

# Measure Theory

Jacob Denson

December 24, 2017

# Contents

<b>1</b>	<b>The Lebesgue Measure</b>	<b>2</b>
1.1	Measuring Elementary Sets . . . . .	5
1.2	Jordan Measurable Sets . . . . .	9
1.3	Lebesgue Measure . . . . .	16
1.4	Appendix: Banach Tarski . . . . .	22
<b>2</b>	<b>Measure Extension</b>	<b>24</b>
<b>3</b>	<b>Product Spaces</b>	<b>29</b>
3.1	Completion of Product Measures . . . . .	34
3.2	Integration in Polar Coordinates . . . . .	36
<b>4</b>	<b>Integration</b>	<b>39</b>
<b>5</b>	<b>Interpolation</b>	<b>42</b>
5.1	Interpolation of Scalars . . . . .	42
5.2	Interpolation of functions . . . . .	43
<b>6</b>	<b>Differentiation and Averages</b>	<b>44</b>
6.1	Maximal Functions and Differentiation . . . . .	44
6.2	Lebesgue Density Theorem . . . . .	48
6.3	Generalizing The Differentiation Theorem . . . . .	50
6.4	Approximations to the Identity . . . . .	51
6.5	Differentiability of Measurable Functions . . . . .	55
6.6	Absolute Continuity . . . . .	60
6.7	Differentiability of Jump Functions . . . . .	63
6.8	Application: Rectifiable Curves . . . . .	65
6.9	The Lebesgue Stieltjes Integral . . . . .	66

6.10 Representation of Complex Measures . . . . .	71
<b>7 The Hausdorff Measure</b>	<b>76</b>

# Chapter 1

## The Lebesgue Measure

Area is one of the most primitive measurements in geometry. Every elementary school student knows that the area of a circle of radius  $r$  is  $\pi r^2$ , and that the area of a rectangle is equal to the product of the lengths of its two distinct 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 more complex 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 shapes enclosed by the shape.

With the advent of Cartesian coordinates and the subsequent modern introduction of set theory, shapes are identified with subsets of  $\mathbf{R}^2$ , and one can obtain strange subsets of  $\mathbf{R}^2$  for which these methods begin 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  $\pi$ , 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 so 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, and volume 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 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 general sets.

- One often wants to measure the length of continuous curves in the plane. In 1890, Peano constructed a continuous curve mapping the unit interval  $[0,1]$  surjectively onto the unit rectangle  $[0,1]^2$ . The length of this curve is certainly suspect, because it is constructed as the limit of piecewise differentiable curves whose lengths tend to infinity. Thus a more general analysis of which curves have a ‘measurable length’ required an invention, and this required the development of measure theory. These curves 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 also does not have a reasonable length, because there are too many ‘bends’ in the path. Similarly, difficulties were found when measuring the length of the perimeter of countries, which are too ‘jagged’ to be measured.
- 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 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 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  $f_1, f_2, \dots$ , 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

$$\int_0^1 f(x) dx = \lim_{n \rightarrow \infty} \int_0^1 f_n(x) dx$$

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  $S_1, \dots, S_n$ , 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  $S_1, S_2, \dots$ , 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  $|S_i| \rightarrow |S|$ .

We will begin by extending the notion of sets to fairly general subsets of  $\mathbf{R}^n$ . 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. We remind the reader that we shall use  $\mu$  for the length of a subset of  $\mathbf{R}$ , the area of a subset of  $\mathbf{R}^2$ , the volume of a subset of  $\mathbf{R}^3$ , and higher dimensional variants.

## 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  $\mathbf{R}^d$ , meaning it is the product of intervals  $I_i$  starting at  $a_i$  and ending at  $b_i$ , then we define

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

The easiest sets to measure the area of 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 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  $R_{i_1 \dots i_d}$ , where

$$R_{i_1 \dots i_d} = I_{i_1}^1 \times \cdots \times I_{i_d}^d$$

and the endpoint of  $I_i^k$  is the startpoint of  $I_{i+1}^k$ . We can then just apply distributivity to conclude that

$$\sum_{i_1, \dots, i_d} |R_{i_1 \dots i_d}| = \sum_{i_1, \dots, i_d} |I_{i_1}^1| \dots |I_{i_d}^d| = \prod_{k=1}^d \left( \sum_j |I_j^k| \right)$$

and all that remains is to show  $\sum_j |I_j^k| = |I_k|$ . 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.  $\square$

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

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

**Lemma 1.3.** *If  $R_1, \dots, R_n$  and  $S_1, \dots, S_m$  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*

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

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

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

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

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



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  $\mathbf{R}^n$  by the formula

$$\mu(E) = \lim_{N \rightarrow \infty} \frac{|\mathbf{Z}^n/N \cap E|}{N^n}$$

However, this definition does run 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 measure is by Monte Carlo theory in 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  $R_1, \dots, R_n$ , 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 unambiguous measure of elementary sets to approximate the measure of arbitrary subsets of the plane.

But first, let's discuss an important structure of the family of elementary sets – its invariance under certain set theoretic operations. This is a way of measuring the ‘power’ of measurement that the family has. One often has to consider the areas of the set formed from the union of two sets, or the intersection, and it is useful to know that we can measure the size of a set if it is the union of measurable sets, or the intersection of measurable sets. Before we move on to measuring the size of more advanced sets, let's dwell on the algebraic power of the family of elementary 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  $E \Delta F = (E - F) \cup (F - E)$  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

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

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 is an **algebra** if it is closed under the union and complement operations. The reason for the name is that, if we view the operation  $\Delta$  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  $\mathbf{F}_2$ . What we have just proven is that the set of elementary

subsets contained in any bounded box, like, for instance,  $[0, 1]^n$ , is an algebra. The fact that elementary sets are not an algebra is one of the reasons we have to enlarge the family of sets we can measure.

The measure of size we have constructed on the family of elementary sets also have certain relations which are very important. First, we note that because of the corresponding property for the measure of boxes, if  $E = \bigcup E_i$  is the union of disjoint elementary sets, then

$$|E| = \sum |E_i|$$

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

$$|E| \leq \sum |E_i|$$

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

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

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  $F$ , which holds 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!

## 1.2 Jordan Measurable Sets

In the last section, we constructed a consistent measure on the family of elementary sets. However, this family is certainly 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. We haven’t

used the method of approximating shapes by simple sets from above and below.

The work of Newton improved on the methods of approximation by Archimedes to measure the area of certain sets by taking Riemann integrals, and applying the fundamental theorem of calculus. 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

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

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  $\int_R f(x) dx$ . If  $E$  is a bounded set, contained in some box  $R$ , then the area of  $E$  can be measured 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.

Of course, the full definition of the Riemann integral is overkill to define the Jordan measure of a set. Given some partition  $P$  of  $R$  into boxes, we can write

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

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

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

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

and a bounded set  $E$  is Jordan measurable precisely when  $J_*(E) = J^*(E)$ , and this common value is then the measure of the set. 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  $J^*(E \cup F) \leq J^*(E) + J^*(F)$ . The next theorem also offers another important criterion for Jordan measurability.

**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  $J^*(E \Delta F) \leq \varepsilon$ .*

*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  $|E_2 - E_1| \leq \varepsilon$ . But  $E_2 - E \subset E_2 - E_1$ , and so the monotonicity of  $J^*$  allows us to conclude that

$$J^*(E \Delta E_2) = J^*(E_2 - E) \leq |E_2 - E_1| \leq \varepsilon$$

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

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

taking  $\varepsilon$  to zero, we conclude  $E$  is Jordan measurable.  $\square$

**Theorem 1.6.** *If  $E$  and  $F$  are Jordan measurable subsets of  $\mathbf{R}^n$ , 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)  $E \Delta F$  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  $J^*(E_1 \Delta E), J^*(F_1 \Delta F) \leq \varepsilon$ , then

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

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

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

which proves the theorem.  $\square$

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  $R_1, R_2, \dots$  covering  $E$  with

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

Thus a measure zero subset can be very efficiently covered by measure zero sets. 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  $f : R \rightarrow \mathbf{R}$  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

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

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

$$A_\varepsilon = \{x : \text{osc}(f, x) \geq \varepsilon\}$$

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

$$\begin{aligned} U(f, P) - L(f, P) &= \sum_{k=1}^M |R_k| \left[ \sup_{x \in R_k} f(x) - \inf_{x \in R_k} f(x) \right] \\ &\leq \sum_{k=1}^N |R_k| 2\|f\|_\infty + \sum_{k=N+1}^M |R_k| \varepsilon \\ &\leq (2\|f\|_\infty + |R|) \varepsilon \end{aligned}$$

and by taking  $\varepsilon \rightarrow 0$ , 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  $\sum_i |U_{ij}| \leq \varepsilon/2^n$ , then  $\sum_{ij} |U_{ij}| \leq \varepsilon$  and covers  $\bigcup E_i$ . Thus, given a Riemann integrable function  $f$ , it suffices to prove that  $A_\varepsilon$  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

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

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

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

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

**Example.** Let  $R$  be a closed box in  $\mathbf{R}^n$ , and  $f : R \rightarrow \mathbf{R}$  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  $\varepsilon$  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

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

If we let  $\varepsilon \rightarrow 0$ , we obtain that  $\Gamma(f)$  has upper measure zero, and thus lower measure zero. The set  $X = \{(x, t) \in R \times \mathbf{R} : 0 \leq f(x) \leq t\}$  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

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

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

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

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  $|x_i| \leq |x| \leq 1$ , so  $B_1$  is contained in the square  $[-1, 1]^d$ . Similarly, if  $|x_i| \leq 1/\sqrt{d}$  for each  $i$ , then  $|x| \leq 1$ , so  $[-1/\sqrt{d}, 1/\sqrt{d}] \subset B_1$ . This yields the bounds

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

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  $2\pi^{d/2}/d(d/2)!$ .

One problem with the Jordan measure is it is too connected to the topology of the underlying space  $\mathbf{R}^d$ . In particular, one can easily verify that for any set  $E$ ,  $J^*(E) = J^*(\bar{E})$ , and  $J_*(E) = J_*(E^\circ)$ . It follows that  $E = \mathbf{Q} \cap [0, 1]$  isn't a Jordan measurable set, because the set is dense in  $[0, 1]$ , but has empty interior, so  $J^*(E) = |[0, 1]| = 1$ , but  $J_*(E) = J_*(\emptyset) = 0$ . Another way to see this is to note that  $\partial E = [0, 1]$ , which is a large set. Because of the connection between Jordan measurable sets and topology, it follows that sets with 'too many holes' cannot be Jordan measurable. To assign these sets a reasonable notion of measure, we require the development of the full theory of Lebesgue measure.

### 1.3 Lebesgue Measure

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

*Proof.* Given any open set  $U$ , let  $\mathcal{U}$  denote the family of intervals contained in  $U$ . If we order this family by inclusion, and if  $\{I_\alpha\}$  is a linear chain of intervals, each contained within one another, then  $\bigcup I_\alpha$  is an interval, though possibly unbounded. It follows by Zorn's lemma that maximal intervals in  $U$  exist. If  $I$  and  $J$  are distinct maximal intervals, then they are disjoint, for otherwise  $I \cup J$  is an interval in  $U$  containing them both. Since every point  $x \in U$  is contained in an interval in  $\mathcal{U}$ , this means that  $U$  is equal to the union of maximal intervals in  $\mathcal{U}$ . Since every interval contains a rational number, and the rational numbers are countable, the family of maximal intervals is at most countable. Uniqueness follows because if  $U = \bigcup I_i = \bigcup J_i$ , then  $J_i = \bigcup J_i \cap I_j$ , and only one  $j$  can have  $J_i \cap I_j \neq \emptyset$ , because the union of disjoint intervals is not an interval.  $\square$

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

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

The definition of measure we will eventually give will give us this identity, but it is not clear how we can generalize this to other sets, nor to subsets of  $\mathbf{R}^d$ , because the uniqueness decomposition does not hold here. Here is the analogous theorem. 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 the *dyadic intervals* in  $\mathbf{R}^d$ , 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  $U_0 = \emptyset$ . 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  $U_m \rightarrow U$  as  $m \rightarrow \infty$ .  $\square$

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  $R_i$ , it makes sense to define

$$|U| = \sum_{k=1}^{\infty} |R_i|$$

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

MERGE LAST AND NEW

If we are able to stick with Jordan measurable sets, we should, because it is here that integration theory works in the best way. However, not all sets are Jordan measurable, and often when studying fractal sets

one runs into unmeasurable sets. As we have seen, Jordan measurable sets are precisely those whose boundaries are very small – for sets with lots of ‘holes’, the boundaries can be rather large, and in these situations we still need a well defined notion of measure. Algebraic, even if a sequence of sets  $E_1, E_2, \dots$  is Jordan measurable, their union  $\bigcup E_i$  and their intersection  $\bigcap E_i$  need not be measurable, even if these sets are bounded. In terms of Riemann integrability, this causes problems with understanding the pointwise limit of functions: A sequence of uniformly bounded Riemann integrable functions  $f_n : [0, 1] \rightarrow \mathbf{R}$  which converges pointwise to a bounded function  $f : [0, 1] \rightarrow \mathbf{R}$  need not be Riemann integrable. If we replace pointwise convergence with uniform convergence, then  $f$  will be Riemann integrable, but this relates to the fact that uniform convergence allows one to converge  $f$  with finitely many rectangles (ELABORATE HERE).

To obtain a family of sets with a well defined measure theory which satisfies countable additivity, we must tinker with how we defined 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 if we replace the finiteness with infinitely many rectangles. We consider the values

$$\mu^*(E) = \inf\{\sum\}$$

**Definition.** If  $A$  is a set of real numbers, then its Lebesgue measure is

$$m(A) = \inf \left\{ \sum_{k=1}^{\infty} m(I_k) : \bigcup_{k=1}^{\infty} I_k \supset A \right\}$$

The end goal of this passage is to find out what it takes to prove that if  $\{A_i\}$  is a disjoint collection of sets, then  $m(\bigcup A_i) = \sum m(A_i)$ . This will give us intuition in the abstract case. In this case, one side of the equality is fairly easy to show.

**Theorem 1.10.** *If  $\{A_i\}$  is a countable collection of sets, then  $m(\bigcup A_i) \leq \sum m(A_i)$ .*

*Proof.* If any  $A_i$  has infinite length, then the theorem is trivial. Thus assume all  $A_i$  have finite measure. Fix some  $\varepsilon > 0$ . For each  $A_i$ , pick a

countable set  $\mathcal{I}_i$  of open intervals such that  $\sum_{I \in \mathcal{I}_i} I \leq m(A_i) + \varepsilon/2^k$ . Then  $\bigcup \mathcal{I}_i$  is a countable collection of open intervals covering  $\bigcup A_i$ , and so

$$m\left(\bigcup A_i\right) \leq \sum_{i=1}^{\infty} \sum_{I \in \mathcal{I}_i} m(I) \leq \sum_{i=1}^{\infty} [m(A_i) + \varepsilon/2^k] = \sum_{i=1}^{\infty} m(A_i) + \varepsilon$$

The proof is completed since  $\varepsilon$  was arbitrary.  $\square$

**Lemma 1.11.** *If  $A \subset B$ ,  $m(A) \leq m(B)$ .*

*Proof.* Any cover of  $B$  is a cover of  $A$ .  $\square$

Let us check the  $m$  is well defined, when passing from lengths of intervals to approximations of arbitrary sets.

**Lemma 1.12.** *For any interval  $I = (a, b)$ ,  $m(I) = b - a$*

*Proof.* First, we will verify that  $m([a, b]) = b - a$ . Let  $\mathcal{I}$  be a collection of open intervals such that  $\bigcup \mathcal{I} \supset [a, b]$ . Without loss of generality, we may choose a finite subcover, since  $[a, b]$  is compact. Using this finiteness, construct a sequence  $(a_1, b_1), \dots, (a_n, b_n)$  from  $\mathcal{I}$  such that  $b_i \geq a_{i+1}$  for each  $i$ ,  $a_1 \leq a$ , and  $b_n \geq b$ . Then

$$\begin{aligned} \sum_{I \in \mathcal{I}} m(I) &\geq \sum_{i=1}^n m((a_i, b_i)) = \sum_{i=1}^n b_i - a_i \\ &\geq (b_n - a_n) + \sum_{i=1}^{n-1} (a_{i+1} - a_i) \\ &= (b_n - a_n) + (a_n - a_1) = b_n - a_1 \geq b - a \end{aligned}$$

Thus  $m([a, b]) \geq b - a$ . Now, fix  $\varepsilon > 0$ . Choose the cover

$$\mathcal{I} = \{(a - \varepsilon, a + \varepsilon), (a, b), (b - \varepsilon, b + \varepsilon)\}$$

Now  $\bigcup \mathcal{I} = (a - \varepsilon, b + \varepsilon) \supset [a, b]$ , so

$$m([a, b]) \leq m((a, b)) + m((a - \varepsilon, a + \varepsilon)) + m((b - \varepsilon, b + \varepsilon)) = b - a + 4\varepsilon$$

Since  $\varepsilon$  was arbitrary,  $m([a, b]) \leq b - a$ , and so  $m([a, b]) = b - a$ .

Surely,  $m((a, b)) \leq m([a, b]) = b - a$ . But also, by Lemma (1.1),

$$m([a, b]) \leq m((a, b)) + m(\{a\}) + m(\{b\}) = m((a, b))$$

since the length of a single point is zero.  $\square$

Now we want to know that measuring the union is the same as measuring the component parts, as our intuition would tell us. However, Banach and Tarski have warned us that this won't be true of all sets. One side of the inequality can be shown for all sets, but we must specialize to obtain equality – defining exactly what it means for a set to be measurable, as we were discussing above.

**Definition.** A set  $A$  is **measurable** (in the manner of Lebesgue), if for any other set  $B$ ,  $m(B) = m(A \cap B) + m(A^c \cap B)$ .

It is simple to verify that  $\mathbf{R}$  is a measurable set, and if  $A$  is measurable, then so is  $A^c$ . More complicated is the fact that open intervals are measurable.

**Lemma 1.13.** *For any real number  $a$ ,  $(a, \infty)$  is measurable.*

*Proof.* Let  $A$  be an arbitrary set. Let  $\mathcal{I}$  be a countable collection of intervals such that  $\sum_{I \in \mathcal{I}} m(I) \leq m(A) + \varepsilon$ . Then, for each  $I$ , either  $I \cap (a, \infty)$  is empty or an interval, as is  $I \cap (-\infty, a]$ , and the measure of  $I$  is equal to the measure of the sum. Thus

$$m(A \cap (a, \infty)) + m(A \cap (-\infty, a]) \leq \sum m(I_k \cap (a, \infty)) + \sum m(I_k \cap (-\infty, a]) = \sum m(I_k) \leq m(A) + \varepsilon$$

So  $m(A \cap (a, \infty)) + m(A \cap (-\infty, a]) \leq m(A)$ , and we have already proved the inequality the other way.  $\square$

**Lemma 1.14.** *If  $A$  and  $B$  are measurable, then so is  $A \cup B$ .*

*Proof.* Let  $S$  be an arbitrary subset of the reals. Then

$$\begin{aligned} m(S) &= m(S \cap A) + m(S \cap A^c) \\ &= m(S \cap A) + m(S \cap A^c \cap B) + m(S \cap A^c \cap B^c) \\ &= m(S \cap A) + m(S \cap B \cap A^c) + m(S \cap [A \cup B]^c) \\ &= m([S \cap A] \cup [S \cap B \cap A^c]) + m(S \cap [A \cup B]^c) \\ &= m(S \cap [A \cup B]) + m(S \cap (A \cup B)^c) \end{aligned}$$

One may get from the second last equation to the third last equation by applying the measurability of  $A$  to the first measured set.  $\square$

**Corollary 1.15.** *If  $A$  and  $B$  are measurable, then  $A \cap B$  and  $A - B$  are measurable.*

*Proof.*  $A \cap B = (A^c \cup B^c)^c$ , and  $A - B = A \cap B^c$ . □

**Corollary 1.16.** *All open intervals are measurable.*

**Corollary 1.17.** *If  $A$  is any set, and  $E_1, \dots, E_n$  is a finite collection of disjoint measurable sets, then*

$$m(A \cap (\bigcup_{k=1}^n E_k)) = \sum_{k=1}^n m(A \cap E_k)$$

What we have shown here is that the set of measurable sets is a Boolean algebra. We can go one further.

**Lemma 1.18.** *If  $E_1, E_2, \dots$  is a countable collection of disjoint measurable sets, then  $E = E_1 \cup E_2 \cup \dots$  is measurable.*

*Proof.* Define  $F_n = \bigcup_{k=1}^n E_k$ . Then  $F_n$  is measurable, and  $F_n^c \supset E^c$ . Hence

$$\begin{aligned} m(A) &= m(A \cap F_n) + m(A \cap F_n^c) \geq m(A \cap F_n) + m(A \cap E^c) \\ &= \sum_{k=1}^n m(A \cap E_k) + m(A \cap E^c) \end{aligned}$$

Since  $n$  was arbitrary,

$$m(A) \geq \sum_{k=1}^{\infty} m(A \cap E_k) + m(A \cap E^c) \geq m(A \cap E) + m(A \cap E^c)$$

But  $m(A) \leq m(A \cap E) + m(A \cap E^c)$ , so  $E$  is measurable. □

**Corollary 1.19.** *The countable union of measurable sets are measurable.*

*Proof.* Simply modify measurable sets by elementary set operations so they are disjoint, and then take their union. □

From this theorem, we can determine that every open set in  $\mathbf{R}$  is open, as every open set is the countable union of open intervals.

We can now show what we originally set out to solve.

**Theorem 1.20.** *If  $\{A_i\}$  is a countable collection of pairwise disjoint measurable sets, then*

$$m(\bigcup A_i) = \sum m(A_i)$$

*Proof.* The calculations in the above theorem show that, letting  $A = \bigcup A_i$ ,

$$m(A) \geq \sum_{k=1}^{\infty} m(A \cap A_i) + m(A \cap A^c) = \sum_{k=1}^{\infty} m(A_i)$$

but this shows equality, since the other direction of inequality always holds.  $\square$

The function  $m$ , restricted to measurable sets, will now be known as the Lebesgue measure on  $\mathbf{R}$ . It is the first in a line of a general class of functions known as measures, defined on subsets of a space and measuring these subset's size. The notions of measurable set will be abstracted to the properties proved above. That is, we can measure the union, intersection, and complement of all measurable sets. Most theorems in measure theory are actually be proved in general just as easily as on the Lebesgue measure we have just described.

## 1.4 Appendix: Banach Tarski

Let us consider the sphere. A nice property of this object is that it is invariant under any rotation - that is, if you take a point, and rotate it around the origin, you will never end up at a point off the unit sphere. Mathematically, we say that the orthogonal group  $O(3)$  acts on the sphere  $S^1$ .

The core technique of this proof can be executed in a simpler form on free groups. Consider the free group  $F_{\{a,b\}}$  on two characters. Let  $S(a)$  be the set of all sequences whose simplest form begins with  $a$ , and define  $S(b)$ ,  $S(a^{-1})$ , and  $S(b^{-1})$  similarly. We have the following equalities:

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

$$F_{\{a,b\}} = S(a) \cup aS(a^{-1})$$

$$F_{\{a,b\}} = S(b) \cup bS(b^{-1})$$

Thus we have partitions  $F_{\{a,b\}}$  into four sets. By 'rotating' two of these partitions, we obtain two copies of the group.



The trick to the Banach-Tarski paradox on the sphere is to find subsets of the orthogonal group that behave like  $F_{\{a,b\}}$ . We will say a subset  $X$  of euclidean space can be **paradoxically decomposed**, if it can be expressed as the disjoint union of subsets,  $X_1 \cup \cdots \cup X_n$ , and, under group actions  $g_1, \dots, g_n \in O(3)$ , we may express  $X = g_1 X_1 \cup \cdots \cup g_i X_i$ , and  $X = g_{i+1} X_{i+1} \cup \cdots \cup g_n X_n$ , for some  $i$ .

**Lemma 1.21.** *There is a subgroup of  $O(3)$  isomorphic to  $F_{\{a,b\}}$ .*

*Proof.* Map  $a$  to a rotation horizontally by  $\sqrt{2}\pi$  radians, and map  $b$  to a rotation vertically by  $\sqrt{2}\pi$  radians. This induces a homomorphism from  $F_{\{a,b\}}$  to  $O(3)$ . We claim this homomorphism is injective.  $\square$

**Theorem 1.22** (Banach Tarski). *The sphere may be paradoxically decomposed.*

=

## Chapter 2

### 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  $\sigma$  algebra, and then to extend this measure to the  $\sigma$  algebra this family generates. If  $X$  is a set, a family  $\Sigma_0$  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*  $A \Delta B = (A - B) \cup (B - A)$  as addition). A  $\sigma$  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  $\sigma$  algebra, and thus the extension was necessary to construct the satisfying theory of the Lebesgue measure. To generalize this construction, we take an algebra  $\Sigma_0$ , on which a sense of measure is defined, and try to extend this to a  $\sigma$  algebra containing  $\Sigma_0$ . The intuitive notion of measure is an additive function  $\mu_0 : \Sigma_0 \rightarrow [0, \infty]$  such that  $\mu_0(\emptyset) = 0$ , which is countably additive when countable additivity is well defined in  $\Sigma_0$ . 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  $\sigma$  subadditive, and  $\mu^*(\emptyset) = 0$ . On any set  $X$ , we call a function  $\mu^*$  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 **Caratheodory measurable** if it satisfies every equation of the form above. Since  $\mu^*$  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 Caratheodory measurable.

**Lemma 2.1.** *The set of Caratheodory measurable sets forms a  $\sigma$  algebra, and  $\mu^*$  is a measure on this family.*

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

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

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 Caratheodory measurable sets is closed under countable unions of disjoint sets, and  $\mu^*$  is countably additive on this family. Consider measurable sets  $E_1, E_2, \dots$ , and let

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

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

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

and thus

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

Letting  $n \rightarrow \infty$  gives that

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

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

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 2.2** (Caratheodory's Extension Theorem). *Every premeasure on an algebra  $\Sigma_0$  can be extended uniquely to a measure on the  $\sigma$  algebra  $\Sigma$  that  $\Sigma_0$  generates.*

*Proof.* Given  $\mu_0$ , define the exterior measure

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

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  $\varepsilon \rightarrow 0$ . 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

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

we conclude

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

This shows in particular that  $\mu^*(\emptyset) = \mu(\emptyset) = 0$ , so  $\mu^*$  is an exterior measure. Because of Caratheodory's lemma, all that remains is to show that all elements  $E$  of  $\Sigma_0$  are Caratheodory 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  $\varepsilon \rightarrow 0$ . If  $\mu^*(A) = \infty$ ,  $\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A \cap E^c)$  is obvious. This shows that the  $\sigma$  algebra of Caratheodory measurable sets contains  $\Sigma_0$ , and the measure  $\mu^*$  defined on it extends  $\mu$ . Note, however, that there may be many more Caratheodory measurable sets than the subsets of  $\Sigma$ , but we can solve the problem by restriction.  $\square$

**Corollary 2.3.** *If  $X$  is  $\sigma$ -finite, the extension is unique on  $\Sigma$ .*

*Proof.* Let  $\mu$  be another extension of  $\mu_0$  to  $\Sigma$  rather than  $\mu^*$ . Choose some  $E \in \Sigma$ , and fix  $\varepsilon > 0$ . For any countable covering of  $E$  by elements  $A_1, \dots \in \Sigma_0$ ,

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

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

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

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.  $\square$

For the purposes of using the Caratheodory extension theorem, it is often useful to use the following method of generating a premeasure. A **semialgebra** on a set  $X$  is a family of 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  $\mu$  defined on a semialgebra with  $\mu(\emptyset) = 0$  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  $\infty$ . 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  $\mu$  is countably additive, which we partook in the first chapter, but once this is done, the Caratheodory extension theorem provides a black box to extend  $\mu$  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 3

## Product Spaces

Let  $X$  and  $Y$  be measure spaces, with respective  $\sigma$  algebras  $\Sigma$  and  $\Pi$ . 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  $\sigma$  algebra on  $X \times Y$  is the one generated by elementary rectangles, and we denote this  $\sigma$  algebra by  $\Sigma \otimes \Pi$  (it isn't *exactly* the tensor product if we view the  $\sigma$  algebras as algebras over  $\mathbf{F}_2$ , but it has analogous properties to the tensor product). Since

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

we conclude the family of elementary rectangles is a  $\pi$  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 3.1.** *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  $y \notin F$ ,  $(E \times F)^y = \emptyset$ , 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  $\lambda$  system containing the  $\pi$  system of elementary rectangles, so

the  $\pi$ - $\lambda$  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  $f : X \times Y \rightarrow Z$  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  $\Sigma$  is a **monotone class** if it is closed under upward limits: If  $E_1 \subset E_2 \subset \dots \in \Sigma$ , then  $\lim E_i \in \Sigma$ , and if  $E_1 \supset E_2 \supset \dots \in \Sigma$ , then  $\lim E_i \in \Sigma$ .

**Theorem 3.2.** *The smallest monotone class containing an algebra is the  $\sigma$  algebra generated by the algebra.*

*Proof.* If  $\Sigma$  is an algebra, then it is a  $\pi$  system. The  $\pi$ – $\lambda$  theorem says that the smallest  $\lambda$  system containing  $\Sigma$  is the  $\sigma$  algebra containing the  $\pi$  system. But if  $\Sigma_*$  is a monotone class containing  $\Sigma$ , then it is almost a  $\lambda$  system, except that we may not have  $B - A \in \Sigma_*$  if  $A \subset B$  are both members of  $\Sigma_*$ . We show that this does hold if  $\Sigma_*$  contains  $\Sigma$ . 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  $F_i \rightarrow F$  (upward or downward), then  $E - (E \cap F_i) \rightarrow E - (E \cap F)$  and  $F_i - (E \cap F_i) \rightarrow F - (E \cap F)$ . Now if  $E \in \Sigma$ , then  $X_E$  and  $Y_E$  both contain  $\Sigma$ , 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  $\Sigma$ , so  $\Sigma_* \subset Y_E$ . This shows that  $\Sigma_*$  is a  $\lambda$  system, and therefore that  $\Sigma_*$  contains the  $\sigma$  algebra generated by  $\Sigma$ .  $\square$

We can use the monotone class theorem to understand the structure of the product  $\sigma$  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*



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 smallest  $\lambda$  system containing the  $\pi$  system of elementary rectangles.

Now suppose we have a positive measure  $\mu$  defined on  $\Sigma$ , and a positive measure  $\lambda$  on  $\Pi$ . 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  $\lambda$  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 Caratheodory extension theorem guarantees that  $\lambda \otimes \mu$  extends to a measure on  $\Sigma \otimes \Pi$ . Provided that we are working over a  $\sigma$  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  $\sigma$  finite spaces.

**Theorem 3.3.** *If  $X$  and  $Y$  are  $\sigma$  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  $E_1 \subset E_2 \subset \dots \rightarrow E$ , then  $\lambda((E_i)_x) \rightarrow \lambda(E_x)$  monotonely upward, and  $\mu(E_i^y) \rightarrow \mu(E^y)$  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  $E_1 \supset E_2 \supset \dots \rightarrow E$ , 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  $\lambda(E_x) \leq \lambda(Y)$ ,  $\mu(E^y) \leq \mu(X)$ , 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  $\sigma$  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  $X_i \rightarrow X$ ,  $Y_i \rightarrow Y$ , and apply monotone convergence.  $\square$

**Theorem 3.4** (Tonelli). *If  $X$  and  $Y$  are  $\sigma$  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  $\alpha, \beta \geq 0$ . 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.  $\square$

**Corollary 3.5.** *If  $f$  is a complex measurable function on  $X \times Y$ , where  $X$  and  $Y$  are both  $\sigma$  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 3.6** (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  $f^-$ .  $\square$

### 3.1 Completion of Product Measures

Just because  $\Sigma$  and  $\Pi$  are complete  $\sigma$  algebras with respect to  $\mu$  and  $\lambda$ , 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 3.7.** *The Lebesgue measure on  $\mathbf{R}^{n+m}$  is the completion of the product of the Lebesgue measures on  $\mathbf{R}^n$  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  $\mathbf{R}^n$  and  $\mathbf{R}^m$ . But this is a  $\pi$  system generating the Borel  $\sigma$  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.  $\square$

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

**Theorem 3.8.** *If  $X$  and  $Y$  are complete  $\sigma$  finite measure spaces with  $\sigma$  algebras  $\Sigma$  and  $\Pi$ , 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  $\Sigma$  measurable for almost all  $x$ ,  $f_y$  is  $\Pi$  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 3.9.** *If  $\mu$  is a positive measure on  $\Sigma$ , and  $f$  is  $\Sigma^*$  measurable, then there is a  $\Sigma$  measurable function  $g$  with  $f = g$  almost everywhere.*

*Proof.* Suppose  $f \geq 0$ . Then there are non-negative  $\Sigma^*$  measurable simple functions  $s_0 \leq s_1 \leq \dots$  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  $\Sigma$  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  $\Sigma$  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 3.10.** *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  $\Sigma$  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  $\Pi$  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

$$\begin{aligned} E_2 \times F_2 - E_1 \times F_1 &= (E_2 - E_1) \times (F_2 \cap F_1) \\ &\quad \cup (E_2 \cap E_1) \times (F_2 - F_1) \\ &\quad \cup (E_2 - E_1) \times (F_2 - F_1) \end{aligned}$$

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

## 3.2 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  $\sigma$  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  $\mu$  on the Lebesgue measurable subsets  $(0, \infty)$  by the equation  $d\mu = r^{d-1} dr$ .

**Lemma 3.11.** *If  $\pi : (0, \infty) \times S^{d-1} \rightarrow \mathbf{R}^d - \{0\}$  is defined by  $\pi(r, x) = rx$ , then  $\pi$  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  $\pi$  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  $\pi$  system generates all Borel subsets of  $\mathbf{R}^d$ , and therefore  $\pi_*(\mu \otimes \sigma)$  agrees with the Lebesgue measure on all Borel subsets.  $\square$

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 similarly 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 3.12.** *If  $f : \mathbf{R}^d \rightarrow \mathbf{R}$  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  $f : \mathbf{R}^d - \{0\} \rightarrow \mathbf{R}$  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  $\pi$  is Borel measurable, and we

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

$$\begin{aligned}
\int_{\mathbf{R}^d} f(x) dx &= \int_{\mathbf{R}^d - \{0\}} f(x) dx = \int_{\mathbf{R}^d - \{0\}} f(x) d\pi_*(\mu \otimes \sigma)(x) \\
&= \int_{(0, \infty) \times S^{d-1}} (f \circ \pi)(r, x) d(\mu \otimes \sigma)(r, x) \\
&= \int_{S^{d-1}} \int_0^\infty f(rx) d\mu(r) d\sigma(x) \\
&= \int_{S^{d-1}} \int_0^\infty f(rx) r^{d-1} dr d\sigma(x)
\end{aligned}$$

If  $f$  is only Lebesgue measurable, then  $f \circ \pi$  is Lebesgue measurable when we *complete* the  $\sigma$  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  $f^r$ .  $\square$



# Chapter 4

## Integration

**Lemma 4.1** (Fatou). *If  $f_1, f_2, \dots \geq 0$ , then*

$$\liminf \int f_n \leq \int \liminf f_n$$

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 4.2** (Scheffe). *If  $f, f_1, f_2, \dots \in L^1$ , and  $f_k \rightarrow f$  pointwise almost everywhere, then*

$$\int |f_n - f| \rightarrow 0 \quad \text{iff} \quad \int |f_n| \rightarrow \int |f|$$

*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  $\min(f_n, f) \leq f$ , and  $\min(f_n, f) \rightarrow f$  almost everywhere, so the dominated convergence theorem guarantees that

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

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

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

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  $\int |f_n| \rightarrow \int |f|$ , 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

$$\begin{aligned} \int f^- + \int f^+ &= \liminf \int |f_n| \leq \liminf \int f_n^- + \liminf \int f_n^+ \\ &\leq \limsup \int f_n^- + \limsup \int f_n^+ \leq \int f^- + \int f^+ \end{aligned}$$

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  $\int |f^+ - f_n^+| \rightarrow 0$  and  $\int |f^- - f_n^-| \rightarrow 0$ , and

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

and so  $\int |f - f_n| \rightarrow 0$ . On the other hand, if  $\int |f - f_n| \rightarrow 0$ , 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.  $\square$

**Exercise 4.1 (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

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

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

□

# Chapter 5

## Interpolation

Interpolation is a core part of the ‘hard’ style of analysis, crunchy quantitative estimates that give strict bounds on quantities. One basic tool here is interpolation, which ‘in essence’ 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  $0 \leq t \leq 1$ .

The most basic example occurs in the theory of real numbers. Suppose  $0 \leq A_0 \leq B_0$  and  $0 \leq A_1 \leq B_1$ . 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  $A_t \leq B_t$ , for the power function  $x \mapsto x^a$  is order preserving for  $a > 0$ . For  $t = 1/2$ , we obtain the geometric mean inequality

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

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.

### 5.1 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  $A \leq B_t$  over all  $0 \leq t \leq 1$ , and this bound can be refined to

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

for  $\varepsilon < \min(t, 1 - t)$ . This tells us that the bound  $A \leq B_t$  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 negligible for  $|n| \gg 0$ .

The inequality  $A_t \leq B_t$  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. Similarly,

## 5.2 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 6

# Differentiation and Averages

If  $f : [a, b] \rightarrow \mathbf{R}$  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 maximal functions introduced by Hardy and Littlewood, which is a general strategy for establishing regularity results for various operators encountered in analysis.

### 6.1 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 continuous, because, as  $f$  is integrable, for every  $\varepsilon > 0$ , there is  $\delta > 0$  such that if  $|E| \leq \delta$ ,  $\int_E |f| \leq \varepsilon$ , and this means that for every

$$x \in [a, b - \delta],$$

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

Thus  $F$  is *uniformly* continuous. 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  $f : \mathbf{R}^d \rightarrow \mathbf{R}$ ,

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

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 6.1.** *If  $f$  is continuous at  $x$ , then*

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

*Proof.* If  $\varepsilon$  is fixed, and  $\delta$  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

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

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**

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

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

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

*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  $B'$  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

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

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



**Theorem 6.3.** *The function  $f^*$  is measurable, and for any  $\alpha > 0$ ,*

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

*Thus  $\|f^*\|_{1,w} \leq 3^d \|f\|_1$ , 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  $\alpha$ , 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  $\alpha$ . 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

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

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

**Theorem 6.4** (Lebesgue Differentiation Theorem). *If  $f$  is integrable on  $\mathbf{R}^d$ , then for almost every  $x$ ,*

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

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

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

is a set of measure zero. If we fix  $\varepsilon$ , 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 converge, and therefore

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

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

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

But  $\varepsilon$  is arbitrary, so  $F_\alpha$ ,  $G_\alpha$ , and therefore  $E_\alpha$ , are sets of measure zero.  $\square$

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  $f$ .

## 6.2 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 density 1** if

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

This means that for every  $\alpha < 1$ , for suitably small balls, we conclude that  $|B \cap E| \geq \alpha|B|$ , 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 6.5** (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

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

then a union bound gives

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

as  $\alpha \rightarrow 1$ ,  $2\alpha - 1 \rightarrow 1$ , 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  $\alpha E$  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 than just the Lebesgue measure.

**Theorem 6.6.** *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

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

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  $x$  converge to  $f(x)$ .

**Theorem 6.7.** *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  $x \in E_p^c$ ,

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

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  $\varepsilon$ , there is a rational  $p$  such that  $|f(x) - p| < \varepsilon$ , and we know the equation above holds, so

$$\begin{aligned} \lim_{\substack{|B| \rightarrow 0 \\ x \in B}} \frac{1}{|B|} \int_B |f(y) - f(x)| \, dy \\ \leq \limsup_{\substack{|B| \rightarrow 0 \\ x \in B}} \frac{1}{|B|} \int_B |f(y) - p| + |p - f(x)| \, dy \leq 2\varepsilon \end{aligned}$$

we can then let  $\varepsilon \rightarrow 0$ . 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  $g$  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.

## 6.3 Generalizing The Differentiation Theorem

A family of sets  $U_\alpha$  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  $U_\alpha$ , there is a ball  $B$  with  $x \in B$ ,  $U_\alpha \subset B$ , and  $|U_\alpha| \geq c|B|$ . Thus  $U_\alpha$  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  $x, y \in \mathbf{R}^d$ ,

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

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 6.8.** 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  $f$ ,

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

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

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

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

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

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

## 6.4 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  $\alpha$ ,

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

and if for every  $\delta > 0$ ,

$$\int_{|x| \geq \delta} |K_\alpha(x)| \, dx \rightarrow 0$$

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

**Theorem 6.9.** *If  $K_\alpha$  is a good kernel, then for any absolutely integrable function  $f$ ,  $f * K_\alpha \rightarrow f$  in the  $L^1$  norm, and  $(f * K_\alpha)(x) \rightarrow f(x)$  for every  $x$  which is a point of continuity of  $f$ .*

*Proof.* Note that

$$\begin{aligned} \|(f * K_\alpha) - f\|_1 &= \int |(f * K_\alpha)(x) - f(x)| dx \\ &= \int \left| \int K_\alpha(y) [f(x-y) - f(x)] dy \right| dx \\ &\leq \int |K_\alpha(y)| \|T_y f - f\|_1 dy \end{aligned}$$

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

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

Since  $\|K_\alpha\|_1$  is universally bounded over  $\alpha$ , we can let  $\varepsilon \rightarrow 0$  to obtain convergence. If  $x$  is a fixed point of continuity, and for a given  $\varepsilon > 0$ , we pick  $\delta > 0$  with  $|f(y) - f(x)| \leq \varepsilon$  for  $|y - x| < \delta$ , then

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

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

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  $\delta$ . These assumptions are stronger than being a good kernel, because if  $K_\delta$  is an approximation to the identity, then

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

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

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

This calculation also implies

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

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

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

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

**Example.** If  $\varphi$  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

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

Because  $\varphi$  is bounded, we find

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

Now  $K_\delta$  is supported on a disk of radius  $\delta$ , this bound also shows

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

and so  $K_\delta$  is an approximation to the identity. If  $\varphi$  is an arbitrary integrable function, then  $K_\delta$  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

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

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

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

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  $\delta \rightarrow 0$ , we may think of the  $K_\delta$  as ‘tending to the unit mass’ Dirac delta function  $\delta$  at the origin.  $\delta$  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 functions  $K_\delta$ .

**Theorem 6.10.** *If  $\{K_\delta\}$  is an approximation to the identity, and  $f$  is integrable on  $L^1(\mathbf{R}^d)$ , then  $(f * K_\delta)(x) \rightarrow f(x)$  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

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

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

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

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

## 6.5 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  $x : [a, b] \rightarrow \mathbf{R}^d$  parameterizes a continuous curve in the plane, then, for a given partition  $P = a \leq t_0 \leq \cdots \leq t_n$ , we can consider an approximate length

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

If  $x$  has a reasonable notion of length, then the straight lines between  $x(t_{k-1})$  and  $x(t_k)$  should be shorter than the length of  $x$  between  $t_{k-1}$  and  $t_k$ . 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 its variation on a partition  $P = a \leq t_0 < \cdots < t_n \leq b$  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  $P$ ,  $V_P(f) \leq M$ . 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 6.11.** *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  $x, y \in \mathbf{R}^d$ ,

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

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

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

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

**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  $|f'| \leq M$ , then for all  $x, y \in [a, b]$ ,*

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

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

**Example.** *Set*

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

*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

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

The **negative variation** is

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

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

**Lemma 6.12.** *If  $f$  is real-valued and has bounded variation on  $[a, b]$ , then for all  $a \leq x \leq b$ ,*

$$\begin{aligned} f(x) - f(a) &= P_{[a,x]}(f) - N_{[a,x]}(f) \\ V(f) &= P_{[a,b]}(f) + N_{[a,b]}(f) \end{aligned}$$

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

$$\begin{aligned} \left| P_{[a,x]}(f) - \sum_{f(t_i) \geq f(t_{i-1})} f(t_i) - f(t_{i-1}) \right| &< \varepsilon \\ \left| N_{[a,x]}(f) + \sum_{f(t_i) \leq f(t_{i-1})} f(t_i) - f(t_{i-1}) \right| &< \varepsilon \end{aligned}$$

It follows that

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

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

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 6.13.** *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 function of  $x$ , so for continuity on the left it suffices to prove that for each  $x$  and  $\varepsilon$ , there is  $x_1 < x$  such that  $V_{[a,x_1]}(f) \geq V_{[a,x]}(f) - \varepsilon$ . If we consider a partition

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

where  $|V_P(f) - V_{[a,x]}(f)| \leq \varepsilon$ , 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

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

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

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

**Lemma 6.14** (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  $f(a_i) \leq f(b_i)$ .*

**Theorem 6.15.** *If  $f$  is an increasing, continuous function, then  $f$  is differentiable almost everywhere. That is,*

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

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

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

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

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

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.  $\square$

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  $f : [0, 1] \rightarrow [0, 1]$  with  $f(0) = 0$ ,  $f(1) = 1$ , but with  $f'(x) = 0$  almost everywhere.  $f$  has bounded variation, but

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

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  $[2/3, 1]$ . Similarly, 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.

## 6.6 Absolute Continuity

A function  $f : [a, b] \rightarrow \mathbf{R}$  is **absolutely continuous** if for every  $\varepsilon > 0$ , there is  $\delta > 0$  such that whenever  $(a_1, b_1), \dots, (a_n, b_n)$  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 6.16.** *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  $B_1, \dots, B_n$  in the covering such that*

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

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

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

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

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

If we pick a compact set  $K_1 \subset E_1$  with  $|K_1| \geq \delta/2$ , then if we remove all sets in the Vitali covering which intersect  $B_1, \dots, B_{n_0}$ , 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  $B_1^1, \dots, B_{n_1}^1$  with  $\sum |B_k^1| \geq 3^{-d} |K_1|$ . Then  $\sum |B_k| + \sum |B_k^1| \geq 2(3^{-d} \delta)$ . 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  $E_2 = K_1 - \bigcup B_k^1$ . 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  $k \geq (|E| - \delta)3^d/\delta$ , then the family of open sets satisfies the requirements of the theorem.  $\square$

**Corollary 6.17.** *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

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

and this gives the required bound.  $\square$

**Theorem 6.18.** *If  $f : [a, b] \rightarrow \mathbf{R}$  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

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

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

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

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

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

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

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



so applying the triangle inequality,

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

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

**Theorem 6.19.** *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.  $\square$

## 6.7 Differentiability of Jump Functions

We now consider the differentiability of not necessarily continuous monotonic 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  $f(x-) \leq f(x) \leq f(x+)$ . If there is a discontinuity, this means we are forced to have a ‘jump discontinuity’ where  $f$  skips an interval. This implies 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

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

so the series defining  $J$  converges absolutely and uniformly.

**Lemma 6.20.** *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  $\sum_y \Delta f(y) j_y(x) \rightarrow J(x)$  uniformly, so we conclude that  $J$  is continuous 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  $x_1, x_2, \dots$ , 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

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

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.  $\square$

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 6.21.**  *$J'$  exists and vanishes almost everywhere.*

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

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

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

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

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

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

by finitely many points. Using inner regularity, find a compact set  $K \subset E'$  with  $|K| \geq \delta/2$ . For each  $x \in K$ , we can find intervals  $(\alpha_x, \beta_x)$  upon which  $J_0(\beta_x) - J_0(\alpha_x) \geq \varepsilon |\beta_x - \alpha_x|$ . But applying the elementary Vitali covering lemma, we can find a disjoint family of such intervals with  $\sum (\beta_{x_i} - \alpha_{x_i}) \geq |K|/3 \geq \delta/6$ . But now we find

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

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

## 6.8 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 6.22.** *If a curve is parameterized by absolutely continuous functions  $x$  and  $y$  on  $[a, b]$ , then the curve is rectifiable, and has length*

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

## 6.9 The Lebesgue Stieltjes Integral

We now use our study to provide a correspondence between positive Borel measures on  $\mathbf{R}$  which are finite on intervals, and non-decreasing functions on the real line. Given a measure  $\mu$ , 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  $\mu$  with  $F_\mu = F$ .

**Theorem 6.23.** *If  $F$  is an increasing real-valued right continuous function on  $\mathbf{R}$ , then there exists a unique Borel measure  $\mu$  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 infimum 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  $\mu^*((a, b]) \leq F(b) - F(a)$ . Next, suppose  $(a, b]$  is covered by  $\bigcup (a_i, b_i]$ . Fix  $\varepsilon > 0$ . If we use the right continuity of  $F$  to pick  $b'_i > b_i$  such that  $F(b'_i) - F(b_i) \leq \varepsilon/2^i$ , then

$$\sum F(b'_i) - F(a_i) \leq \sum F(b_i) - F(a_i) + \varepsilon$$

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

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

Taking  $\varepsilon \rightarrow 0$  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

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

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

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  $\mu$  is the Borel measure calculated from the theorem above, and  $f$  is integrable with respect to  $\mu$ . Because of how we defined  $\mu$ , 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  $\mu$  if  $\sum \alpha_k |f(x_k)| < \infty$ . A particular case occurs when  $F$  is the **Heaviside step function**  $H(x) = \mathbf{I}(x \geq 0)$ , in which case

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

so  $dH$  is the Dirac delta function.

**Theorem 6.24.** *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  $\mu$  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  $\pi$  system generating all the Borel sets, so we find that  $dF = F'(x) dx$  for all Borel sets, and therefore the integrals above agree.  $\square$

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

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

*Proof.* Let  $F$  be associated the measure  $\mu$  and  $G$  associated the measure  $\nu$ , 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.  $\square$

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

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

*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

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

and then the theorem is obvious. Suppose  $f, g \in A$ . Then  $(f \circ G)(g \circ G)$  has finite variation, because if  $P = \{x_0, \dots, x_n\}$  is any partition, then we can

find  $M$  such that  $f \circ G, g \circ G \leq M$  on  $[a, b]$ , and so

$$\begin{aligned} & \sum |g(G(x_i))f(G(x_i)) - g(G(x_{i-1}))f(G(x_{i-1}))| \\ & \leq \sum |g(G(x_i)) - g(G(x_{i-1}))||f(G(x_i))| \\ & \quad + |g(G(x_{i-1}))||f(G(x_i)) - f(G(x_{i-1}))| \\ & \leq M[V_{[a,b]}(g \circ G) + V_{[a,b]}(f \circ G)] \end{aligned}$$

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

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

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

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

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.  $\square$

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

$$\begin{aligned} f(G(b)) - f(G(a)) &= \int_a^b \partial_i f(G(s-)) dG_s^i \\ &\quad + \sum_{a < s \leq b} \{\Delta f(G)(s) - \partial_i f(G(s-))\Delta G_s^i\} \end{aligned}$$

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.



## 6.10 Representation of Complex Measures

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

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

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 strategy 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  $\mu$  by a positive measure  $\nu$ , in the sense that for every set  $E$ ,  $|\mu(E)| \leq \nu(E)$ . We shall find that we can find a *minimal*  $\nu$  bounding  $\mu$ . If  $E$  has a partition into measurable sets  $E_1, E_2, \dots$ , then we must have

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

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  $\mu$ , or, to be specific, the total variation measure. Besides being a measure, we also find that  $|\mu|(X) < \infty$ , which implies that the values  $|\mu(E)| \leq |\mu|(X)$  lie within a compact disk in the plane, which means  $\mu$  has **bounded variation**.

**Theorem 6.27.**  $|\mu|$  is a positive measure on  $X$  with  $|\mu|(X) < \infty$ .

*Proof.* Let  $E = \bigcup E_i$ , where the  $E_i$  are disjoint. Let  $t_1, t_2, \dots$  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  $|\mu|(E) \geq \sum |\mu|(E_i)$ . 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

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

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

**Lemma 6.28.** *Every finite set of complex numbers  $z_1, \dots, z_n$  has a subset  $z_{i_1}, \dots, z_{i_m}$  such that*

$$|\sum z_{i_k}| \geq \frac{1}{\pi} \sum |z_k|$$

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

**Theorem 6.29.**  $|\mu|(X) < \infty$ .

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

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

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  $(A_1, B_1), (A_2, B_2), \dots$  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  $|\mu|(E) < \infty$ .  $\square$

If  $\mu$  and  $\lambda$  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  $\mu$  such that  $\mu(E) \in \mathbf{R}$  for all measurable sets  $E$ , in which case we say  $\mu$  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  $\mu$ . The decomposition  $\mu = \mu^+ - \mu^-$  is known as the **Jordan decomposition** of  $\mu$ .

We say that a positive measure  $\lambda$  is **absolutely continuous** with respect to a positive measure  $\mu$  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  $\lambda$  is **concentrated on  $A$**  (this is equivalent to saying  $\lambda(E) = 0$  if  $E$  is disjoint from  $A$ ). If two measures  $\mu$  and  $\lambda$  are concentrated on disjoint sets, we say they are **mutually singular**, and write  $\mu \perp \lambda$ .

**Theorem 6.30.** *Suppose  $\mu$  is a positive measure, and  $\lambda, \lambda_1$ , and  $\lambda_2$  are also measures, then*

- *If  $\lambda$  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. □

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 6.31.** *If  $\mu$  is a positive  $\sigma$ -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  $E_1, E_2, \dots$ , with  $\mu(E_i) < \infty$ . By taking successive unions, we may assume  $\mu(E_i) \neq 0$  for all  $i$  (or  $\mu$  is already a finite measure, and we can define  $w = 1/2$ ). Define

$$w(x) = \sum \mathbf{I}(x \in E_k) \frac{1}{\mu(E_k)2^k}$$

Then  $w$  is the limit of step functions, hence measurable, and it is easy verified to be in  $L^1(X)$ .  $\square$

Thus the measures  $\mu$  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 6.32** (Radon-Nikodym). *Let  $\mu$  be a positive  $\sigma$  finite measure on a  $\sigma$  algebra  $X$ , and let  $\lambda$  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  $\lambda$ , with  $\lambda_a \ll \mu$ , and  $\lambda_s \perp \mu$ . If  $\lambda$  is positive and finite, then so too are  $\lambda_a$  and  $\lambda_s$ . 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  $\lambda_a$  w.r.t  $\mu$ .*

*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  $\lambda$  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 bounded. 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

$$\left| \int f d\lambda \right| \leq \int |f| d\lambda \leq \int |f| d\varphi \leq \|f\|_{L^2(\varphi)} \sqrt{\varphi(X)}$$

It follows that integration with respect to  $\lambda$  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  $\lambda(E) \leq \varphi(E)$ , we conclude that

$$0 \leq \varphi(E)^{-1} \int_E g d\varphi \leq 1$$

so  $0 \leq g \leq 1$  almost surely with respect to  $\varphi$ . We can rearrange the equation defining  $g$  to read

$$\int f(1 - g) d\lambda = \int f g d\mu$$

Set  $A = \{0 \leq g < 1\}$ , and  $B = \{g = 1\}$ , and define

$$\lambda_a(E) = \lambda(A \cap E) \quad \lambda_s(E) = \lambda(B \cap E)$$

s

□

## Chapter 7

### 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  $\alpha$  such that  $nX$  is the union of  $n^\alpha$  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  $\log_3 2 = 0.6309\dots$ ’.