(0,\infty) \times S^{d-1}} (f \circ \pi)(r,x)d(\mu \otimes \sigma)(r,x)$$

$$= \int_{S^{d-1}} \int_0^\infty f(rx)d\mu(r)d\sigma(x)$$

$$= \int_{S^{d-1}} \int_0^\infty f(rx)r^{d-1}drd\sigma(x)$$

If f is only Lebesgue measurable, then $f \circ \pi$ is Lebesgue measurable when we *complete* the σ algebra on $(0,\infty) \times S^{d-1}$, and so we conclude f^x is Lebesgue measurable for almost all x. Similar results hold for f^r. $\qquad\square$

Chapter 6

Integration

Lemma 6.1 (Fatou)**.** *If $f_1, f_2, \cdots \geqslant 0$, then*

$$\liminf \int f_n \leqslant \int \liminf f_n$$

One way to interpret Fatou's lemma is to view the f_i as mass distributions. Whereas it is possible for some of the mass of the f_i to slip through the cracks when taking the lim infs, we can never gain mass via the liminf process.

Lemma 6.2 (Scheffe)**.** *If $f, f_1, f_2, \cdots \in L^1$, and $f_k \to f$ pointwise almost everywhere, then*

$$\int |f_n - f| \to 0 \quad iff \int |f_n| \to \int |f|$$

Proof. First, assume f and f_n are positive functions. If we write

$$|f_n - f| = \max(f_n, f) - \min(f_n, f)$$

then $\min(f_n, f) \leqslant f$, and $\min(f_n, f) \to f$ almost everywhere, so the dominated convergence theorem guarantees that

$$\int \min(f_n, f) \to \int f$$

We can also write $\max(f_n, f) = f + f_n - \min(f_n, f)$. Thus

$$\lim_{n \to \infty} \max(f_n, f) = \lim_{n \to \infty} \int f_n$$

70

Then $\int |f_n - f| = 0$ if and only if $\int \max(f_n, f) = f$, which is equivalent to saying $\int f_n = f$. In general, if $\int |f_n| \to \int |f|$, we can apply Fatou's lemma to conclude that

$$\limsup \int f_n^- \leq \int f^- \quad \limsup \int f_n^+ \leq \int f^+$$

and

$$\int f^- + \int f^+ = \liminf \int |f_n| \leq \liminf \int f_n^- + \liminf \int f_n^+$$

$$\leq \limsup \int f_n^- + \limsup \int f_n^+ \leq \int f^- + \int f^+$$

and this shows the liminf and limsup of $\int f_n^-$ and $\int f_n^+$ must be equal to one another, and that $\lim \int f_n^+ = \int f^+$, $\lim \int f_n^- = \int f^-$. But now applying the theorem for positive functions implies that both $\int |f^+ - f_n^+| \to 0$ and $\int |f^- - f_n^-| \to 0$, and

$$|f - f_n| = |f^+ - f_n^+ + f_n^- - f^-| \leq |f^+ - f_n^+| + |f^- - f_n^-|$$

and so $\int |f - f_n| \to 0$. On the other hand, if $\int |f - f_n| \to 0$, then the inequality $|f - f_n| \geq |f| - |f_n|$ guarantees that

$$\int |f| - \liminf \int |f_n| = \limsup \int |f| - |f_n| \leq 0$$

and the inequality $|f - f_n| \geq |f_n| - |f|$ guarantees that

$$\limsup \int |f_n| - \int |f| = \limsup \int |f_n| - |f| \leq 0$$

Hence the required limit of $\int |f_n|$ exists. $\qquad\square$

Exercise 6.1 (Tao). *Use the dominated convergence theorem to prove that the harmonic series diverges, and to understand the rate of divergence of the harmonic series.*

Proof. If $\sum_{k=1}^{\infty} 1/k < \infty$, then the function $f = \sum n^{-1} \chi_{[n-1,n]}$ would be in $L^1(\mathbf{R})$. But if we define $f_n = n^{-1} \chi_{[0,n]}$, then $f_n \leq f$, so the dominated

71

convergence theorem would imply that, since we have a pointwise limit $\lim f_n = 0$,

$$0 = \int \lim f_n = \lim \int f_n = 1$$

and this is impossible. We can also use this technique to understand how the harmonic series diverges. If we define

$$g_m = \sum_{k=1}^{m} 2^{-k} \chi_{[2^{k-1}, 2^k]}$$

Then $g_m \leq f_{2^m}$, and so

$$S_{2^m} = \int f_{2^m} \geq \int g_m = \sum_{k=1}^{m} 2^{-1} = m/2$$

$\square$

Chapter 7

Interpolation

Interpolation is a core part of the 'hard' style of analysis, crunchy quantitative estimates that give strict bounds on quantitities. One basic tool here is interpolation, which 'in essense' enables one to take two results A_0 and A_1, and via combining them together obtain a family of results' between' the two results, of the form A_t for $0 \leqslant t \leqslant 1$.

The most basic example occurs in the theory of real numbers. Suppose $0 \leqslant A_0 \leqslant B_0$ and $0 \leqslant A_1 \leqslant B_1$. If we define $A_t = A_0^{1-t} A_1^t$, and $B_t = B_0^{1-t} B_1^t$, then it is trivial to verify that $A_t \leqslant B_t$, for the power function $x \mapsto x^a$ is order preserving for $a > 0$. For $t = 1/2$, we obtain the geometric mean inequality

$$\sqrt{A_0 A_1} \leqslant \sqrt{B_0 B_1}$$

Another way to see that the bound is trivial is to notice that the points $\log A_t$ are just the straight line from $\log A_0$ to $\log A_1$, and the result is established geometrically once we notice that log is order preserving.

7.1 Interpolation of Scalars

Let's consider some examples. If $A_0 = AX^{1/p}$, and $A_1 = AX^{1/q}$, then $A_t = AX^{(1-t)/p+t/q}$. These deductions are trivial, but we can still learn about the general inequality from them. For instance, if $A_0 = A_1$, then we find a lower bound $A \leqslant B_t$ over all $0 \leqslant t \leqslant 1$, and this bound can be refined to

$$A \leqslant B_t \min(B_0/B_1, B_1/B_0)^\varepsilon$$

for $\varepsilon < \min(t, 1 - t)$. This tells us that the bound $A \leqslant B_t$ can only be sharp when B_0 and B_1 are roughly equal to one another. If $B_0 = 2^n B_1$, then we can improve the standard bound by a factor of $2^{-|n|\varepsilon}$. Since $2^{-|n|\varepsilon}$ is absolutely summable, it is a good heuristic to imply that the needed interpolation bound is negligable for $|n| \gg 0$.

The inequality $A_t \leqslant B_t$ can be easily generalized to the case where the A_t are defined in such a way that $t \mapsto \log A_t$ is a convex function (we say that the A_t are *log convex*), and $t \mapsto \log B_t$ is concave (though we normally always assume the B_t *are constant*). Thus one can interpolate upper bounds for log convex functions to obtain upper bounds across the domain. However, we cannot use interpolation to lower bound log-concave functions, nor can we extrapolate bounds from interior points (bounding A_0 and $A_{1/2}$ give us no info about A_1), but upper bounding A_0 and lower bounding A_t do gives us a lower bound on A_1. This is just the contraposition of the interpolation inequality.

Application of interpolation relies on the existence of a large class of log convex functions. If f and g are log convex, then fg is log convex because the sum of two convex functions is convex. Similarily,

7.2 Interpolation of functions

If we consider a step function $f = A\chi_Y$, then $\|f\|_p = A\mu(Y)^{1/p}$. Notice this is a log convex function of p, because for any $C > 0$, $\log(C)/p$

Chapter 8

The Lebesgue Stieltjes Integral

We now use our study to provide a correspondence between positive Borel measures on **R** which are finite on intervals, and non-decreasing functions on the real line. Given a measure μ, we define the distribution function

$$
F_\mu(x) = \begin{cases} \mu((0, x]) & : x > 0 \\ 0 & : x = 0 \\ -\mu((-x, 0]) & : x < 0 \end{cases}
$$

Then F is verified to be an increasing right continuous function. On the other hand, if F is an increasing right continuous function with $F(0) = 0$, the next theorem says that there is a Borel measure μ with $F_\mu = F$.

Theorem 8.1. *If F is an increasing real-valued right continuous function on* **R***, then there exists a unique Borel measure μ such that $\mu((a, b]) = F(b) - F(a)$.*

Proof. Define

$$
\mu^*(E) = \inf \sum_{j=1}^{n} F(b_j) - F(a_j)
$$

where the infinum is taken over covers of E by intervals $(a_j, b_j]$. Furthermore, $\mu^*((a, b]) = F(b) - F(a)$. It is easy to see that $\mu^*((a, b]) \leqslant F(b) - F(a)$. Next, suppose $(a, b]$ is covered by $\bigcup(a_i, b_j]$. Fix $\varepsilon > 0$. If we use the right continuity of F to pick $b_j' > b_j$ such that $F(b_j') - F(b_j) \leqslant \varepsilon/2^j$, then

$$
\sum F(b_j') - F(a_j) \leqslant \sum F(b_j) - F(a_j) + \varepsilon
$$

Then (a_i, b_i') is an open cover of $[a', b]$ for any $a < a'$, and therefore has a finite subcover. Since it is easy to verify $F(b) - F(a)$ is optimal for finite covers, we conclude that

$$\sum F(b_j) - F(a_j) + \varepsilon \geqslant \sum F(b_j') - F(a_j) \geqslant \sum F(b) - F(a)$$

Taking $\varepsilon \to 0$ gives optimality. To verify that the Caratheodory extension theorem gives a Borel measure, we need to verify the metric separation condition. Suppose E and F are two sets with $d(E, F) = \delta > 0$. Let $\bigcup (a_i, b_i]$ be a cover of $E \cup F$. By subdividing the intervals, we may assume that each interval has length $< \delta$. But this means that either an interval is redundant, and doesn't cover any points in E or F, covers only points in E, or only covers points in F. Thus the cover can be partitioned into a cover of E and a cover of F, and so

$$\mu^*(E) + \mu^*(F) \leqslant \sum F(b_i) - F(a_i)$$

since the cover was arbitrary, we conclude $\mu^*(E) + \mu^*(F) \leqslant \mu^*(E \cup F)$, which implies μ^* is a metric measure. This completes the proof. $\qquad\square$

Given an increasing real-valued function F, we define the **Lebesgue Stieltjes integral** with respect to F by the equation

$$\int_a^b f(x) \, dF(x) = \int_a^b f(x) d\mu(x)$$

where μ is the Borel measure calculated from the theorem above, and f is integrable with respect to μ. Because of how we defined μ, we will assume the integral is over $(a, b]$. If G is a function of bounded variation, then we can write G as the span of increasing real-valued functions, and we can define the integral by linearity here. We can feasibly compute the integral in a couple special cases.

Example. *For any function F of bounded variation,*

$$\int_a^b dF(x) = F(b) - F(a)$$

which follows from the fact that $\mu((a, b]) = F(b) - F(a)$.

Example. *Let $F = \sum \alpha_k j_k$ be a jump function, with the j_k discontinuous at x_k. Then one verifies that $dF = \sum \alpha_k d\delta_{x_k}$, and so*

$$\int f(x) dF(x) = \sum \alpha_k f(x_k)$$

and f is integrable with respect to μ if $\sum \alpha_k |f(x_k)| < \infty$. A particular case occurs when F is the **Heaviside step function** *$H(x) = \mathbf{I}(x \geqslant 0)$, in which case*

$$\int f(x) dH(x) = f(0)$$

so dH is the Dirac delta function.

Theorem 8.2. *If F is absolutely continuous, then*

$$\int_a^b f(x) \, dF(x) = \int_a^b f(x) F'(x) \, dx$$

for all f integrable with respect to F.

Proof. Suppose that F is increasing, and let μ be the measure corresponding to F. Then we have shown that

$$\mu((a,b]) = F(b) - F(a) = \int_a^b F'(x) \, dx$$

But the class of $(a,b]$ is a π system generating all the Borel sets, so we find that $dF = F'(x) \, dx$ for all Borel sets, and therefore the integrals above agree. $\qquad\square$

Even if the integrator isn't absolutely continuous, we can still obtain formulas analogous to those obtained as consequences of the fundamental theorem of calculus.

Theorem 8.3 (Integration by Parts)**.** *If F and G have finite variation, then*

$$F(b)G(b) - F(a)G(a) = \int_a^b F(x-)dG(x) + \int_a^b G(x-)dF(x) + \sum_{x \in (a,b]} \Delta G(x) \Delta F(x)$$

77

Proof. Let F be associated the measure μ and G associated the measure ν, then we find

$$(\mu \otimes \nu)(a,b]^2 = [F(b) - F(a)][G(b) - G(a)]$$

but

$$\int_a^b [F(x) - F(a)]\, dG(x) = \int_a^b \int_a^y dF(x)dG(y)$$

$$\int_a^b [G(x) - G(a)]\, dF(x) = \int_a^b \int_a^x dG(y)dF(x)$$

and by Fubini's theorem, these two integrals form the upper and lower triangles of $(a,b]^2$, hence they add to $[F(b) - F(a)][G(b) - G(a)]$, except that we have counted the diagonal of $[0,1]^2$, and the integral over the diagonal is given by the sum of discontinuity points. We then just simplify our result to obtain the formula above. $\qquad\square$

Next, we have a version of the chain rule, which in this case has an interesting extra term which doesn't occur in the classical case, where we have no jumps. In differential form, the formula is written

$$d(FG) = F^- dG + G^- dF + d[F,G]$$

where we define $[F,G](x) = \sum_{x<a} \Delta F(x)\Delta G(x)$ the **quadratic covariation** of F and G. If F and G are continuous, then $[F,G] = 0$, and we get the normal chain rule for differentials. These ideas become very important in the study of the stochastic integration of processes with not necessarily finite variation, where in certain circumstances we can interpret $[F,G]$ as a reasonable quantity.

Theorem 8.4. *If f is a C^1 function, and G has finite variation, then $f \circ G$ has finite variation, and*

$$f(G(b)) - f(G(a)) = \int_a^b f'(G(x-))dG(x)$$
$$+ \sum_{x \in (a,b]} [\Delta(f \circ G)(x) - f'(G(x-))\Delta G(x)]$$

Proof. Let A denote the algebra of all functions f for which $f \circ G$ has finite variation, and such that the theorem holds. Clearly, the theorem holds for

$f(x) = 1$. If $f(x) = x$, then $f \circ G = G$ obviously has finite variation, and the formula reads

$$\int_a^b dG(x) = G(b) - G(a) - \sum_{a \leqslant x \leqslant b} [\Delta G(x) - \Delta G(x)]$$

and then the theorem is obvious. Suppose $f, g \in A$. Then $(f \circ G)(g \circ G)$ has finite variation, because if $P = \{x_0, \ldots, x_n\}$ is any partition, then we can find M such that $f \circ G, g \circ G \leqslant M$ on $[a, b]$, and so

$$\sum |g(G(x_i))f(G(x_i)) - g(G(x_{i-1}))f(G(x_{i-1}))|$$
$$\leqslant \sum |g(G(x_i)) - g(G(x_{i-1}))||f(G(x_i))|$$
$$+ |g(G(x_{i-1}))||f(G(x_i)) - f(G(x_{i-1}))|$$
$$\leqslant M[V_{[a,b]}(g \circ G) + V_{[a,b]}(f \circ G)]$$

which is a universal bound independent of the partition. Now by assumption,

$$df(G) = f'(G^-)dG + \{\Delta f(G) - f'(G^-)\Delta G\}$$
$$dg(G) = g'(G^-)dG + \{\Delta g(G) - g'(G^-)\Delta G\}$$

If $h(x) = f(x)g(x)$, then

$$\Delta h(G) = f(G^-)\Delta g(G) + g(G^-)\Delta f(G^-) + \Delta f(G)\Delta g(G)$$

and an integration by parts gives

$$d(h(G)) = f(G^-)dg(G) + g(G^-)df(G) + d[f(G), g(G)]$$
$$= h'(G^-)dG + \{\Delta h(G) - h'(G^-)\Delta G\}$$

Thus h also satisfies the theorem. Our discussion indicates the formula holds for all polynomials. But now if $f \in C^1(\mathbf{R})$ is arbitrary, then we can approximate f and f' uniformly by polynomials, and then we need only verify the formula above is preserved under uniform convergence of integrands. $\qquad \square$

For Riemann Stieltjes integrals over $\mathbf{R}^d$, if $G = (G^1, \ldots, G^d)$, where each G^i is a finite variation function, and if $f \in C^1(\mathbf{R}^d)$, then

$$f(G(b)) - f(G(a)) = \int_a^b \partial_i f(G(s-))dG_s^i$$
$$+ \sum_{a < s \leqslant b} \{\Delta f(G)(s) - \partial_i f(G(s-))\Delta G_s^i\}$$

79

using Einstein notation. We skip the proof, because it is essentially the method used to prove the theorem above. These theorems become very useful in the context of stochastic integration, where they can be extended to where the integrands are random processes.

8.1 Representation of Complex Measures

Let X be a measure space. We define a **complex measure** on X to be map μ on the Σ algebra defining the measure space such that for disjoint collection $E_1, \ldots, E_n$

$$\mu\left(\bigcup E_i\right) = \sum \mu(E_i)$$

we could take the sum above in terms of pointwise convergence, but since the sum must converge to the same value regardless of how the E_i are rearranged, we conclude the sum must converge absolutely to the left hand side. This absolute convergence will show that the behaviour of a complex valued measure is incredibly regular, and we will find that the space of complex measures work very similarly to the space $L^1(X)$, in that the measures are 'integrable' over the entire space. What's more, we can decompose complex measures in a way that is an abstract version of the decomposition of a function of bounded variation into continuous part.

The general stategy by which we analyze integrals of functions with respect to positive measures is to dominate functions by positive functions, for which convergence theory works pretty nicely (Tonelli's theorem, Fatou's lemma, and the monotone convergence theorem provide examples). We would expect similar nice results to hold if we can bound a complex measure μ by a positive measure ν, in the sense that for every set E, $|\mu(E)| \leqslant \nu(E)$. We shall find that we can find a *minimal* ν bounding μ. If E has a partition into measurable sets $E_1, E_2 \ldots$, then we must have

$$\nu(E) = \sum \nu(E_i) \geqslant \sum |\mu(E_i)|$$

Thus it makes sense to *define* a function $|\mu|$ by the formula

$$|\mu|(E) = \sup_{\substack{E = \bigcup E_i \\ E_i \text{ disjoint}}} \sum |\mu(E_i)|$$

The function $|\mu|$ is known as the **total variation** of μ, or, to be specific, the total variation measure. Besides being a measure, we also find that

$|\mu|(X) < \infty$, which implies that the values $|\mu(E)| \leqslant |\mu|(X)$ lie within a compact disk in the plane, which means μ has **bounded variation**.

Theorem 8.5. *$|\mu|$ is a positive measure on X with $|\mu|(X) < \infty$.*

Proof. Let $E = \bigcup E_i$, where the E_i are disjoint. Let $t_1, t_2, \ldots$ be real numbers with $t_i < |\mu|(E_i)$. Then each E_i has a partition A_{ij} with $\sum_j |\mu(A_{ij})| > t_i$, and summing up over all i and j, we conclude that $|\mu|(E) \geqslant \sum |\mu|(E_i)$. To prove the reverse inequality, if A_j is a disjoint partition of E, then $E_i \cap A_j$ is a partition of E_i for each i, and the triangle inequality gives

$$\sum |\mu(A_j)| \leqslant \sum |\mu(E_i \cap A_j)| \leqslant \sum |\mu|(E_i)$$

Since A_j was arbitrary, we conclude $|\mu|(E) \leqslant \sum |\mu|(E_i)$, so we have countable additivity. $\qquad\square$

Lemma 8.6. *Every finite set of complex numbers $z_1, \ldots, z_n$ has a subset $z_{i_1}, \ldots, z_{i_m}$ such that*

$$\left|\sum z_{i_k}\right| \geqslant \frac{1}{\pi} \sum |z_k|$$

Proof. Write $z_i = r_i e^{it}$. $\qquad\square$

Theorem 8.7. *$|\mu|(X) < \infty$.*

Proof. Suppose $|\mu|(E) = \infty$. If $E = \bigcup E_i$, with $\sum |\mu(E_i)| > N$, then we can choose a subset $E_{i_1}, \ldots, E_{i_n}$ with

$$\left|\sum \mu(E_{i_k})\right| = |\mu(\bigcup E_{i_k})| = |\mu(A)| > \frac{N}{\pi}$$

If $B = E - \bigcup E_{i_k}$, then

$$|\mu(B)| = |\mu(E) - \mu(A)| \geqslant |\mu(A)| - |\mu(E)| > \frac{N}{\pi} - |\mu(E)|$$

If N is chosen large enough, we conclude $|\mu(A)|$ and $|\mu(B)|$ are greater than one. This means that we have split E into A and B, and so either $|\mu|(A) = \infty$ or $|\mu|(B) = \infty$. Continuing this process, we may find $(A_1, B_1), (A_2, B_2), \ldots$ with $A_{i+1}, B_{i+1} \subset A_i$, and $|\mu|(A_i) = \infty$, and $|\mu(B_i)| > 1$. Now

$$\mu\left(\bigcup B_i\right) = \sum \mu(B_i)$$

But the sum on the right cannot converge absolutely, because $|\mu(B_i)| > 1$. Thus we converge that $|\mu|(E) < \infty$. $\qquad\square$

81

If μ and λ are complex measures, then it is easy to see that the measures $(\mu + \lambda)(E) = \mu(E) + \lambda(E)$ and $(z\mu)(E) = z\mu(E)$ are also complex measures, for any $z \in \mathbf{C}$. Thus the set of complex measures form a complex vector space. If we define a norm on this space by setting $\|\mu\| = |\mu|(X)$, then we find this is a normed vector space. We shall denote the space of complex measures on a measure space by $\text{cm}(X)$.

As a particular example of this process, consider a complex measure μ such that $\mu(E) \in \mathbf{R}$ for all measurable sets E, in which case we say μ is a **signed measure**. If $|\mu|$ is defined as above, then the functions

$$\mu^+ = \frac{|\mu| + \mu}{2} \quad \mu^- = \frac{|\mu| - \mu}{2}$$

are both positive measures, and are known as the **positive** and **negative variations** of μ. The decomposition $\mu = \mu^+ - \mu^-$ is known as the **Jordan decomposition** of μ.

We say that a positive measure λ is **absolutely continuous** with respect to a positive meaure μ if $\lambda(E) = 0$ whenever $\mu(E) = 0$, and we write $\lambda \ll \mu$. If there is a measurable set A such that $\lambda(E) = \lambda(A \cap E)$ for every E, we say that λ is **concentrated on** A (this is equivalent to saying $\lambda(E) = 0$ if E is disjoint from A). If two measures μ and λ are concentrated on disjoint sets, we say they are **mutually singular**, and write $\mu \perp \lambda$.

Theorem 8.8. *Suppose μ is a positive measure, and λ, λ_1, and λ_2 are also measures, then*

- *If λ is concentrated on A, so too is $|\lambda|$.*

- *If $\lambda_1 \perp \lambda_2$, then $|\lambda_1| \perp |\lambda_2|$.*

- *If $\lambda_1 \perp \mu$ and $\lambda_2 \perp \mu$, then $\lambda_1 + \lambda_2 \perp \mu$.*

- *If $\lambda_1, \lambda_2 \ll \mu$, then $\lambda_1 + \lambda_2 \ll \mu$.*

- *If $\lambda \ll \mu$, $|\lambda| \ll \mu$.*

- *If $\lambda_1 \ll \mu$, and $\mu \perp \lambda_2$, then $\mu_1 \perp \mu_2$*

Proof. Elementary. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

The next theorem is perhaps the most important theorem in measure, allowing us to reduce the study of a measure on a space to considering particular functions in $L^1(X)$.

Lemma 8.9. *If μ is a positive σ-finite measure on X, then there is a function $w \in L^1(X)$ such that $0 < w(x) < 1$ for every $x \in X$.*

Proof. Let X be the disjoint union of $E_1, E_2, \ldots$, with $\mu(E_i) < \infty$. By taking successive unions, we may assume $\mu(E_i) \neq 0$ for all i (or μ is already a finite measure, and we can define $w = 1/2$). Define

$$w(x) = \sum \mathbf{I}(x \in E_k) \frac{1}{\mu(E_i)2^k}$$

Then w is the limit of step functions, hence measurable, and it is easy verified to be in $L^1(X)$. $\qquad\square$

Thus the measures μ and $w\mu$ both have the same sets of measure zero, except that $w\mu$ is now a *finite* measure on X. This simplifies the next theorem considerably using the theory of Hilbert spaces, in particular, L^2 functions on X. The proof is due to Von Neumann.

Theorem 8.10 (Radon-Nikodym). *Let μ be a positive σ finite measure on a σ algebra X, and let λ be a complex measure on X. Then there is a unique decomposition $\lambda = \lambda_a + \lambda_s$, known as the **absolute** and **singular** part of λ, with $\lambda_a \ll \mu$, and $\lambda_s \perp \mu$. If λ is positive and finite, then so too are λ_a and λ_s. There is also a unique $f \in L^1(X)$ such that $\lambda_a = f\mu$, and we often write $f = d\lambda_a/d\mu$, and call it the **Radon-Nikodym derivative** of λ_a w.r.t μ.*

Proof. The uniqueness follows because 0 is the only absolutely continuous and singular measure. The uniqueness of f follows from the fact that the integral of f is determined on every measurable set. The existence is the important part of the theorem.

Assume at first that λ is positive and bounded. Given the function w constructed in the last lemma, we can construct the measure $\varphi = \lambda + w\mu$, which is positive and buonded. Because of the construction, the standard machine verifies that

$$\int f\, d\varphi = \int f\, d\lambda + \int f w\, d\mu$$

for every non-negative measurable f. If $f \in L^2(\varphi)$, then

$$\left| \int f\, d\lambda \right| \leqslant \int |f|\, d\lambda \leqslant \int |f|\, d\varphi \leqslant \|f\|_{L^2(\varphi)} \sqrt{\varphi(X)}$$

83

It follows that integration with respect to λ is a bounded functional on $L^2(\varphi)$, and it follows that there exists $g \in L^2(\varphi)$ such that for every $f \in L^2(\varphi)$,

$$\int f d\lambda = \int f g d\varphi$$

If $f = \chi_E$ we conclude that

$$\lambda(E) = \int_E g d\varphi$$

Since $\lambda(E) \leqslant \varphi(E)$, we conclude that

$$0 \leqslant \varphi(E)^{-1} \int_E g d\varphi \leqslant 1$$

so $0 \leqslant g \leqslant 1$ almost surely with respect to φ. We can rearrange the equation defining g to read

$$\int f(1-g) d\lambda = \int f g w d\mu$$

Set $A = \{0 \leqslant g < 1\}$, and $B = \{g = 1\}$, and define

$$\lambda_a(E) = \lambda(A \cap E) \quad \lambda_s(E) = \lambda(B \cap E)$$

s $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Chapter 9

The Hausdorff Measure

The expression of geometric properties of subsets of $\mathbf{R}^d$ requires more than can be expressed using the Lebesgue measure. For instance, curves and surfaces all have measure zero in two and three dimensions respectively, and thus we cannot distinguish them by the Lebesgue measure from any of the other nasty Lebesgue measurable subsets of measure zero. Hausdorff showed that there is a notion of 'dimension' of measure zero subsets of $\mathbf{R}^d$ which matches the dimension of corresponding curves and surfaces. Even more interestingly, Hausdorff's theory of dimension gives certain fractal subsets fractional dimension!

Here is the general idea. If $X = [0,1)$ is a unit interval, then $nX = [0,n)$ is the union of n disjoint translates of $[0,1)$. If we instead consider the unit square $X = [0,1) \times [0,1)$, then $nX = [0,n) \times [0,n)$ is the union of n^2 disjoint translates of $[0,1)$. If X is a unit cube, the nX is the union of n^3 disjoint translates of $[0,1)$, and so on and so forth. Thus, it makes sense to define the dimension of X to be the value α such that nX is the union of n^α disjoint copies of X. Note that if X is the Cantor set, then $3X$ is the union of two translates of X, so we would have to be willing to say that the Cantor set 'has dimension $\log_3 2 = 0.6309\ldots$.