

# Measure Theory

Jacob Denson

August 3, 2019

# Table Of 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	The Invention of the Cantor Set . . . . .	28
1.5	Symmetries of the Lebesgue Measure . . . . .	29
1.6	The Borel-Cantelli Lemma . . . . .	30
1.7	Banach Tarski and Nonmeasurable Sets . . . . .	31
<b>2</b>	<b>Lebesgue Measurable Functions</b>	<b>34</b>
2.1	Monotone Rearrangement Invariant Norms . . . . .	37
2.2	The $L^p$ norms . . . . .	38
2.3	Weak and Lorentz Norms . . . . .	46
2.4	Metric Structure of Norms . . . . .	52
<b>3</b>	<b>Differentiation and Averages</b>	<b>55</b>
3.1	Maximal Functions and Differentiation . . . . .	56
3.2	Lebesgue Density Theorem . . . . .	60
3.3	Generalizing The Differentiation Theorem . . . . .	62
3.4	The Calderon Zygmund Decomposition . . . . .	63
3.5	Approximations to the Identity . . . . .	65
3.6	Differentiability of Measurable Functions . . . . .	70
3.7	Absolute Continuity . . . . .	77
3.8	Differentiability of Jump Functions . . . . .	83
3.9	Rectifiable Curves . . . . .	85
3.10	Bounded Variation in Higher Dimensions . . . . .	87
3.11	Minkowski Content . . . . .	89
3.12	The Isoperimetric Inequality . . . . .	92

<b>4</b>	<b>Measure Extension</b>	<b>95</b>
<b>5</b>	<b>Product Spaces</b>	<b>100</b>
5.1	Completion of Product Measures . . . . .	104
5.2	Integration in Polar Coordinates . . . . .	106
<b>6</b>	<b>Integration</b>	<b>109</b>
<b>7</b>	<b>The Lebesgue Stieltjes Integral</b>	<b>112</b>
7.1	Representation of Complex Measures . . . . .	117

# Chapter 1

## The Lebesgue Measure

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

- If a shape can be cut up into finitely many components, and then rearranged into the form of a different shape by a series of rigid motions, then the new shape has the same area as the old.
- We can obtain upper bounds on the area of the shape by enveloping the shape by another shape with an already known area, and lower bounds by finding the area of shapes enclosed by the shape.

Naive applications of these techniques sufficed for much of the classical era of mathematics. Nonetheless, with the advent of Cartesian coordinates and the subsequent modern introduction of set theory, shapes in the Euclidean plane are identified with subsets of  $\mathbf{R}^2$ . Set theory enables us to obtain rather strange subsets of  $\mathbf{R}^2$ , for which the methods of Archimedes 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 we call the general study of these objects *measure theory*. In line with this unification, we will let  $|S|$ , or  $\mu(S)$ , denote the length, area, volume, and higher dimensional measures of a given shape  $S$  in  $\mathbf{R}^d$ , regardless of the dimension.

*Remark.* 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 higher dimensional volumes of even classical geometric sets.

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

- Given a continuous planar curve, it is natural to try and measure its length. In 1890, Peano constructed a continuous curve mapping the unit interval  $[0,1]$  surjectively onto the unit rectangle  $[0,1]^2$ . It is suspect as to whether we can measure the length of this curve, because it is constructed as the limit of curves whose lengths tend to infinity. Thus a more particular analysis is required, and this required the theory of measure. Curves without a well defined length are not purely artificial, for if we take a typical continuous curve obtained by following the random motion of a dust particle in air, then we find the path does not have a reasonable length, because there are too many ‘bends’ in the path. In the mid 20th century, difficulties were found when measuring the length of the perimeter of countries, which are too ‘jagged’ to have a well defined length.
- 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 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:

- 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  $|S_i| \rightarrow |S|$ .

What is useful is that you have to try very hard to find a non measurable set. Almost every set one ever considers in mathematics is measurable, so the theory is suitably general to be used in applications. We will begin by extending the notion of sets to fairly general subsets of  $\mathbf{R}^d$ . We do this not only because this is classically how measure theory was introduced, but also because it brings to light the many intricate parts of the theory which we consider when we build measures on more general ‘measure spaces’, which is required for understanding the modern theory of probability and certain other ‘size’ metrics which occur in analytical mathematics, most importantly, the modern foundations of probability theory.

## 1.1 Measuring Elementary Sets

Let’s begin by using basic ideas of Euclidean geometry to find a basic class of sets, which are suitably elementary that the paradoxes we have seen above cannot occur. It is unarguable that the length of an interval with start point  $a$  and end point  $b$ , whether closed, open, or half open, is  $b - a$ . We know from elementary geometry that the area of a rectangle is the product of the length of the intervals which define it, and if we are suitably relaxed, we can take this as the definition of the volume of a rectangle in higher dimensions. That is, if  $I = I_1 \times \cdots \times I_d$  is a box in  $\mathbf{R}^d$ , in the sense that it is the product of intervals  $I_i$  between  $a_i$  and  $b_i$ , then we define  $|I| = (b_1 - a_1) \cdots (b_d - a_d) = |I_1| \cdots |I_d|$ . The easiest sets to measure are those covered by finitely many *disjoint* boxes, and we will show this family has a well defined theory of area.

**Lemma 1.1.** *If a box  $R$  is a disjoint union of boxes  $R_i$ , then  $|R| = \sum |R_i|$ .*

*Proof.* We proceed by means of a grid decomposition, which refines the objects in the problem in such a way that the proof of the proposition is trivial. Write  $R = I_1 \times \cdots \times I_d$ . Suppose first that the rectangular decomposition forms a grid, in the sense that we can index the decomposition as  $R_{i_1 \dots i_d}$ , 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  $I_{i+1}^k$ . Then we find

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

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 proves the following.

**Lemma 1.2.** *If  $R_1, \dots, R_n$  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  $\mathbf{R}^d$  by the formula

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



However, this definition runs into problems for more general sets than intervals. One can find sets where this limit doesn't exist, and even if the limit does exist, we might not even have translation invariance;  $\mathbf{Q} \cap [0, 1]$  has length 1 with respect to this metric, whereas  $\mathbf{Q} \cap [0, 1] + \sqrt{2}$  has length 0. However, for the class of Jordan measurable sets, which we will soon discuss, this definition works. A more suitable way to obtain a continuous uniform measure from some discrete quantity is by the theory of Monte Carlo 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 unarguably *the* choice for the volume of this family of sets. We call such a set  $E$  an **elementary set**. We will use the unarguable measure of elementary sets to approximate the measure of arbitrary subsets of the plane.

The measure of size on elementary sets have the following fundamental properties, which are *very important* going forward. One can summarize the development of measure theory as trying to measure the most general sets possible, while still preserving these properties:

- If  $E = E_1 \cup \dots \cup E_N$  is the union of disjoint elementary sets, then  $|E| = |E_1| + \dots + |E_N|$ . This is the property of *finite additivity*. If the union is not disjoint, then we have *subadditivity*, i.e.  $|E| \leq |E_1| + \dots + |E_N|$ .
- The measure is *translation invariant*, so  $|E + x| = |E|$  for any elementary set  $E$  and  $x \in \mathbf{R}^d$ .
- 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  $F$ , 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 an elementary set up to a scalar factor. Since all the properties are intuitive to us, this tells us that our construction is going in the right direction!

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

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

- (a) *If  $E$  is the finite union of boxes, then  $E$  is the finite union of disjoint boxes. Thus the union of two elementary sets is also an elementary set.*
- (b) *The intersection of two boxes is a box, and the intersection of two elementary sets is an elementary set.*
- (c)  *$R - S$  is an elementary set, and so the difference of two elementary sets is an elementary set. The family of elementary sets is not closed under the complement operation, but this proposition says the family is ‘almost’ closed under this operation.*
- (d)  *$x + E$  is an elementary set, so the family of elementary sets is translation invariant.*
- (e) *The symmetric difference  $E \triangle F = (E - F) \cup (F - E)$  is an elementary set.*
- (f) *If  $E \subset \mathbf{R}^d$  and  $F \subset \mathbf{R}^s$ , then  $E \times F$  is an elementary subset of  $\mathbf{R}^{d+s}$ , and  $|E \times F| = |E||F|$ .*

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

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

which is a box. If  $E = \bigcup R_i$ , and  $F = \bigcup S_j$  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 over some set  $X$  is an **algebra** if it is closed under the union and complement operations, and contains  $\emptyset$ . 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  $\mathbf{F}_2$ , with  $\emptyset$  as the zero element. If we consider a box, and consider the family of elementary sets contained within the box, then our arguments above result in showing this family forms an algebra. Nonetheless, the family of all elementary subsets in  $\mathbf{R}^d$  is not an algebra, and this is one of the reasons for us to enlarge the family of sets we can measure.

## 1.2 Jordan Measurable Sets

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

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

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  $\int_R f(x) dx$ . Applying this technique to our scenario, if  $E$  is a bounded set, contained in some box  $R$ , then the area of  $E$  can be defined by the formula

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

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

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

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

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

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

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

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

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

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

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

*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  $\mathbf{R}^d$ , then

- (a)  $E \cup F$  is Jordan measurable.
- (b)  $E \cap F$  is Jordan measurable.
- (c)  $F - E$  is Jordan measurable. If  $E \subset F$ , then  $|F - E| = |F| - |E|$ .
- (d)  $E \Delta F$  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

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

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

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

The  $R_i$ , for  $i \in S$ , cover  $A_\delta$ , 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  $\mathbf{R}^d$ , 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  $[-1/\sqrt{d}, 1/\sqrt{d}]^d \subset B_1$ . This yields the bounds

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

These bounds are pretty crude, however. Later we will be able to establish precise calculations for the volume of a ball of radius one.

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

### 1.3 Lebesgue Measure

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Then if  $a \in \inf A$  and  $b = \sup B$ , then  $(a, b) \in U$ . Because  $U$  is open,  $a, b \notin U$ , and therefore no open interval which is contained in  $U$  contains  $(a, b)$ , so the interval is maximal. This argument shows that if we let  $\mathcal{U}$  denote the family of maximal intervals contained in  $U$ , then  $\mathcal{U}$  covers  $U$ . If  $I$  and  $J$  are maximal open intervals, then  $I$  must be disjoint from  $J$ , and therefore  $U$  is the disjoint union of the intervals in  $\mathcal{U}$ . Since each interval contains a rational number, and distinct maximal intervals are disjoint, we can index maximal intervals by rational numbers, and so the set  $\mathcal{U}$  is countable. This proves the theorem.  $\square$

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

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

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

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

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

*Proof.* Let  $U$  be an open set. For each  $m$ , we will successively define  $U_m \subset U$  by adding dyadic boxes such that  $\bigcup U_m = U$ . Set  $U_0 = \emptyset$ . 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  $R_n$ , it makes sense to define

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

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

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

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

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

Since the  $S_n$  are arbitrary, we conclude that

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

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

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

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

so the distance is the minimal distance between two points in the set.

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

*Proof.* By subadditivity, it suffices to prove that  $\mu^*(E \cup F) \geq \mu^*(E) + \mu^*(F)$ . Let  $E \cup F$  be covered by rectangles  $R_1, R_2, \dots$ . By subdividing each rectangle, we may assume that each is contained within circles of radius  $d(E, F)/2$ . Then each  $R_i$  either only contains points in  $E$ , only contains points in  $F$ , or contains points in neither set, and so we can partition the rectangles into a cover  $S_1, S_2, \dots$  of  $E$ , and a cover  $S'_1, S'_2, \dots$  of  $F$ , hence

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

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

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

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

*Letting  $N \rightarrow \infty$ , we conclude by monotonicity that*

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

*Letting  $\varepsilon \rightarrow 0$ , we conclude that  $\mu^*(E) \geq \sum |R_n|$ . But we already know that  $\mu^*(E) \leq \sum |R_n|$  by countable subadditivity, so  $\mu^*(E) = \sum |R_n|$ .*

As with the theory of Jordan measure, it is not in general true that for any two sets  $E$  and  $F$ ,  $\mu^*(E \cup F) = \mu^*(E) + \mu^*(F)$ , as Banach and Tarski warned us. This means we cannot define the Lebesgue measure  $|E|$  for all sets. For the class of Lebesgue measurable sets, we will show that we can obtain additivity. There isn't a great definition of the Lebesgue inner measure, because fractal sets may not even contain rectangles, let alone countably many. But we saw that a good alternate definition for Jordan measurability was the fact that these sets were well approximated by Jordan measurable sets, and we use this for the Lebesgue measure definition.

We say a set  $E$  is **Lebesgue measurable** if, for every  $\varepsilon > 0$ , there is an open set  $U$  with  $\mu^*(U - E) \leq \varepsilon$ . Note that for any set  $E$  with  $\mu^*(E) < \infty$ , there is an open set  $U \supset E$  with  $\mu^*(U) - \mu^*(E) \leq \varepsilon$ , but this doesn't necessarily imply  $\mu^*(U - E) \leq \varepsilon$ .

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

- (a) *Every open set is measurable.*
- (b) *If  $\mu^*(E) = 0$ , then  $E$  is Lebesgue measurable.*
- (c) *Countable unions of measurable sets are measurable.*
- (d) *Closed sets are measurable.*
- (e) *The complement of a measurable set is measurable.*
- (f) *Countable intersections of measurable sets are measurable.*
- (g) *Every Jordan measurable set is Lebesgue measurable.*

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

$$\mu^*\left(\bigcup R_i \triangle E\right) = \mu^*\left(\bigcup R_i - E\right) \leq \mu^*\left(\bigcup R_i\right) \leq \varepsilon$$

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

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

so

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

which proves (c). To prove (d), it suffices to show that a compact set  $K$  is measurable, because every closed set is the countable union of compact sets. We know  $\mu^*(K) < \infty$ , because  $K$  is contained in some open box  $R$ , and monotonicity implies  $\mu^*(K) \leq |R| < \infty$ . Consider some open set  $U$  containing  $K$  with  $\mu^*(U) - \mu^*(K) \leq \varepsilon$ . Then  $U - K$  is open, and therefore

the union of almost disjoint closed rectangles  $R_1, R_2, \dots$ . But this means that  $\mu^*(U - K) = \sum |R_i|$ , and by subadditivity,

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

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

$$\mu^*(E^c \Delta V) \leq \mu^*(E^c \Delta U^c) + \mu^*(U^c \Delta V) = \mu^*(E \Delta U) + \mu^*(U^c \Delta V) \leq 2\varepsilon$$

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

Here are some alternate characterizations of measurability.

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

- (a) *For any  $\varepsilon > 0$ , there is an open set  $U \supset E$  with  $\mu^*(U - E) \leq \varepsilon$ . This is our definition of Lebesgue measurability.*
- (b) *For every  $\varepsilon > 0$ , there is an open set  $U$  with  $\mu^*(E \Delta U) \leq \varepsilon$ .*
- (c) *For all  $\varepsilon > 0$ , there is a closed set  $F \subset E$  with  $\mu^*(E - F) \leq \varepsilon$ .*
- (d) *For all  $\varepsilon > 0$ , there is a closed set  $F$  with  $\mu^*(E \Delta F) \leq \varepsilon$ .*
- (e) *For every  $\varepsilon > 0$ , there is a Lebesgue measurable set  $F$  with  $\mu^*(E \Delta F) \leq \varepsilon$ .*

*Thus these conditions provide an alternate definition of measurability.*

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

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



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

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

and so

$$\mu^*(U \cup V \Delta E) \leq \mu^*(U \Delta F) + \mu^*(V \Delta F) + \mu^*(F \Delta E) \leq 3\varepsilon$$

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

If  $E$  is Lebesgue measurable, we define  $|E| = \mu^*(E)$  to be the **Lebesgue measure** of the set  $E$ . Unlike the value  $\mu^*(E)$ , which we view as a potential candidate for measuring the size of a set, which may fail for sets which aren't well approximated, the Lebesgue measure has proven itself to be *the* correct definition of length, area, and volume in mathematical analysis. We now show that that Lebesgue measurable sets cannot be entangled enough for countable additivity to break (they reach all depths of the set, so that features of the sets not reached by the measure cannot be added up countably to obtain a significant feature of the set).

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

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

we can choose  $N$  large enough that

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

If we now consider the elementary set  $R = R_1 \cup \dots \cup R_N$ , then  $\mu^*(U - R) \leq \varepsilon$ , and so  $\mu^*(E \Delta R) \leq 2\varepsilon$ . Thus, with respect to area, the Lebesgue measurable

sets are those which can be ‘well pixelated’ by elementary sets. This is different from Jordan measurable sets, which can be both well encapsulated by elementary sets, and cover an elementary set well – here we allow  $R$  to contain new points, and to avoid certain points of  $E$ .

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

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

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

We define the families  $G_\delta$ , and  $F_\sigma$ , which consist of countable intersections of open sets, and countable unions of closed sets, respectively. Our proof above shows not only that Lebesgue measurable sets differ from Borel sets by a set of measure zero, but that they differ from a  $G_\delta$  or an  $F_\sigma$  set by a set of measure zero. This is just one of many properties of the Lebesgue measure which can be encompassed under the banner ‘measurable sets are just fuzzier open/closed sets’. Here’s another example, which shows measurable sets contain an arbitrary large fraction of some interval.

**Theorem 1.15.** *If  $E$  is a measurable subset of  $\mathbf{R}$ , with  $|E| > 0$ , then for any  $\alpha < 1$  there exists an interval  $I$  with  $|E \cap I| \geq \alpha|I|$ .*

*Proof.* Consider a cover of  $E$  by intervals  $I_1, I_2, \dots$  with

$$\sum_{n=1}^{\infty} |I_n| - |E| \leq \varepsilon$$

Now if there exists  $\alpha$  such that for any interval  $I$ ,  $|I \cap E| \leq \alpha |I|$  then

$$|E| = \sum_{n=1}^{\infty} |E \cap I_n| \leq \alpha \sum_{n=1}^{\infty} |I_n|$$

so

$$\sum_{n=1}^{\infty} |I_n| - |E| \geq (1 - \alpha) \sum_{n=1}^{\infty} |I_n| \geq (1 - \alpha) |E|$$

Thus  $(1 - \alpha)|E| \leq \varepsilon$ , and so by taking  $\varepsilon \downarrow 0$  we conclude that  $|E| = 0$ .  $\square$

Note that by splitting the cover of intervals up into smaller and smaller intervals in the argument above, we may assume each interval is centered at a point in  $E$ , for otherwise the resultant intervals may be discarded from the cover. Furthermore, since we can always assume the intervals above have rational length, we can assume the interval well covered by  $E$  has rational length. But if we now take an interval  $I$  such that  $E$  contains a fraction  $\alpha$  of it, and split  $I$  into smaller intervals, then an  $\alpha$  fraction of one these interval must be covered by  $E$ . In particular, if two sets  $E$  and  $F$  are considered, by splitting up two rational  $\alpha$  intervals for  $E$  and  $F$  into comparable ratios, we may assume that there is a common interval  $I$  centered at the origin, with  $x \in E$ ,  $y \in F$ , and

$$|E \cap (x + I)|, |F \cap (y + I)| \geq \alpha |I|$$

This will help us in the corollary below.

**Corollary 1.16.** *If  $E$  and  $F$  have positive measure, then  $E + F$  contains an open interval. In particular, if  $E$  has positive measure, then the difference set  $E - E = \{x - y : x, y \in E\}$  contains an open interval about the origin.*

*Proof.* More generally, if  $E$  and  $F$  are sets of positive measure, and choose  $x \in E$ ,  $y \in F$ , and an interval  $I$  centered at the origin such that  $|E \cap (I + x)|, |F \cap (I + y)| \geq \alpha |I|$ . Suppose there are arbitrarily small  $t$  such that  $(x + y) + t \notin E + F$ . Then for each  $u$ , only one of  $x + u \in E$  and  $y + (t - u) \in F$

can hold at a given time. But this means that  $E - x$  and  $(t + y) - F$  are disjoint. Thus

$$\begin{aligned}
|I| &\geq |I \cap (E - x)| + |I \cap (t + y) - F| \\
&= |I \cap (E - x)| + |I + t + (F - y)| \\
&\geq |I \cap (E - x)| + |I + (F - y)| - 2t \\
&\geq 2\alpha|I| - 2t
\end{aligned}$$

Taking  $t \downarrow 0$  shows that  $|I| \geq 2\alpha|I|$ , and so if  $\alpha > 1/2$ , we obtain a contradiction. Selecting  $x = y = 0$  in this argument in the case of the difference set shows that  $E - E$  contains an open interval around the origin.  $\square$

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

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

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

**Theorem 1.17.** *If  $E_1, E_2, \dots$  are disjoint Lebesgue measurable sets, and  $E = \bigcup E_n$ , then*

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

*So the Lebesgue measure has the property of countable additivity.*

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

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

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

and we can then let  $\varepsilon \rightarrow 0$ .  $\square$

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

$$\lim_{n \rightarrow \infty} E_n = \bigcup_{n=1}^{\infty} E_n \quad \lim_{n \rightarrow \infty} F_n = \bigcap_{n=1}^{\infty} F_n$$

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

**Theorem 1.18.** *If each  $E_n$  is a Lebesgue measurable set, then*

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

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

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

*Proof.* Define the sets  $E'_n = E_n - E_{n-1}$ , with  $E'_1 = E_1$ . Then the  $E'_n$  are disjoint and measurable, and  $\lim E_n = \bigcup E'_n$ , and  $E_N = \bigcup_{n=1}^N E'_n$ , so that

$$\lim_{N \rightarrow \infty} |E_N| = \lim_{N \rightarrow \infty} \sum_{n=1}^N |E'_n| = \sum_{n=1}^{\infty} |E'_n| = \left| \lim E_n \right|$$

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

$$|F_N| - \left| \lim F_n \right| = |F_N - \lim F_n| = \left| \lim F'_n \right| = \lim |F'_n| = |F_N| - \lim |F_n|$$

and we can rearrange this equation to get the equation required.  $\square$

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

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

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

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

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

## 1.4 The Invention of the Cantor Set

One of the strangest subsets of the real line is the Cantor set, our first example of an uncountable set with Lebesgue measure zero. The set  $C$  is obtained from the interval  $[0, 1]$  by dividing it into three parts, removing the open inner interval, and repeating this process ad infinitum on all the subintervals contained in this interval. In particular, if we start with the unit interval  $C_0 = [0, 1]$ , then we will have  $C_1 = [0, 1/3] \cup [2/3, 1]$ ,  $C_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$ , and so on and so forth, with the Cantor set  $C$  being defined as  $\lim C_n$ . The Cantor set possesses many paradoxical properties which are good for honing our intuitions. For instance, we know that all the points on the boundary of the intervals in each  $C_n$  are contained in  $C$ , and as a perfect, infinite set, it is uncountable, so it must contain more than just the points on the boundary of intervals. In fact, if one looks at how the set is constructed, we can define the points in  $C$  as those numbers  $x \in [0, 1]$  which have a base three expansion  $0.a_1a_2\dots$  with each  $a_n \in \{0, 2\}$ , i.e.

$$x = \sum_{n=1}^{\infty} \frac{a_n}{3^n}$$

The Cantor set is Jordan measurable as well as Lebesgue measurable, since  $\partial C = C$  has measure zero. To see this, we note that we can define  $C_{n+1} = C_n/3 \cup (C_n/3 + 2/3)$ . Thus we obtain the recursive formula  $|C_{n+1}| = |C_n|/3 + |C_n|/3 = (2/3)|C_n|$ , and this means  $|C_n| = (2/3)^n$ . But now we find  $|C| = \lim |C_n| = \lim (2/3)^n = 0$ . Similar properties hold for other ‘Cantor-like sets’, if, rather than moving the ‘middle third’ of the intervals at each stage, we instead remove a middle length  $\xi \in (0, 1)$ , the set obtained often denote  $C_\xi$ . However, if we vary the lengths that we cut out of the intervals appropriately, we can actually find a ‘Cantor-like’ set which has positive measure, which therefore cannot be Jordan measurable. Introducing parameters, suppose that  $X_0 = [0, 1]$ , and we remove a fraction  $l_n$  from the center of each interval in  $X_{n-1}$  to obtain the set  $X_n$ . Then we find  $|X_n| = (1 - l_n)|X_{n-1}|$ , so  $|X_n| = \prod (1 - l_n)$ . If  $l_n \downarrow 0$  suitably fast, then the limiting set  $X = \lim X_n$  will have positive measure, i.e. if  $l_n = 3^{-n}$ . The Cantor set seems like a set introduced just to form counterexamples to intuitive properties of measures, but the set emerged instead from explorations into trigonometric series. This makes sense, because if, for instance, we consider the functions which satisfy  $\cos(nx) > 1/3$  or  $\sin(nx) > 1/3$  for all  $n$ , we get a Cantor like set.

## 1.5 Symmetries of the Lebesgue Measure

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

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

then  $\mu^*(x + E) = \mu^*(E)$ , and  $E$  is measurable if and only if  $x + E$  is measurable. In fact, the Lebesgue measure is essentially uniquely characterized by translation invariance, up to a constant, as a countably additive function on the  $\sigma$  algebra of Borel sets. Another important symmetry is obtained from the fact that  $\mathbf{R}^d$  is a vector space, and so we can consider invertible linear operators  $T : \mathbf{R}^d \rightarrow \mathbf{R}^d$ , which have a nonzero determinant. The chain rule in basic multivariate calculus then says that, if  $f$  is Riemann integrable, then

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

In particular, this implies that if  $R$  is any box, then  $|TR| = |\det T||R|$ . This then implies that  $|TU| = |\det T||U|$  for any open set  $U$ , which is the countable union of boxes. But now we find that  $\mu^*(TE) = |\det T||E|$ , because  $\mu^*$  can be defined in terms of an outer approximation by open sets, and so  $|TE| = |\det T||E|$  for any open set. In particular, if we define  $\delta E = \{\delta x : x \in E\}$  for  $\delta > 0$ , which is the set  $E$  scaled down or up by some constant amount, then  $|\delta E| = \delta^d \mu^*(E)$ .

## 1.6 The Borel-Cantelli Lemma

The Borel-Cantelli lemma is a crude way to prove a certain set, obtained from a particular solution, has measure zero. Without the methods of measure theory, the proof is quite difficult to obtain.

**Lemma 1.19.** *If  $E_1, E_2, \dots$  are a sequence of Lebesgue measurable sets, with  $\sum |E_i| < \infty$ , then the set*

$$\limsup_{n \rightarrow \infty} E_n = \{x : x \in E_k \text{ for infinitely many } k\} = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} E_m = \lim_{n \rightarrow \infty} \bigcup_{m=n}^{\infty} E_m$$

*has measure zero.*

*Proof.* Because of the limit theorems we established, we find

$$\left| \limsup_{n \rightarrow \infty} E_n \right| = \lim_{n \rightarrow \infty} \left| \bigcup_{m=n}^{\infty} E_m \right|$$

But now Boole's inequality shows

$$\left| \bigcup_{m=n}^{\infty} E_m \right| \leq \sum_{m=n}^{\infty} |E_m|$$

and since the series is summable, for large  $n$  this value becomes smaller than any positive constant.  $\square$

**Example.** *For any irrational number  $x$ , there exists infinitely many rational numbers of the form  $p/q$ , with  $p$  and  $q$  relatively prime, such that*

$$|x - p/q| \leq 1/q^2$$



There also exists numbers  $x$  such that there are infinitely many  $p/q$  with

$$|x - p/q| \leq 1/q^3$$

Nonetheless, this family of numbers has Lebesgue measure zero. To see this, note that it suffices to show the set of numbers in  $[0, 1]$  satisfying this property have Lebesgue measure zero, because the property is invariant under integer translation, and the entire set is the countable union of the intersection with some interval of length one. Consider an enumeration  $p_1/q_1, p_2/q_2, \dots$  of the rational numbers between 0 and 1, and define the sets  $E_n = [p_n/q_n - 1/q_n^3, p_n/q_n + 1/q_n^3] \cap [0, 1]$ . The set of numbers  $x \in [0, 1]$  satisfying the property is then precisely  $\limsup E_n$ . But now we find

$$\sum_{n=1}^{\infty} |E_n| \leq \sum_{n=1}^{\infty} \frac{2}{q_n^3} = \sum_{q=1}^{\infty} \sum_{p=0}^q \frac{2}{q^3} = \sum_{q=1}^{\infty} \frac{2(q+1)}{q^3} \lesssim \sum_{q=1}^{\infty} \frac{1}{q^2} < \infty$$

and so  $\limsup E_n$  has Lebesgue measure zero.

## 1.7 Banach Tarski and Nonmeasurable Sets

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

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

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

**Lemma 1.20.** *There is a subgroup of  $O(3)$  isomorphic to  $F(a, b)$ .*

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

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

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

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

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

**Theorem 1.21** (Banach Tarski). *The sphere may be paradoxically decomposed.*

*Proof.* Partition  $S^2$  into orbits for the faithful group action of  $F(a, b)$  we constructed, and let  $M$  be a set obtained by selecting a single point from every orbit. Also define  $B = a^{-1}M \cup a^{-2}M \cup \dots$ . Then consider the partition of the sphere into four disjoint sets

$$E_1 = S(a^{-1})M - B \quad E_2 = S(b^{-1})M \quad E_3 = F(a)M \quad E_4 = S(b)M$$

We find

$$aE_1 = F(a, b)M - S(a)M - aB = S^2 - F(a)M = S^2 - E_3$$

$$bE_2 = F(a, b)M - S(b)M = S^2 - E_4$$

Thus we have our paradoxical decomposition.  $\square$

All the sets considered, as a subset of  $\mathbf{R}^3$ , are actually measurable, despite this decomposition, but this is only because of the fact that it is contained within  $S^2$ , which has measure zero in  $\mathbf{R}^3$ . But if we now think of the sets  $E_1, E_2, E_3$ , and  $E_4$ , and  $M$  as being identified with the subsets of unit ball obtained by taking rays through the origin, then the set  $M$  cannot be measurable. If it were, then the sets  $E_1, E_2, E_3$ , and  $E_4$  would also be measurable as countable unions of rotations of  $M$ , and so the decompositions above imply

$$|E_1| = |aE_1| = |S^2 - E_3| = |E_1| + |E_2| + |E_4|$$

$$|E_2| = |bE_2| = |S^2 - E_4| = |E_1| + |E_2| + |E_3|$$

The only possible finite positive solutions to these equations  $|E_1| = |E_2| = |E_3| = |E_4| = 0$ , which is impossible, since the union of these sets has measure one. Thus  $M$  must not be measurable.

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

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

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

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

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

## Chapter 2

# Lebesgue Measurable Functions

The initial way we measured the area of nontrivial shapes  $E \subset \mathbf{R}^n$  was through the Riemann integral

$$|E| = \int \chi_E(x) dx$$

Which reflects the fact that, for  $f \geq 0$ , we interpret the Riemann integral of  $f$  as defining the volume under the hypersurface defined by  $f$ . A useful fact about the Riemann integral is that if  $f$  is Riemann integral, then we can write  $f = f^+ - f^-$ , where  $f^+ = \max(f, 0)$  and  $f^- = -\min(f, 0)$  are both Riemann integrable, and then  $\int f = \int f^+ - \int f^-$ , so the general Riemann integral is just the ‘oriented’ area lying above and below the hypersurface defined by  $f$ . Of course, Jordan measure is very limited. If  $f \geq 0$  is a function chosen such that the set

$$X_f = \{(x, y) : 0 \leq y \leq f(x)\}$$

is Lebesgue measurable, then we say  $f$  itself is **Lebesgue measurable**, and then we define the Lebesgue integral as the quantity

$$\int f(x) dx = |X_f|$$

so the integral measures the Lebesgue measure of the set lying below the hypersurface. More generally, if  $f^+$  and  $f^-$  are Lebesgue integrable, and the area under both functions is not both infinite, then we define the general Lebesgue integral

$$\int f(x) dx = \int f^+(x) dx - \int f^-(x) dx$$

and if  $f$  is complex valued, with  $f = u + iv$ , then we say  $f$  is Lebesgue measurable if  $u$  and  $v$  are both Lebesgue measurable (with finite integral), and then we set

$$\int f(x) dx = \int u(x) dx + i \int v(x) dx$$

If  $f$  is only defined on a subset of  $\mathbf{R}^n$ , then we extend it globally by making it vanish outside the domain of definition, and use this to make a more general definition of the Lebesgue integral. Thus we set

$$\int_E f(x) dx = \int f(x) \chi_E(x) dx$$

The utility of this extension relates to the fact that the Lebesgue integral behaves much more nicely with respect to convergence of functions than the Riemann integral.

**Theorem 2.1.** *A function  $f$  is Lebesgue measurable if and only if, for each  $M \in \mathbf{R}$ , the set  $\{x : f(x) \leq M\}$  is Lebesgue measurable.*

**Lemma 2.2.** *If  $f$  is a continuous function in  $n$  variables, and  $g_1, \dots, g_n$  are measurable, then  $f(g_1, \dots, g_n)$  is measurable.*

*Proof.* If  $U$  is open, then  $f^{-1}(U)$  is open, and is therefore the countable union of open rectangles of the form  $(a_1, b_1) \times \dots \times (a_n, b_n)$ , and it suffices to prove that  $(g_1, \dots, g_n)^{-1}((a_1, b_1) \times \dots \times (a_n, b_n))$  is open. But this is precisely the intersection of  $g_m^{-1}((a_m, b_m))$ , which is measurable, completing the proof, since a function is measurable if and only if the inverse image of every open set is measurable.  $\square$

A consequence of this lemma is that if  $f_1, \dots, f_n$  are measurable functions, then so too is  $f_1 + \dots + f_n$  and  $f_1 \dots f_n$ , since the functions  $g(x_1, \dots, x_n) = \sum x_i$  and  $g(x_1, \dots, x_n) = \prod x_i$  are continuous. Other variants show that every nice combination of measurable functions is again measurable, which we take for granted in the sequel.

An incredibly nice feature of Lebesgue measurable functions is their closure under pointwise convergence, a feature not present in the theory of Riemann integrable functions. If  $f_1, f_2, \dots$  is a family of measurable functions, then  $\sup f_n$ ,  $\inf f_n$ ,  $\limsup f_n$ , and  $\liminf f_n$  are all measurable functions as well. This follows by noticing that  $X_{\inf f_n} = \bigcap X_{f_n}$ ,

and  $X_{\sup f_n} = \bigcup X_{f_n}$ . It then follows that  $\limsup f_n = \inf_N \sup_{n \geq N} f_n$  and  $\liminf f_n = \sup_N \inf_{n \geq N} f_n$  are measurable by iterating this argument. Since the set of points  $x$  where  $\lim f_n(x)$  exists is the same as the set of points where  $\limsup f_n(x) = \liminf f_n(x)$ , and the function  $\limsup f_n - \liminf f_n$  is measurable, we conclude the set of points where the limit exists is measurable. Since the function agrees with  $\limsup f_n$  and  $\liminf f_n$  on its domain of definition, the limit of any sequence of measurable functions is also measurable. The next theorem is one of the three main measure theoretic results related to pointwise convergence.

**Theorem 2.3** (Monotone convergence theorem). *If  $0 \leq f_1 \leq f_2 \leq \dots$  converge to a function  $f$  pointwise, then  $\int f_n \rightarrow \int f$  pointwise.*

*Proof.* We have  $X_f = \lim X_{f_n}$ , and so the countable additivity of the Lebesgue measure comes into play to determine that  $|X_f| = \lim |X_{f_n}|$ .  $\square$

The simplest measurable functions are the *step functions*, which are functions which take the form of a finite sum  $\sum a_n \chi_{E_n}$ , where each  $E_n$  is a rectangle, and the  $a_n$  are scalars. Next most simplest are the *simple functions*, which have the same definition as step functions but which allow  $E_n$  to take the form of any measurable set. For both such functions we have

$$\int \sum a_n \chi_{E_n} = \sum a_n |E_n|$$

assuming that the  $E_n$  all have finite measure. If  $f$  is any positive measurable function, then there exists a family of simple functions  $f_1 \leq f_2 \leq \dots$  converging to  $f$  pointwise (approximating the range of  $f_n$  by discrete segments at length  $1/N$  spaces, for instance). If we only want pointwise convergence almost everywhere rather than monotone convergence, we can approximate the function by step functions. If  $f = \lim f_n$ , where the  $f_n$  are monotone converging to  $f$  pointwise. For any measurable set  $E$ , we may find a finite union of rectangles  $F$  with  $|E \Delta F| \leq \varepsilon$ . Thus for each  $n$  and  $\varepsilon$  we can find a step function  $g_n$  which agrees with  $f_n$  except on a set  $F_n$  of size  $1/2^n$ . The functions  $g_n$  converge to  $f$  except on the set

$$\bigcap_{N} \bigcup_{n \geq N} F_n$$

which has measure zero, completing the proof.

If  $f$  is an integrable function, then it is an obvious fact that  $f(x) \neq \infty$  for almost every  $x$ . Surprisingly, this fact is usable in a great many scenarios to show a function is almost everywhere finite.

**Example.** Let  $C$  be a closed set in  $\mathbf{R}$  whose complement has finite measure, we consider the energy at a point to be

$$I(x) = \int \frac{d(y, C)}{|x - y|^2} dy$$

If  $x \notin C$ , then  $d(x, C) > 0$ , and since this changes continuously in  $y$ ,  $I(x) = \infty$ . The function  $d(y, C)$  is Lipschitz, so we should be able to eliminate a factor of  $|x - y|$ , yet the function should still be expected to integrate to infinity because  $1/|x - y|$  is not integrable. Nonetheless, this energy is finite for almost every point of  $C$ . We can see this because

$$\int_C I(x) = \int_C \int_{C^c} \frac{d(y, C)}{|x - y|^2} dy dx = \int_{C^c} d(y, C) \int_C \frac{dx}{|x - y|^2} dy \lesssim \int_{C^c} 1 < \infty$$

so  $I(x)$  must be finite almost everywhere on  $C$ . Later, we will see why this should be expected using the theory of Lebesgue sets – for almost every  $x$ , for each  $\varepsilon > 0$ , there is  $\delta > 0$  such that if  $r \leq \delta$ ,  $|C \cap [x - r, x + r]| \geq 2(1 - \varepsilon)r$ . In particular, any point  $y \in C^c$  that is a distance  $r$  from  $x$  must actually be at a distance  $2\varepsilon r$  from a point in  $C$ . This examples shows that if we are only caring about the Hausdorff distance there is a tighter dependence of  $\delta$  on  $\varepsilon$  which causes the energy to be integrable.

The Lebesgue integral is obtained by breaking up the *range* of a function into separate parts, and measuring the size of the function in the domain. Conversely, the Riemann integral is obtained by breaking up the *domain*. If Riemann were to count the change in his wallet, he would pick out his coins one by one, totalling up the amount one by one. Lebesgue would splay out all his coins, sort them in terms of denomination, and then count the number of coins of each denomination, thus determining the total.

## 2.1 Monotone Rearrangement Invariant Norms

Often in analysis, a problem occurs where one must understand how a quantitative control on the properties of a function, e.g. its oscillation or

growth at  $\infty$ , implies a control on the properties of a related object, i.e. the decay of it's derivative. The most analytically sound way to understand these controls is with respect to a *norm*, a quantitative measure of some property of a function which behaves well under the operations of addition and multiplication. Here we discuss the *translation invariant* norms, those norms such that if  $\Phi : Y \rightarrow X$  is a measure preserving bijection of the measure spaces  $X$  and  $Y$ , then  $f$  and  $f \circ \Phi$  should have the same norm. Furthermore, we study *monotone* norms, so if  $f \leq g$ , the norm of  $f$  should be smaller than the norm of  $g$ .

Really, the only properties a monotone, translation-invariant norm can reflect are the functions typical amplitude, or *height*, and it's *width*, or the measure of the set where the majority of the amplitude is distributed. These two properties are not precisely defined, and the translation invariant norms quantify them in various different ways. In a particular problem, one picks the norm best emphasizing the features useful in a problem, often chosen by optimizing over a large family of norms.

## 2.2 The $L^p$ norms

The most important family of measurable function are those which are integrable, i.e. such that

$$\int |f(x)| dx < \infty.$$

This condition already places a control over the height and width of  $f$ ; the function cannot be too large, except on a small set, and also must have the large majority of it's mass on a small set. The condition

$$\int |f(x)|^2 dx < \infty$$

also places a certain amount of control over the height and width of a function  $f$ , but now the height is controlled more tightly than the width, since the portions of the domain where  $f$  is large are magnified in  $|f(x)|^2$ , and the portions of the domain where  $f$  is small are made more negligible. More generally, for any positive value of  $p$ , the condition

$$\int |f(x)|^p dx < \infty \tag{2.1}$$



controls the height and width of a function, and we define the space  $L^p(X)$  as the family of measurable functions  $f : X \rightarrow \mathbf{C}$  satisfying (2.1). The norm

$$\|f\|_p = \left( \int |f|^p \right)^{1/p}$$

provides a quantity measuring  $f$ 's width and height. The importance of taking the  $p$ 'th root is useful for the stability of the norm, so that  $\|\lambda f\|_p = |\lambda| \|f\|_p$  for any constant  $\lambda$ .

**Example.** If  $f(x) = |x|^{-\alpha}$  for  $x \in \mathbf{R}^d$ , then integration by radial coordinates shows there exists constants  $C_d$  such that

$$\int_{\varepsilon \leq |x| \leq M} |f(x)|^p dx = C_d \int_{\varepsilon}^M r^{d-1-p\alpha} dr = C_d \cdot \frac{M^{d-p\alpha} - \varepsilon^{d-p\alpha}}{d-p\alpha}.$$

This quantity remains finite as  $\varepsilon \rightarrow 0$  if and only if  $d > p\alpha$ , and finite as we let  $M \rightarrow \infty$  if and only if  $d < p\alpha$ . Thus if  $p < d/\alpha$ ,  $f$  is locally in  $L^p$ , or in  $L^p(B)$  for every bounded  $B \subset \mathbf{R}^d$ . Conversely, if  $p > d/\alpha$  for every domain  $B$  with  $d(B, 0) > 0$ . Notice that for large  $\alpha$ ,  $f$  is locally in  $L^p$  for fewer and fewer values of  $p$ , since  $f$  becomes more peaked at the origin, whereas  $f$  is in  $L^p$  away from the origin for more and more  $L^p$  norms as  $\alpha \rightarrow \infty$ , since  $f$  then decays faster and faster away from the origin.

**Example.** Let  $p < q$ , and suppose  $f \in L^p(X) \cap L^q(X)$ . For any  $r \in (p, q)$ , the  $L^r$  norm emphasizes the height of  $f$  less than the  $L^q$  norm, and emphasizes the width of  $f$  less than the  $L^p$  norm. In particular, we find that for any  $\lambda \geq 0$ ,

$$\begin{aligned} \|f\|_r^r &= \int_{\mathbf{R}} |f(x)|^r dx = \int_{|f(x)| \leq 1} |f(x)|^r dx + \int_{|f(x)| > 1} |f(x)|^r dx \\ &\leq \int_{|f(x)| \leq 1} |f(x)|^p dx + \int_{|f(x)| > 1} |f(x)|^q dx \\ &\leq \|f\|_p^p + \|f\|_q^q < \infty. \end{aligned}$$

In particular, this shows  $f \in L^r(X)$ .

The bound obtained in the last example can be improved by using scaling symmetries. For any  $A > 0$ ,

$$\|f\|_r^r = \frac{\|Af\|_r^r}{A^r} \leq \frac{\|Af\|_p^p + \|Af\|_q^q}{A^r} \leq \frac{A^p \|f\|_p^p + A^q \|f\|_q^q}{A^r}.$$

If  $1/r = \theta/p + (1-\theta)/q$ , and we set  $A = \|f\|_q^{q/(p-q)} / \|f\|_p^{p/(p-q)}$ , then the above inequality implies  $\|f\|_r \leq 2\|f\|_p^\theta \|f\|_q^{1-\theta}$ , which is a homogenous equality. The constant 2 can be removed in the equation using the *tensor power trick*. If we consider the function on  $(\mathbf{R}^d)^n$  defined by  $f^{\otimes n}(x_1, \dots, x_n) = f(x_1) \dots f(x_n)$ , then  $\|f^{\otimes n}\|_t = \|f\|_t^n$ , and so

$$\|f\|_r = \|f^{\otimes n}\|_r^{1/n} \leq (2\|f^{\otimes n}\|_p \|f^{\otimes n}\|_q)^{1/n} = 2^{1/n} \|f\|_p \|f\|_q.$$

We can then take  $n \rightarrow \infty$  to conclude that  $\|f\|_r \leq \|f\|_p \|f\|_q$ . This is an instance of *real interpolation*; In order to conclude some fact about a function which lies ‘between’ two other functions we know how to deal with, we split the function up into two parts lying in the other spaces, deal with them separately, and then put them back together to get some equality. One can then apply various symmetry considerations (homogeneity and the tensor power trick being two examples) to eliminate extraneous constants unneeded to the problem. We will later prove this theorem as a special case of some inequalities involving convexity, but it is useful to know the techniques used here for other, more complex scenarios.

**Example.** If  $f = A\chi_E$ , and we set  $H = |A|$  and  $W = |E|$ , then  $\|f\|_p = W^{1/p}H$ . As  $p \rightarrow \infty$ , the value of  $\|f\|_p$  depends more and more on  $H$ , and less on  $W$ , and in fact  $\lim_{p \rightarrow \infty} \|f\|_p = H$ . If  $f = \sum A_n \chi_{E_n}$ , and  $|A_m|$  is the largest constant from all other values  $A_n$ , then as  $p$  becomes large,  $|A_m|^p$  overwhelms all other terms. We calculate that as  $p \rightarrow \infty$ ,

$$\|f\|_p = \left( \sum |E_n| |A_n|^p \right)^{1/p} = |A_m|^p (|E_m| + o(1))^{1/p} = |A_m| (1 + o(1)).$$

This implies  $\|f\|_p \rightarrow |A_m|$  as  $p \rightarrow \infty$ . But as  $p \rightarrow 0$ , the power of  $1/p$  works against us, so that  $\|f\|_p$  rarely converges to any universal quantity. We can conclude that  $\lim_{p \rightarrow 0} \|f\|_p^p = \sum |E_n|$ , so the width of the function becomes more and more emphasized. In the case where  $\sum |E_i| = 1$ , we can use L’Hopitals rule as  $p \rightarrow 0$  to conclude that

$$\lim_{p \rightarrow 0} \log \|f\|_p = \lim_{p \rightarrow 0} \frac{\log(\sum |a_i|^p |E_i|)}{p} = \sum |E_i| \log |a_i| = \int_{|f(x)| > 0} \log |f(x)| dx.$$

Taking exponentials on both sides, we conclude that

$$\lim_{p \rightarrow 0} \|f\|_p = \exp \left( \int_{|f(x)| > 0} \log |f(x)| dx \right).$$

However, the fact that  $\log |f(x)|$  can take negative values makes this norm difficult to work with, unless  $f(x)$  is suitably granular.

As  $p \rightarrow \infty$ , the width of a function is disregarded completely by the  $L^p$  norm, motivating the definition of the  $L^\infty$  norm; Given a measurable  $f$ , we define  $\|f\|_\infty$  to be the smallest number such that  $|f| \leq \|f\|_\infty$  almost surely. We then define  $L^\infty(X)$  to be the space of measurable functions  $f$  for which  $\|f\|_\infty < \infty$ . We have already shown  $\|f\|_p \rightarrow \|f\|_\infty$  if  $s$  is a simple function, and the density of such functions, combined with some functional analysis, shows that if  $f \in L^p(X) \cap L^\infty(X)$  for any value of  $p$ ,  $\|f\|_q \rightarrow \|f\|_\infty$  as  $q \rightarrow \infty$ . Sometimes, authors also define  $\|f\|_0 = |\text{supp } f| = |\{x : f(x) \neq 0\}|$ , but this is really a ‘zeroeth power’ of the  $L^0$  norm. Because of this, most of the nice norm properties of the  $L^p$  spaces are broken. For instance,  $\|\cdot\|_0$  is no longer homogenous, i.e. if  $\lambda \neq 0$ ,  $\|\lambda f\|_0 = \|f\|_0$ .

An incredibly useful properties of the  $L^p$  norms is provided by the Hölder inequality, which describes a way to estimate the  $L^p$  norm of the product of two functions in terms of the  $L^p$  norms of the original functions. For the purposes of the next argument, we let  $1/\infty = 0$ .

**Theorem 2.4 (Hölder).** *If  $0 < p, q \leq \infty$  and  $1/p + 1/q = 1/r$ ,  $\|fg\|_r \leq \|f\|_p \|g\|_q$ .*

*Proof.* The case where  $p$  or  $q$  is  $\infty$  is left as an exercise to the reader. In the other case, by moving around exponents, we may simplify to the case where  $r = 1$ . The theorem depends on the log convexity inequality, such that for  $A, B \geq 0$  and  $0 \leq \theta \leq 1$ ,  $A^\theta B^{1-\theta} \leq \theta A + (1 - \theta)B$ . But since the logarithm is concave, we calculate

$$\log(A^\theta B^{1-\theta}) = \theta \log A + (1 - \theta) \log B \leq \log(\theta A + (1 - \theta)B),$$

and we can then exponentiate. To prove Hölder’s inequality, by scaling  $f$  and  $g$ , which is fine by homogeneity, we may assume that  $\|f\|_p = \|g\|_q = 1$ . Then we calculate

$$\begin{aligned} \|fg\|_1 &= \int |f(x)| |g(x)| = \int |f(x)|^{p/p} |g(x)|^{q/q} \\ &\leq \int \frac{|f(x)|^p}{p} + \frac{|g(x)|^q}{q} = \frac{1}{p} + \frac{1}{q} = 1 \end{aligned}$$

so we have proved the inequality. If  $p = \infty$ ,  $q = 1$ , then the inequality is trivial, since we have the pointwise inequality  $|f(x)g(x)| \leq \|f\|_\infty |g(x)|$ , which we can then integrate.  $\square$

*Remark.* Note that  $A^\theta B^{1-\theta} \leq \theta A + (1-\theta)B$  is an *equality* if and only if  $A = B$ , or  $\theta \in \{0,1\}$ . In particular, following through the proof above shows that if  $\|f\|_p = \|g\|_q = 1$ , we must have  $|f(x)|^{1/p} = |g(x)|^{1/q}$  almost everywhere. In general, this means Hölder's inequality is sharp if and only if  $|f(x)|^{1/p}$  is a constant multiple of  $|g(x)|^{1/q}$ .

The next inequality is known as the *triangle inequality*.

**Corollary 2.5.** *Given  $f, g$ , and  $p \geq 1$ ,  $\|f + g\|_p \leq \|f\|_p + \|g\|_p$ .*

*Proof.* The inequality when  $p = 1$  is obtained by integrating the inequality  $|f(x) + g(x)| \leq |f(x)| + |g(x)|$ , and the case  $p = \infty$  is equally trivial. When  $1 < p < \infty$ , by scaling we can assume that  $\|f\|_p + \|g\|_p = 1$ . Then we can apply Hölder's inequality combined with the  $p = 1$  case to conclude

$$\begin{aligned} \int |f(x) + g(x)|^p &\leq \int |f(x)| |f(x) + g(x)|^{p-1} + \int |g(x)| |f(x) + g(x)|^{p-1} \\ &\leq \|f\|_p \|f + g\|_q^{p-1} + \|g\|_p \|f + g\|_q^{p-1} = \|f + g\|_p^{p-1} \end{aligned}$$

Thus  $\|f + g\|_p^p \leq \|f + g\|_p^{p-1}$ , and simplifying gives  $\|f + g\|_p \leq 1$ .  $\square$

*Remark.* Suppose  $\|f + g\|_p = \|f\|_p + \|g\|_p$ . Following through the proof given above shows that both applications of Hölder's inequality must be sharp. And this is true if and only if  $|f(x)|^p$  and  $|g(x)|^p$  are scalar multiples of  $|f(x) + g(x)|^p$  almost everywhere. But this means  $|f(x)|$  and  $|g(x)|$  are scalar multiples of  $|f(x) + g(x)|$ . If  $|f(x)| = A|f(x) + g(x)|$  and  $|g(x)| = B|f(x) + g(x)|$ . If  $g \neq 0$ , this implies there is  $C$  such that  $|f(x)| = C|g(x)|$  for some  $C > 0$ . Thus we can write  $f(x) = Ce^{i\theta(x)}g(x)$ , and we must have

$$\|f + g\|_p^p = \int |1 + Ce^{i\theta(x)}|^p |g(x)|^p = (1 + C)^p \int |g(x)|^p$$

so  $|1 + Ce^{i\theta(x)}| = |1 + C|$  almost everywhere but this can only be true if  $e^{i\theta(x)} = 1$  almost everywhere, so  $f = Cg$ . Thus the triangle inequality is only sharp if  $f$  and  $g$  are positive scalar multiples of one another.

This discussion leads to a useful heuristic: Unless  $f$  and  $g$  are 'aligned' in a certain way, the triangle inequality is rarely sharp. For instance, if  $f$  and  $g$  have disjoint support, we calculate that

$$\|f + g\|_p = \left( \|f\|_p^p + \|g\|_p^p \right)^{1/p}$$

For  $p > 1$ , this is always sharper than the triangle inequality.

*Remark.* If  $p < 1$ , then the proof above no longer works, and in fact, is no longer true. In fact, if  $f$  and  $g$  are non-negative functions, then we actually have the *anti* triangle inequality

$$\|f + g\|_p \geq \|f\|_p + \|g\|_p.$$

On the other hand, we do have a *quasi* triangle inequality. We note that for any  $A, B > 0$ , and  $0 < p \leq 1$ ,

$$(A + B)^p \leq A^p + B^p.$$

$$(A + B)^{1/p} \leq 2^{1/p-1} (A^{1/p} + B^{1/p}).$$

The first inequality follows easily by normalizing  $A + B = 1$ , in which case  $A^p > A$ ,  $B^p > B$ . The second follows from Hölder's inequality for sums (or Hölder's inequality for the counting measure). Put together, we conclude

$$\begin{aligned} \|f + g\|_p &= \left( \int |f(x) + g(x)|^p dx \right)^{1/p} \\ &\leq \left( \int |f(x)|^p dx + \int |g(x)|^p dx \right)^{1/p} \\ &\leq 2^{1/p-1} (\|f\|_p + \|g\|_p). \end{aligned}$$

More generally, we have

$$(A_1 + \cdots + A_N)^{1/p} \leq N^{1/p-1} (A_1^{1/p} + A_2^{1/p} + \cdots + A_N^{1/p}),$$

which leads to the inequality

$$\|f_1 + \cdots + f_n\|_p \leq N^{1/p-1} (\|f_1\|_p + \cdots + \|f_n\|_p).$$

Thus  $L^p(X)$  is a quasinorm space.

As we have mentioned, as  $p \rightarrow \infty$ , the  $L^p$  norm excludes functions with large peaks, or large height, and as  $p \rightarrow 0$ , the  $L^p$  norm excludes functions with large tails, or large width. They form a continuously changing family of functions as  $p$  ranges over the positive numbers. In general, there is no inclusion of  $L^p(X)$  in  $L^q(X)$  for any  $p, q$ , except in two circumstances which occur often enough to be mentioned.

**Example.** If  $X$  is a finite measure space, and  $0 < p \leq q \leq \infty$ ,  $L^p(X) \subset L^q(X)$ . Hölder's inequality implies  $\|f\|_p = \|f \chi_X\|_p \leq \|f\|_q |X|^{1/p-1/q}$ . Taking  $q \rightarrow \infty$ , we conclude  $\|f\|_p \leq |X|^{1/p} \|f\|_\infty$ . On the other hand, we can rearrange the inequality to read more symmetrically as  $\|f\|_p |X|^{-1/p} \leq \|f\|_q |X|^{-1/q}$ . In particular, the  $L^p$  norms increase as  $p$  increases if  $X$  is a probability space.

**Example.** On the other hand, suppose the measure space is granular, in the sense that there is a universal  $\varepsilon$  such that either  $|E| = 0$  or  $|E| \geq \varepsilon$  for any measurable set  $E$ . Then  $L^p(X) \supset L^q(X)$  for  $0 < p \leq q \leq \infty$ . First we check the  $q = \infty$  case, which follows by the trivial estimate

$$\int |f(x)|^p \geq \varepsilon \|f\|_\infty^p,$$

so  $\|f\|_\infty \leq \|f\|_p \varepsilon^{-1/p}$ . But then applying log convexity, if  $p \leq q < \infty$ , we can write  $1/q = \theta/p$  for  $0 < \theta \leq 1$ , and then log convexity shows

$$\|f\|_q = \|f\|_p^\theta \|f\|_\infty^{1-\theta} \leq \varepsilon^{-(1-\theta)/p} \|f\|_p = \varepsilon^{-1/p-1/q} \|f\|_p.$$

If  $\varepsilon = 1$ , which occurs if  $X = \mathbf{Z}$ , then the  $L^p$  norms are decreasing in  $p$ .

**Example.** Controlling additional properties of the function offers similar properties as for control on the measure space. If  $|f(x)| \leq M$  for almost all  $x$ , then for  $p \leq q$ ,

$$\|f\|_q \leq \|f\|_p^{p/q} M^{1-p/q}.$$

Conversely, if  $|f(x)| \geq M$  whenever  $f(x) \neq 0$ , then

$$\|f\|_p \leq \|f\|_q^{q/p} M^{1-q/p}.$$

*Remark.* We can often use such results in spaces which are not granular by coarsening the sigma algebra. For instance, the Lebesgue measure is  $\varepsilon^d$  granular over the sigma algebra generated by the length  $\varepsilon$  cubes whose corner's lie on the lattice  $(\mathbf{Z}/\varepsilon)^d$ , and if a function is measurable with respect to such a  $\sigma$  algebra we call the function  $\varepsilon$  granular.

*Remark.* If we let  $X = \{1, \dots, N\}$ , then  $X$  is both finite and granular, so all  $L^p$  norms are comparable. In particular, if  $p \leq q$ ,

$$\|f\|_q \leq \|f\|_p \leq N^{1/p-1/q} \|f\|_q.$$

The left hand side of this inequality becomes sharp when  $f$  is concentrated at a single point, i.e.  $f(n) = \mathbf{I}(n = 1)$ . On the other hand, the left hand side becomes sharp when  $f$  is constant, i.e.  $f(n) = 1$  for all  $n$ .

**Theorem 2.6.** *If  $p_\theta$  lies between  $p_0$  and  $p_1$ , then*

$$L^{p_0}(X) \cap L^{p_1}(X) \subset L^{p_\theta}(X) \subset L^{p_0}(X) + L^{p_1}(X)$$

*Proof.* If  $\|f\|_{p_0}, \|f\|_{p_1} < \infty$ , then for any  $p_\theta$  between  $p_0$  and  $p_1$ ,

$$\|f \chi_{|f| \leq 1}\|_{p_\theta}^{p_\theta} = \int_{|f| \leq 1} |f|^{p_\theta} \leq \int_{|f| \leq 1} |f|^{p_0} < \infty$$

$$\|f \chi_{|f| > 1}\|_{p_\theta}^{p_\theta} = \int_{|f| > 1} |f|^{p_\theta} \leq \int_{|f| > 1} |f|^{p_1} < \infty$$

Applying the triangle inequality, we conclude that  $\|f\|_{p_\theta} < \infty$ . In the case where  $p_1 = \infty$ , then  $f \chi_{|f| > 1}$  is bounded, and must have finite support if  $p_0 < \infty$ , which shows this integral is bounded. Note the inequalities above show that we can split any function with finite  $L^{p_\theta}$  norm into the sum of a function with finite  $L^{p_0}$  norm and another with finite  $L^{p_1}$  norm.  $\square$

This theorem is important in the study of interpolation theory, because if we have two linear operators  $T_{p_0}$  defined on  $L^{p_0}(X)$  and  $T_{p_1}$  on  $L^{p_1}(X)$ , and they agree on  $L^{p_0}(X) \cap L^{p_1}(X)$ , then there is a unique linear operator  $T_{p_\theta}$  on  $L^{p_\theta}(X)$  which agrees with these two functions, and we can consider the boundedness of such a function with respect to the  $L^{p_\theta}$  norms.

**Example.** *It is easy to prove that if  $\hat{f}$  is the Fourier transform of a Schwartz function  $f$ , then  $\|\hat{f}\|_2 = \|f\|_2$ , and  $\|\hat{f}\|_\infty = \|f\|_1$ . In particular, this enables us to uniquely extend the Fourier transform to unique operators from  $L^1(\mathbf{R}^n)$  to  $L^\infty(\mathbf{R}^n)$ , and from  $L^2(\mathbf{R}^n) \rightarrow L^2(\mathbf{R}^n)$ . They agree on the intersection upon which they are defined, so in particular the Fourier transform of any function in  $L^p(\mathbf{R}^n)$  is well defined, for  $1 \leq p \leq 2$ , and is an element of  $L^\infty(\mathbf{R}^n) + L^2(\mathbf{R}^n)$ . Later techniques in this chapter will enable us to justify that the two inequalities we have just given immediately imply that  $\|\hat{f}\|_q \leq \|f\|_p$  for  $1 \leq p \leq 2$*

The triangle inequality does not hold when  $0 < p < 1$ . In fact, if  $f, g \geq 0$ ,  $\|f + g\|_p \geq \|f\| + \|g\|$ . Thus the  $L^p$  norm here is a *quasinorm* rather than a norm. This is one of the deficiencies which leads the  $L^p$  theories for  $0 < p < 1$  to be rather deficient. This can be fixed in many ways, most notably to harmonic analysts, using the theory of Hardy spaces.

**Theorem 2.7.** *If  $0 < p < 1$ ,*

$$\|f_1 + \cdots + f_N\|_p \leq N^{1/p-1} [\|f_1\|_p + \cdots + \|f_N\|_p]$$

*If  $f_1, \dots, f_N \geq 0$ , we have the anti triangle inequality*

$$\|f_1 + \cdots + f_N\|_p \geq \|f_1\|_p + \cdots + \|f_N\|_p$$

*Proof.* If  $0 < p < 1$ , by the norm properties of  $l^q(\mathbf{C}^n)$  we have pointwise subadditivity

$$|f_1(x) + \cdots + f_N(x)|^p \leq |f_1(x)|^p + \cdots + |f_N(x)|^p$$

Integrating this gives  $\|f_1 + \cdots + f_N\|_p^p \leq \|f_1\|_p^p + \cdots + \|f_N\|_p^p$ . But now we can apply Hölder's inequality with  $1/p$  and  $1/(1-p)$  to conclude that

$$1 \cdot \|f_1\|_p^p + \cdots + 1 \cdot \|f_N\|_p^p \leq N^{1-p} (\|f_1\|_p + \cdots + \|f_N\|_p)^p$$

and then we can take  $p$ 'th roots. But the function  $x \mapsto x^p$  is concave, so for any  $a_1, \dots, a_N \geq 0$  with  $a_1 + \cdots + a_N = 1$ ,

$$(f_1(x) + \cdots + f_N(x))^p \geq a_1^{1-p} f_1(x)^p + \cdots + a_N^{1-p} f_N(x)^p$$

Integrating both sides, we conclude

$$\|f_1 + \cdots + f_N\|_p^p \geq a_1^{1-p} \|f_1\|_p^p + \cdots + a_N^{1-p} \|f_N\|_p^p$$

But setting  $a_i = \|f_i\|_p / (\sum \|f_i\|_p)$ , we conclude

$$\|f_1 + \cdots + f_N\|_p^p \geq \left( \sum \|f_i\|_p \right)^p$$

Taking  $p$ 'th roots completes the proof. □

## 2.3 Weak and Lorentz Norms

The weak  $L^p$  norms are obtained as a slight 'refinement' of the  $L^p$  norms. The first idea to understand these norms is that any translation invariant property of a function  $f$  should express itself as a property of its distribution function  $F_f : [0, \infty) \rightarrow [0, \infty)$ , denoted by  $F$  if the function is clear, and defined by  $F(\lambda) = |\{ |f| > \lambda \}|$ . Then  $F$  is right-continuous and decreasing.



**Theorem 2.8.** *If  $\phi$  is an increasing, differentiable function on the real line with  $\phi(0) = 0$ , then*

$$\int_X \phi(|f(x)|) = \int_0^\infty \phi'(t) F(t) dt$$

*Proof.* An application of Fubini's theorem is all that is needed to show

$$\begin{aligned} \int_X \phi(|f(x)|) dx &= \int_X \int_0^{|f(x)|} \phi'(t) dt dx \\ &= \int_0^\infty \phi'(t) \int_{|f(x)| > t} dx du \\ &= \int_0^\infty \phi'(t) F(t) dt \end{aligned}$$

□

As a special case we find

$$\|f\|_p = \left( p \int_0^{|f(x)|} \lambda^p \frac{d\lambda}{\lambda} dx \right)^{1/p} = \left( p \int_0^\infty F(\lambda) \lambda^p \frac{d\lambda}{\lambda} \right)^{1/p}$$

The multiplicative invariance of the integral implies that each dyadic interval has equal mass, and so one verifies that  $\|f\|_p \sim p (\sum F(2^n) 2^{np})^{1/p}$ . Thus  $F(\lambda)$  in some sense must decay slightly faster than  $1/\lambda^p$  for  $f$  to have finite  $L^p$  norm. Indeed,

$$F(\lambda) = |\{|f|^p > \lambda^p\}| \leq \frac{1}{\lambda^p} \int |f|^p = \frac{\|f\|_p^p}{\lambda^p}$$

This is known as Chebyshev's inequality. The **Weak  $L^p$  space**  $L^{p,\infty}(X)$  consists of all functions  $f$  for which  $F(\lambda) \lesssim 1/\lambda^p$ . The smallest value  $A$  for which  $F(\lambda) \leq A/\lambda^p$  is the  $p$ 'th power of the weak  $L^p$  norm, denoted  $\|f\|_{p,\infty}^p$ . The weak  $L^p$  space contains  $L^p(X)$ , because of Chebyshev's inequality, and the weak  $L^p$  norm is a contraction of the standard  $L^p$  norm. We let  $L^{\infty,\infty} = L^\infty$ , since  $\|f\|_{p,\infty} \rightarrow \|f\|_\infty$  as  $p \rightarrow \infty$ . The weak norm provides another monotone translation invariant measure, and it often comes up when finer tuning is needed in certain interpolation arguments, especially when dealing with maximal functions.

**Example.** If  $f = H\chi_E$ , with  $|E| = W$ , then to determine  $\|f\|_{p,\infty}$ , we must find the smallest value such that

$$W = F(\lambda) \leq \frac{\|f\|_{p,\infty}^p}{\lambda^p}$$

for  $\lambda \leq H$ . This gives  $\|f\|_{p,\infty} = W^{1/p}H = \|f\|_p$ . More generally, for  $f = \sum A_n\chi_{E_n}$ , with  $|A_1| \leq \dots \leq |A_N|$ , we have

$$\|f\|_{p,\infty} = \left( \max_{m \leq N} |A_m|^p \sum_{n=m}^N |E_n| \right)^{1/p}$$

which is certainly a more discrete inequality than the  $L^p$  norm.

**Example.** The  $L^p$  norms are unable to classify the functions  $1/|x|^\alpha$ , but the weak  $L^p$  norms are. In fact,  $1/|x|^\alpha$  is in  $L^{p,\infty}$  if and only if  $p\alpha = d$ , since

$$|\{1/|x|^{p\alpha} > \lambda\}| = \left| \left\{ |x| \leq \frac{1}{\lambda^{1/p\alpha}} \right\} \right| \propto_d \frac{1}{\lambda^{d/p\alpha}}$$

Notice that  $1/|x|^\alpha$  is only barely not contained in  $L^p$  if  $p\alpha = d$ , so the weak norm provides a fine tune change enough to make the norm of the function finite.

In some cases, we need an even more fine tuned norm. We have proved that  $\|f\|_p = p^{1/p} \|\lambda F^{1/p}\|_{L^p(\mathbf{R}^+)}$ , where  $\mathbf{R}^+$  is given the multiplicatively invariant Haar measure. We also have  $\|f\|_{p,\infty} = \|\lambda F^{1/p}\|_{L^\infty(\mathbf{R}^+)}$ . In general, we define the **Lorentz spaces**  $L^{p,q}(X)$  to be the space of functions with finite Lorentz norm

$$\|f\|_{p,q} = p^{1/q} \|\lambda F^{1/p}\|_{L^q(\mathbf{R}^+)} = \left( p \int_0^\infty \lambda^{q-1} F(\lambda)^{q/p} d\lambda \right)^{1/q}$$

Thus  $q = \infty$  gives the weak  $L^p$  norm, and  $q = p$  gives the strong  $L^p$  norm.

**Lemma 2.9.** Give a function  $f$  and  $g$ ,  $\alpha, \beta > 0$ , then

- If  $|g| \leq |f|$ , then  $F_g \leq F_f$ .
- $F_{\alpha f}(\lambda) = F_f(\lambda/|\alpha|)$ .
- $F_{f+g}(\lambda + \gamma) \leq F_f(\lambda) + F_g(\gamma)$ .

- $F_{fg}(\lambda\gamma) \leq F_f(\lambda) + F_g(\gamma).$

Notice that multiplying the function by a constant dilates the distribution function, so that a norm defined with respect to the distribution function must have a way to turn this dilation into multiplication by a constant. This is given by the use of the multiplicatively invariant measure on  $\mathbf{R}^+$ . Thus we see easily that  $\|\alpha f\|_{p,q} = |\alpha| \|f\|_{p,q}$ . We do not have the triangle inequality for the Lorentz and weak norms, but we have a quasi triangle inequality. This is given by the calculation

$$\begin{aligned} \|f + g\|_{p,q}^q &= p \int \lambda^q F_{f+g}^{q/p}(\lambda) \frac{d\lambda}{\lambda} \leq p \int \lambda^q (F_f(\lambda/2) + F_g(\lambda/2))^{q/p} \\ &\leq p 2^{q/p} \int \left( \lambda^q F_f(\lambda/2)^{q/p} + \lambda^q F_g(\lambda/2)^{q/p} \right) \frac{d\lambda}{\lambda} \\ &= 2^{q+q/p} \left( \|f\|_{p,q}^q + \|g\|_{p,q}^q \right) \end{aligned}$$

Taking  $q$ 'th roots gives

$$\|f + g\|_{p,q} \leq 2^{1+1/p+1/q} (\|f\|_{p,q} + \|g\|_{p,q})$$

Though the Lorentz norms do not satisfy the triangle inequality, they still induce a topology which can also be given by a complete norm.

A useful way to work with the  $L^{p,q}$  norms is by discretizing. When  $p = q$ , if we set  $E_k = \{x : 2^k \leq f(x) \leq 2^{k+1}\}$ , then

$$\|f\|_p \sim \left( \sum_{k=-\infty}^{\infty} |E_k| 2^{pk} \right)^{1/p} \sim_p \left( \sum_{k=-\infty}^{\infty} F(2^k) 2^{pk} \right)^{1/p}$$

where the dependence on  $p$  in the second equivalence can be removed if we know  $p \geq \varepsilon$  for some  $\varepsilon$ . Thus one way to study  $f$  is to consider the integrals of a dyadic function  $\sum 2^k \chi_{E_k}$ . We now consider a generalization of this process to the Lorentz norms. We say a function  $f$  is a *quasi-step function* of height  $H$  and width  $W$  if  $f$  is supported on a set  $E$  with  $|E| \sim W$ , and  $|f(x)| \sim H$  on  $E$ . If we merely have  $|f(x)| \lesssim H$  on  $E$ , and  $|E| \lesssim W$ , then we call  $f$  a *sub step function*. The hidden constants are normally taken to be universally quantified.

**Theorem 2.10.** *Let  $f$  be measurable,  $0 < p < \infty$ ,  $1 \leq q \leq \infty$ , and  $0 < A < \infty$ . Then the following are equivalent up to changes in universal constants:*

- $\|f\|_{p,q} \lesssim_{p,q} A$ .
- We can write  $f = \sum_{n \in \mathbf{Z}} f_n$ , where the  $f_n$  are quasi-step functions with height  $2^n$  and with  $W_n$ , have disjoint support, and  $\|2^n W_n^{1/p}\|_{l^q(\mathbf{Z})} \lesssim_{p,q} A$ .
- We have  $|f| \leq \sum_{n \in \mathbf{Z}} 2^n \chi_{E_n}$ , with  $\|2^n |E_n|^{1/p}\|_{l^q(\mathbf{Z})} \lesssim_{p,q} A$ .
- We can write  $f = \sum_{n \in \mathbf{Z}} f_n$ , where the  $f_n$  have disjoint supports,  $f_n$  is a sub-step function of width  $2^n$  and height  $H_n$ , the heights non-increasing as  $n$  increases, with  $H_{n+1} \leq |f_n| \leq H_n$  on the support of  $f_n$ , and with the bound  $\|H_n 2^{n/p}\|_{l^q(\mathbf{Z})} \lesssim_{p,q} A$ .
- We have  $|f| \leq \sum_{n \in \mathbf{Z}} H_n \chi_{E_n}$ , where  $|E_n| \lesssim_{p,q} 2^n$  and  $\|H_n 2^{n/p}\|_{l_n^q(\mathbf{Z})} \lesssim_{p,q} A$ .

The second and fourth points are useful when using  $L^{p,q}$  bounds, whereas the third and fifth are more useful for obtaining  $L^{p,q}$  bounds.

*Proof.* The fact that the second point implies the third, and the fourth the fifth, are obvious. To show that the first point implies the second, we perform a vertical dyadic layer cake decomposition. Fix  $f$ , and define  $f_n = f \chi_{2^n < |f(x)| \leq 2^{n+1}}$ , and  $W_n = |\{2^n < |f| \leq 2^{n+1}\}|$ . Then all the constraints are satisfied in point two for the  $f_n$ , and

$$\|2^n W_n^{1/p}\|_{l^q(\mathbf{Z})} = \left( \sum \left( 2^n W_n^{1/p} \right)^q \right)^{1/q} \leq \left( \sum 2^{nq} F(2^n)^{q/p} \right)^{1/q}$$

But

$$2^{nq} F(2^n)^{q/p} \leq \frac{1}{\log 2} \int_{2^{n-1}}^{2^n} (2^q \lambda) F(\lambda)^{q/p} \frac{d\lambda}{\lambda}$$

This shows that

$$\|2^n W_n^{1/p}\|_{l^q(\mathbf{Z})} \leq \frac{2}{(\log 2)^{1/q}} \|f\|_{p,q}$$

completing the argument. To show that the first point implies the fourth, we perform a horizontal dyadic layer cake decomposition. We let  $H_n$  be the infimum of  $\lambda$  such that  $F(\lambda) \leq 2^n$ , and define  $f_n = f \chi_{H_{n+1} < |f| \leq H_n}$ . Then  $f_n$  is supported on a set of measure  $F(H_{n+1}) - F(H_n) \leq 2^{n-1}$ , and by a

telescoping estimate,

$$\begin{aligned}
H_n 2^{n/p} &= \left( \sum_{k=0}^{\infty} (H_{n+k}^q - H_{n+k+1}^q) 2^{nq/p} \right)^{1/q} \\
&= \left( q^{-1} \sum_{k=0}^{\infty} 2^{-kq/p} \|\lambda 2^{(n+k)/p}\|_{L^q([H_{n+k+1}, H_{n+k}], \mathbf{R}^+)}^q \right)^{1/q} \\
&\leq \left( q^{-1} \sum_{k=0}^{\infty} 2^{-kq/p} \|\lambda F(\lambda)^{1/p}\|_{L^q([H_{n+k+1}, H_{n+k}], \mathbf{R}^+)}^q \right)^{1/q}
\end{aligned}$$

But now we calculate that

$$\begin{aligned}
\|H_n^q 2^{nq/p}\|_{l^q(\mathbf{Z})} &\leq \left( q^{-1} \sum_{n \in \mathbf{Z}} \sum_{k=0}^{\infty} 2^{-kq/p} \|\lambda F(\lambda)^{1/p}\|_{L^q([H_{n+k+1}, H_{n+k}], \mathbf{R}^+)}^q \right)^{1/q} \\
&\leq q^{-1/q} \left( \sum_{k=0}^{\infty} 2^{-kq/p} \|\lambda F(\lambda)^{1/p}\|_{L^q(\mathbf{R}^+)}^q \right)^{1/q} \\
&= \frac{q^{-1/q}}{1 - 2^{-q/p}} \|\lambda F(\lambda)^{1/p}\|_{L^q(\mathbf{R}^+)} \lesssim \frac{\|\lambda F(\lambda)^{1/p}\|_{L^q(\mathbf{R}^+)}}{1 - 2^{-1/p}}
\end{aligned}$$

All that remains is to show that property three and five each imply property one. Suppose  $|f| \leq \sum 2^n \chi_{E_n}$ . Then  $F(2^m) \leq \sum_{n \geq m} |E_n|$ . Now

$$\|\lambda F(\lambda)^{1/p}\|_{L^q((2^m, 2^{m+1}], \mathbf{R}^+)} \leq 2q^{-1/q} \left( \sum_{k=0}^{\infty} |E_{m+k}| \right)^{1/p} 2^m$$

Using Minkowski's integral inequality, provided  $q \geq p$ , we find

$$\begin{aligned}
\|\lambda F(\lambda)^{1/p}\|_{L^q(\mathbf{R}^+)}^p &\leq 2^p q^{-p/q} \left( \sum_{m \in \mathbf{Z}} \left( \sum_{k=0}^{\infty} 2^{pm} |E_{m+k}| \right)^{q/p} \right)^{p/q} \\
&\leq 2^p q^{-p/q} \sum_{k=0}^{\infty} \left( \sum_{m \in \mathbf{Z}} 2^{qm} |E_{m+k}|^{q/p} \right)^{p/q} \\
&\leq \left( 2^p q^{-p/q} \sum_{k=0}^{\infty} 2^{-pk} \right) \|2^m |E_m|^{1/p}\|_{l^q(\mathbf{Z})}^p
\end{aligned}$$

Thus  $\|\lambda F(\lambda)^{1/p}\|_{L^q(\mathbf{R}^+)} \leq q^{-1/q} \|2^m |E_m|^{1/p}\|_{l^q(\mathbf{Z})}$  □

From this theorem, we see that as  $q \rightarrow 0$ , the  $L^{p,q}$  norms become stronger and stronger. However, they only differ from the  $L^p$  norms ‘by a logarithmic factor’, which explains why they only seem to appear in the fine tuning of arguments. To see one example of this phenomenon, let  $f$  be a function on  $\{1, \dots, N\}$  with  $\|f\|_{p,\infty} = 1$ . Then  $\|f\|_\infty \leq 1$ , since for each  $\varepsilon$ , there are  $1/(1+\varepsilon)^p < 1$  points taking a value exceeding  $1+\varepsilon$ . If for a fixed  $\lambda$ , we set  $E = \{x : |f(x)| < \lambda\}$  be points that are too small, this means that

$$\|f\|_p^p \lesssim_p |E| \lambda^p + \sum_n 2^{np} F(2^n) \lesssim N \lambda^p + \log(1/\lambda)$$

If  $\lambda = 1/(1+N)^{1/p}$ , we get  $\|f\|_p^p \lesssim_p 1 + \log(1+N)/p \lesssim \log(1+N)$ , so  $\|f\|_p \lesssim \log(1+N)^{1/p}$ . Thus in general, we find that  $\|f\|_p \lesssim \log(1+N)^{1/p} \|f\|_{p,\infty}$ , which shows that the weak norms differ by a logarithm in the width of the space. Of course, for spaces with an ‘infinite width’, we might have  $\|f\|_{p,\infty} < \infty$ , but  $\|f\|_p = \infty$ , but this only happens if  $f$  ‘barely’ fails to be integrable by a logarithmic factor.

**Theorem 2.11.** *If  $1/p = 1/p_0 + 1/p_1$ , and  $1/q = 1/q_1 + 1/q_2$ , then*

$$\|fg\|_{p,q} \lesssim_{p_1,q_1,p_2,q_2} \|f\|_{p_0,q_0} \|g\|_{p_1,q_1}$$

## 2.4 Metric Structure of Norms

The geometric interest of these quantitative norms are that we can define a ‘distance’ between two functions  $f$  and  $g$  by the magnitude of a norm of the function  $f - g$ . Thus, for instance, the  $L^1(X)$  can be given a geometric structure by defining the distance between two functions  $f$  and  $g$  as  $\|f - g\|_p$ . The fact that we can treat functions as points in an ‘infinite dimensional geometry’ is the fundamental foundation of the field of functional analysis. For  $p \geq 1$ , these norms do actually satisfy the triangle inequality, giving rise to a metric. For  $0 < p < 1$ , the norms give rise to a quasimetric.

One reason to discuss norms in the context of measure theory is that it enlightens us on the sense that the Lebesgue integral is that it is a ‘completion’ of the Riemann integral. We have already seen this principle in the

discussion of the metric  $d(E, F) = |E \triangle F|$ , for which the family of Lebesgue measurable sets ‘complete’ the metric on Jordan measurable sets or elementary sets. Norms enable us to see this principle in a deeper context, for a norm gives rise to a metric on the family of measurable functions, and the Riemann integrable functions in this metric are dense in this space, so that the Lebesgue functions ‘fill the holes’ with respect to any of the given norms. The triangle inequality for the norms says that the norm works like a metric on space.

The triangle inequality is only ever sharp if  $f$  and  $g$  are correlated in some manner. For instance, if the supports of  $f$  and  $g$  are disjoint, then we find

$$\|f + g\|_p^p = \int (f + g)^p = \int f^p + g^p = \|f\|_p^p + \|g\|_p^p$$

so  $\|f + g\|_p = (\|f\|_p^p + \|g\|_p^p)^{1/p}$ , for any  $0 < p < \infty$ , with the natural infinite extension being  $\|f + g\|_\infty = \max(\|f\|_\infty, \|g\|_\infty)$ . In fact, the only time where  $\|f + g\|_p = \|f\|_p + \|g\|_p$  is when  $f$  and  $g$  are positive scalar multiples of one another (up to a set of measure zero). To see this, note that if equality held, then we would have, except on a set of measure zero,  $|f + g|^p = \theta^{1-p}|f|^p + (1 - \theta)^{1-p}|g|^p$ . We claim that this forces  $g = (1 - \theta)/\theta f$  (TODO).

If  $0 < p < 1$ , the triangle inequality isn’t satisfied, which can be seen in the case of the  $l_p$  norm on  $\mathbf{R}^n$ , which isn’t convex. However, the discrete sum implies we have a pointwise estimate  $|A + B|^p \leq |A|^p + |B|^p$ , so that we have  $\|f + g\|_p^p \leq \|f\|_p^p + \|g\|_p^p$ . In fact, for  $p < 1$  we have

$$\frac{1}{N^{1/p}} \left\| \sum_{n=1}^N f_n \right\|_p \leq \left\| \frac{1}{N} \sum_{n=1}^N |f_n| \right\|_p \leq \frac{1}{N} \sum_{n=1}^N \|f_n\|_p$$

To prove this, since  $1/p \geq 1$ , we can apply convexity to conclude

$$\left( \frac{1}{N} \int \left( \sum_{n=1}^N |f_n| \right)^p \right)^{1/p} \leq \left( \frac{1}{N} \int \sum_{n=1}^N |f_n|^p \right)^{1/p} \leq \frac{1}{N} \sum_{n=1}^N \left( \int |f_n|^p \right)^{1/p}$$

This is ‘almost’ as good as a triangle inequality for the purposes of defining a geometric space. We call such a quantity a quasinorm.

The next most important inequality in the analysis of measurable functions is Hölder’s inequality, which enables us to quantify the norm of the product of two functions in terms of the norm of the original two functions.

**Theorem 2.12.** *If  $0 < p, q, r \leq \infty$ , and  $1/p + 1/q = 1/r$ , then  $\|fg\|_r \leq \|f\|_p \|g\|_q$ .*

*Proof.* We may assume  $p, q$ , and  $r$  are finite, for the other cases are trivial. By normalization, we may assume  $\|f\|_p = \|g\|_q = 1$ . If  $\theta = r/q$ ,  $1 - \theta = r/p$ , then we must show

$$\int |f|^{p(1-\theta)} |g|^{q\theta} \leq 1$$

But this follows by convexity of the map  $\theta \mapsto A^{1-\theta} B^\theta$ , for any  $A, B$ , we can apply a pointwise estimate, and integrate.  $\square$

*Remark.* The symmetry of the inequality under the homogenities of functions explains why the two must occur the same number of times on both sides of the equation. The homogeneity under the scaling of the underlying measure is the reason why  $1/p + 1/q = 1/r$ .



## Chapter 3

# 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 property to more general classes 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 where does  $F'(x) = f(x)$ . In analyzing this problem, we shall use a fundamental tool in modern analysis, the technique of maximal functions. This technique enables one to generalize convergence results for families of operators which have already been established for a dense family of functions. We summarize the idea of the technique in the next theorem.

**Theorem 3.1.** *If  $T_t$  is a family of linear operators on  $L^p(X)$ , then we can define the pointwise maximal operator  $(T^*f)(x) = \sup_t |T_t f(x)|$ . If there exists  $q < \infty$  such that  $T^*$  is weak  $(p, q)$ , then the set  $\{f \in L^p(X) : \lim_t T_t f(x) = f(x) \text{ a.e.}\}$  is closed in  $L^p(X)$ .*

*Proof.* Let  $f_n$  be a sequence of functions such that  $\lim_t (T_t f_n)(x) = f_n(x)$  almost everywhere, which converge in  $L^p(X)$  to some function  $f$ . Then for

each  $\lambda > 0$ , if  $\|T^*g\|_{(q,\infty)} \leq C\|g\|_p$ , for all  $g$ , then

$$\begin{aligned}
& |\{x : \limsup_t |T_t f(x) - f(x)| > \lambda\}| \\
& \leq |\{x : \limsup_t |T_t(f - f_n)(x) - (f - f_n)(x)| > \lambda\}| \\
& \leq |\{x : |T^*(f - f_n)(x)| > \lambda/2\}| + |\{x : |(f - f_n)(x)| > \lambda/2\}| \\
& \leq \left(\frac{2C\|f - f_n\|_p}{\lambda}\right)^q + \left(\frac{2\|f - f_n\|_p}{\lambda}\right)^p
\end{aligned}$$

as  $n \rightarrow \infty$ , this quantity tends to zero. Thus for all  $\lambda > 0$ ,

$$|\{x : \limsup_t |T_t f(x) - f(x)| > \lambda\}| = 0$$

Taking  $\lambda \rightarrow 0$  gives that  $\limsup_t |T_t f(x) - f(x)| = 0$  almost everywhere, and this was the result we were required to prove.  $\square$

The main scenario in which this theorem can be applied is when we can easily establish a pointwise convergence result for suitably regular functions, which are norm dense in the class of all functions we want to obtain pointwise convergence for. While we might not be able to get pointwise convergence everywhere for all such functions, one can use the theorem above to get almost everywhere pointwise convergence. All this requires is to understand the properties of the maximal operator  $T^*$ , which is often the most difficult part of solving the problem.

### 3.1 Maximal Functions and Differentiation

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

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

Since  $f$  is integrable, for every  $\varepsilon > 0$ , there is  $\delta > 0$  such that if  $|E| \leq \delta$ ,  $\int_E |f| \leq \varepsilon$ . This implies that  $F$  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 whether, 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 3.2.** *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  $f \in L^1[a, b]$ , we employ maximal function techniques. Thus we introduce the **Hardy-Littlewood maximal function**

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

This function provides an upper bound on the averages of a function  $f$ . We would like to conclude that  $Mf$  is an integrable function, so we can apply dominated convergence-type results, but in fact,  $Mf$  is almost never integrable.

**Example.** Let  $f \in L^1(\mathbf{R}^d)$  be nonzero. Then, for suitably large  $R \geq 1$ ,

$$\int_{B_R(0)} |f(x)| dx \geq \|f\|_1 - \varepsilon$$

For each  $x \in \mathbf{R}^d$ ,  $B_R(0) \subset B_{|x|+R}(x)$  and so

$$Mf(x) \geq \int_{B_{|x|+R}(x)} |f(y)| dy \gtrsim \frac{\|f\|_1 - \varepsilon}{(|x| + R)^d} \gtrsim \frac{1}{|x|^d}$$

But this means,  $Mf$  is non-integrable. Note, however, that it appears to be on the ‘border’ of integrability. This is why we introduce the weak  $L^1$  inequalities below. In certain cases,  $Mf$  isn’t even locally integrable. If  $f(x) = 1/|x| \log |x|^2$ , then the fact that for  $x \geq 0$

$$\begin{aligned} \frac{1}{2h} \int_{x-h}^{x+h} \frac{dy}{|y| \log |y|^2} &= \frac{1}{2h} \left( \frac{1}{\log(x-h)} - \frac{1}{\log(x+h)} \right) \\ &= \frac{1}{2x \log x} + O\left(\frac{h}{\log x}\right) \end{aligned}$$

Implies that

$$Mf(x) \geq \frac{1}{2x \log x}$$

and so  $Mf$  isn’t integrable about the origin.

It is easy to see that  $M$  is  $(\infty, \infty)$  bounded, since the average of a function on a ball is bounded by its maximum on that ball. It is not  $(1, 1)$  bounded, but we will soon show it is weak  $(1, 1)$ , which is sufficient to apply maximal function techniques. Furthermore, by the Marcinkiewitz interpolation theorem, this implies  $M$  is  $(p, p)$  bounded for all  $1 < p \leq \infty$ . Proving  $M$  is weak  $(1, 1)$  relies on a combinatorial result about coverings of sets, known as the Vitali covering lemma.

**Lemma 3.3** (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 3.4.** *The function  $Mf$  is measurable, and for any  $\alpha > 0$ ,*

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

Thus  $Mf$  has a weak  $(1, 1)$  bound  $\|Mf\|_{1,\infty} \leq 3^d \|f\|_1$ .

*Proof.* The measurability of  $Mf$  is easy, because  $Mf$  is a lower semicontinuous function, and all lower semicontinuous functions are measurable. If  $Mf(x) > \alpha$ , then 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  $y \in B$ ,  $Mf(y) > \alpha$ . Set  $E_\alpha = \{x \in \mathbf{R}^d : Mf(x) > \alpha\}$ . 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

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

Since  $K$  is arbitrary, inner regularity says  $|E_\alpha| \leq 3^d \|f\|_1 / \alpha$ .  $\square$

We immediately obtain the Lebesgue differentiation theorem from our discussion of maximal function techniques.

**Corollary 3.5.** *If  $f \in L^1(\mathbf{R}^d)$ , then for almost every  $x$ ,*

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

The integrability of  $f$  over all of  $\mathbf{R}^d$  isn't *really* applied anywhere in the proof of the Lebesgue differentiation theorem. 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  $f$ .

## 3.2 Lebesgue Density Theorem

If  $E$  is a measurable subset of  $\mathbf{R}^d$ , and  $x \in \mathbf{R}^d$ , we say  $x$  is a point of **Lebesgue density** of  $E$ , or has **metric 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 local points around  $x$  as is possible. Since  $\chi_E \in L^1_{\text{loc}}(\mathbf{R}^d)$ , we can apply the Lebesgue differentiation theorem to immediately obtain an interesting result.

**Theorem 3.6** (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  $E$ . Note that if  $|B \cap E|, |B \cap F| \geq \alpha|B|$ , then a union bound gives  $|B \cap E \cap F| \geq (2\alpha - 1)|B|$ . 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  $\alpha_1, \dots, \alpha_n$ ,  $0$  is a point of Lebesgue density for  $\alpha_1 E \cap \dots \cap \alpha_n E$ , which in particular implies the existence of  $x_i \rightarrow 0$  with  $x_i \in \alpha_1 E \cap \dots \cap \alpha_n E$ , hence

$\alpha_1^{-1}x_i, \dots, \alpha_n^{-1}x_i \in E$ . This is very difficult to prove without the existence of the Lebesgue density theorem, because of the discreteness of the equations. Applying these results with  $\alpha_1 = 1$ ,  $\alpha_2 = 1/2$ ,  $\alpha_3 = 1/3$ , and so on, we conclude that sets of positive measure in  $\mathbf{R}$  contain 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

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

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

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

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

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. However, the Lebesgue set of the universal function constructed above is the largest of any function in the equivalence class, which is sometimes taken as the canonical Lebesgue set of the class. Alternatively, this version of the Lebesgue set can be taken as the points  $x$  such that there exists  $a_x$  with

$$\lim_{\substack{|B| \rightarrow 0 \\ x \in B}} \oint_B |f(y) - a_x| = 0$$

Then it is clear that the Lebesgue set of two functions agree if they are equal almost everywhere.

### 3.3 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 certain balls  $B$  around  $x$ . In particular, if we define

$$M_U(f) = \sup_{U_\alpha} \oint_{U_\alpha} |f(x)| dx$$

Then

$$M_U(f) = \sup_{U_\alpha} \oint_{U_\alpha} |f(x)| dx \leq c^{-1} \sup_B \oint_B |f(x)| dx = c^{-1}(Mf)(x)$$

In particular,  $M_U \lesssim M$ , which implies  $M_U$  is weak  $(1,1)$ . We therefore conclude that for any locally integrable  $f$ ,

$$\lim_{\substack{U_\alpha \rightarrow 0 \\ x \in U_\alpha}} \oint_{U_\alpha} f(y) dy = f(x)$$



Thus the differentiation theorem easily generalizes to averages over any sets which don't differ too much from a ball.

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

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

*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.  $\square$

### 3.4 The Calderon Zygmund Decomposition

For a fixed integer  $n$ , we let  $\mathcal{B}(n)$  denote the family of all sidelength  $1/2^n$  half open cubes with corners on the lattice  $(\mathbf{Z}/2^n)^d$ , i.e.

$$\mathcal{B}(n) = \{[a_1, b_1) \times \cdots \times [a_d, b_d) : a_i, b_i \in \mathbf{Z}/2^n\}$$

Thus  $\mathcal{B}(n)$  partitions  $\mathbf{R}^d$  for each integer  $n$ . We refer to  $\mathcal{B} = \bigcup \mathcal{B}(n)$  as the family of all **dyadic** cubes. An important property is that if  $I$  and  $J$  are dyadic, either one cube contains the other, or they are completely disjoint. This makes Vitali type covering arguments redundant, because every union of dyadic cubes can be written as the disjoint union of dyadic cubes (except if the union is  $\mathbf{R}^d$ ). Thus we can obtain weak type bounds for the corresponding ‘dyadic maximal function’ very easily, and ignore the dimensional constants previously required.

**Theorem 3.9.** *If*

$$(Mf)(x) = \sup_{\substack{I \in \mathcal{B} \\ x \in I}} \int_I |f(x)| dx$$

*Then*  $\|Mf\|_{(1,\infty)} \leq \|f\|_1$ .

*Proof.* Fix  $\varepsilon > 0$ , and let  $E_{\alpha,\varepsilon} = \{x \in \mathbf{R}^d : Mf(x) > \alpha\}$ . Since  $f$  is integrable,  $E_{\alpha} \neq \mathbf{R}^d$ . Then  $E_{\alpha}$  decomposes into union of dyadic cubes  $I_1, I_2, \dots \in \mathcal{B}$ , which we can assume are disjoint. Thus

$$|E_{\alpha}| = \sum_{k=1}^{\infty} |I_k| \leq \frac{1}{\alpha} \sum_{k=1}^{\infty} \int_{I_k} |f(y)| dy \leq \frac{\|f\|_1}{\alpha}. \quad \square$$

Thus we find that if we consider the conditional expectation operators

$$E_k f(x) = \sum_{I \in \mathcal{B}(k)} \left( \int_I f \right) \chi_I$$

Then  $E_k f(x) \rightarrow f(x)$  for almost all  $x$ , if  $f \in L^1_{\text{loc}}(\mathbf{R}^d)$ . Using this, we can obtain a very useful way to discretize  $f$  on dyadic cubes, known as the Calderon-Zygmund decomposition.

**Theorem 3.10.** *Given an integrable, non-negative  $f$ , and  $\lambda > 0$ , there is a disjoint collection of dyadic cubes  $I_1, \dots, I_N$  such that  $\sum |I_k| \leq \lambda^{-1} \|f\|_1$ ,  $f(x) \leq \lambda$  for almost every  $x \notin \bigcup I_k$ , and for each  $k$ ,*

$$\int_{I_k} f(x) dx = 0 \quad \text{and} \quad \lambda < \int_{I_k} |f(x)| < 2^d \lambda$$

*Proof.* Let  $M$  be the dyadic maximal operator. Write  $E = \{x : Mf(x) > \lambda\}$ . Because  $f$  is integrable,  $E \neq \mathbf{R}^d$ . Thus  $E$  is the disjoint union of dyadic cubes  $I_1, \dots, I_N$ , and we can assume  $N$  is minimal, so that if  $I_k$  has sidelength  $2^i$ , then the unique sidelength  $2^{i+1}$  dyadic cube  $J_k$  containing  $I_k$  is not a subset of  $E$ . By construction, we know that

$$\lambda < \oint_{I_k} f(x)$$

But also

$$|I_k| \oint_{I_k} f(x) \leq |J_k| \oint_{J_k} f(x) \leq (2^d |I_k|) \lambda$$

which gives the upper bound on the averages. If  $x$  is not contained in any of the cubes  $I_1, \dots, I_N$ , then  $(E_k f)(x) \leq \lambda$  for all  $k$ , and thus for almost every  $x$  not in  $I_1, \dots, I_N$ ,  $f(x) \leq \lambda$ . The weak bound for the maximal function gives that  $\sum |I_k| \leq \lambda^{-1} \|f\|_1$ , which completes the proof.  $\square$

*A useful way to think about the Calderon Zygmund decomposition is a way to decompose integrable functions  $f = f_0 + f_1$ , where*

$$f_0(x) = \begin{cases} \oint_{I_k} f(x) & x \in I_k \\ f(x) & x \notin I_k \end{cases}$$

*Then  $\|f_0\|_\infty \leq 2^d \lambda$ , which combined with the fact that  $f_0$  is integrable, implies that  $f_0$  is fairly analytically stable. It is the ‘good part’ of  $f$ . On the other hand,  $f_1$  is the ‘bad part’ of  $f$ , but it is supported on a set with size  $\lambda^{-1} \|f\|_1$ , and its average on each of the  $I_k$  is equal to zero.*

### 3.5 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  $\delta > 0$ , as  $\alpha \rightarrow \infty$ ,

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

It requires only basic analysis to verify good kernel convergence.

**Theorem 3.11.** *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

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

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

$$\frac{F_N(x)}{2\pi} = \begin{cases} \frac{1}{2\pi N} \frac{\sin^2(Nx/2)}{\sin^2(x/2)} \end{cases}$$

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 be given a precise meaning, either in the theory of Lebesgue-Stieltjes measures or as a ‘generalized function’, but we don’t need it to discuss the actual convergence results of the functions  $K_\delta$ .

**Theorem 3.12.** 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.  $\square$

### 3.6 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.

If  $f$  is a complex valued function on  $[a, b]$ , and  $P$  is a partition, we can consider its variation on a partition  $P = a \leq t_0 < \dots < t_n \leq b$  to be

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

we say  $f$  has **bounded variation** if there is a constant  $M$  such that for any partition  $P$ ,  $V(f, P) \leq M$ . This implies that, since the net  $P \mapsto V(f, P)$  is increasing, the net converges to a value  $V(f) = V(f, a, b)$ , the **total variation** of  $f$  on  $[a, b]$ . The problem of variation 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  $P = a \leq t_0 \leq \dots \leq t_n$ , 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  $x(t_{i-1})$  and  $x(t_i)$  should be shorter than the length of  $x$  between  $t_{i-1}$  and  $t_i$ . It therefore makes sense to define the **length** of  $x$  as

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

The triangle inequality implies that the map  $P \mapsto L_P(x)$  is an increasing net, so  $L$  is also the limit of the meshes as they become finer and finer. If  $L(x) < \infty$ , we



say  $x$  is a **rectifiable curve**. One problem is to determine what analytic conditions one must place on  $x$  in order to guarantee regularity, and what further conditions guarantee that, if  $x_i$  is differentiable almost everywhere,

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

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

**Theorem 3.13.** 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  $A \sum V(x_i, P) \leq L_P(x) \leq B \sum V(x_i, P)$  gives the required result.  $\square$

**Example.** If  $f$  is a real-valued, monotonic, increasing 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  $V(f, P) \leq M(b - a)$  for all partitions  $P$ .

**Example.** Consider the functions  $f$  defined on  $[0, 1]$  with

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

Then  $f$  has bounded variation on  $[0, 1]$  if and only if  $a > b$ . The function oscillates from increasing to decreasing on numbers of the form  $x = (n\pi)^{-1/b}$ , so the total variation is described as

$$V(f) = 1 + \sum_{n=1}^{\infty} (n\pi)^{-a/b} + ((n+1)\pi)^{-a/b}$$

This sum is finite precisely when  $a/b > 1$ . Thus functions of bounded variation cannot oscillate too widely, too often.

The next result is a decomposition theorem for bounded variation functions into bounded increasing and decreasing functions. We define the **positive variation** of a real valued function  $f$  on  $[a, b]$  to be

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

The **negative variation** is

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

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

**Lemma 3.14.** If  $f$  is real-valued and has bounded variation on  $[a, b]$ , then for all  $a \leq x \leq b$ ,

$$f(x) - f(a) = P(f, a, x) - N(f, a, x)$$

$$V(f) = P(f, a, b) + N(f, a, b)$$

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

$$\left| P(f, a, x) - \sum_{f(t_i) \geq f(t_{i-1})} f(t_i) - f(t_{i-1}) \right| < \varepsilon$$

$$\left| N(f, a, x) + \sum_{f(t_i) \leq f(t_{i-1})} f(t_i) - f(t_{i-1}) \right| < \varepsilon$$

It follows that

$$|f(x) - f(a) - [P(f, a, x) - N(f, a, x)]| < 2\varepsilon$$

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

$$f(x) = [f(a) + P(f, a, x)] - N(f, a, x)$$

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

**Theorem 3.15.** *If  $f$  is a continuous function of bounded variation, then*

$$x \mapsto V(f, a, x) \quad x \mapsto V(x, b)$$

*are continuous functions.*

*Proof.*  $V(f, a, x)$  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  $V(f, a, x_1) \geq V(f, a, x) - \varepsilon$ . If we consider a partition

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

where  $|V(f, P) - V(f, a, x)| \leq \varepsilon$ , 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

$$\begin{aligned} V(f, a, x_1) &\geq V(f, Q) = V(f, P) - |f(x) - f(t_{n-1})| + |f(x_1) - f(t_{n-1})| \\ &\geq V(f, P) - \varepsilon \geq V(f, a, x) - \varepsilon \end{aligned}$$

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 3.16** (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  $(a_n, b_n)$ , where  $f(b_n) = f(a_n)$ . 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  $f(b_n) = f(a_n)$  except if  $a_n = a$ , where we can only conclude  $f(a_n) \leq f(b_n)$ .*

*Proof.* The openness is clear, and the fact that  $E$  can be broken into disjoint intervals follows because of the characterization of open sets in  $\mathbf{R}$ . If

$$E = \bigcup (a_n, b_n)$$

Then  $f(a_n + h) \leq f(a_n)$  and  $f(b_n + h) \leq f(b_n)$ , implying in particular that  $f(b_n) \leq f(a_n)$ . If  $f(b_n) < f(a_n)$ , then choose  $f(b_n) < c < f(a_n)$ . The intermediate value theorem implies there is  $x$  with  $f(x) = c$ , and we may choose the largest  $x \in [a_n, b_n]$  for which this is true. Then since  $x \in (a_n, b_n)$ , there is  $y \in (x, b_n)$  with  $f(x) < f(y)$ , and by the intermediate value theorem, since  $f(b_n) < f(x) < f(y)$ , there must be  $x' \in (y, b_n)$  with  $f(x') = c$ , contradicting that  $x$  was chosen maximally. The proof for closed intervals operates on the same principles.  $\square$

**Theorem 3.17.** *If  $f$  is increasing and continuous on  $[a, b]$ , 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.* the theorem in the It suffices to assume that  $f$  is increasing, and we shall start by proving case assuming  $f$  is continuous. We define

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

and the four Dini derivatives

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

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

Clearly,  $D_+ f \leq D^+ f$  and  $D_- f \leq D^- f$ . It suffices to show  $D^+ f(x) < \infty$  for almost every  $x$ , and  $D^+ f(x) \leq D_- f(x)$  for almost every  $x$ , because if we

consider the function  $g(x) = -g(-x)$ , then we obtain  $D^-f(x) \leq D_+f(x)$  for almost every  $x$ , so

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

for almost every  $x$ , implying all values are equal, and that the derivative exists at  $x$ .

For a fixed  $\gamma > 0$ , consider  $E_\gamma = \{x : D^+f(x) > \gamma\}$ . Since each  $\Delta_h f$  is continuous, the supremum of the  $\Delta_h f$  over any index set is lower semicontinuous, and since

$$D^+f(x) = \lim_{h \rightarrow 0} \sup_{0 \leq s \leq h} \Delta_h f(x+s)$$

can be expressed as the countable limit of these lower semicontinuous functions,  $D^+f$  is measurable, hence  $E_\gamma$  is measurable. Now consider the shifted function  $g(x) = f(x) - \gamma x$ . If  $\bigcup (a_i, b_i)$  is the set obtainable from  $g$  from the rising sun lemma, then  $E_\gamma \subset \bigcup (a_i, b_i)$ , for if  $D^+f(x) > \gamma$ , then there is  $h > 0$  arbitrarily small with  $\Delta_h f(x) > \gamma$ , hence  $f(x+h) - f(x) > \gamma h$ , hence  $g(x+h) > g(x)$ . We know that  $g(a_k) \leq g(b_k)$ , so  $f(b_k) - f(a_k) \geq \gamma(b_k - a_k)$ , so

$$|E_\gamma| \leq \sum (b_k - a_k) \leq \frac{1}{\gamma} \sum f(b_k) - f(a_k) \leq \frac{f(b) - f(a)}{\gamma}$$

Thus  $|E_\gamma| \rightarrow 0$  as  $\gamma \downarrow 0$ , implying  $D^+f(x) = \infty$  only on a set of measure zero.

Now for two real numbers  $r < R$ , we will now show

$$E = \{a \leq x \leq b : D^+f(x) > R \quad D_-f(x) < r\}$$

is a set of measure zero. Letting  $r$  and  $R$  range over all rational numbers establishes that  $D^+f(x) \leq D_-f(x)$  almost surely. We assume  $|E| > 0$  and derive a contradiction. By regularity, we may consider an open subset  $U$  in  $[a, b]$  containing  $E$  such that  $|U| < |E|(R/r)$ . We can write  $U$  as the union of disjoint intervals  $I_n$ . For a fixed  $I_N$ , apply the rising sun lemma to the function  $rx - f(-x)$  on the interval  $-I_N$ , yielding a union of intervals  $(a_n, b_n)$ . If we now apply the rising sun lemma to the function  $f(x) - Rx$  on  $(a_n, b_n)$ , we get intervals  $(a_{nm}, b_{nm})$ , whose union we denote  $U_N$ . Then

$$R(b_{nm} - a_{nm}) \leq f(b_{nm}) - f(a_{nm}) \quad f(b_n) - f(a_n) \leq r(b_n - a_n)$$

then, because  $f$  is increasing,

$$\begin{aligned} |U_N| &= \sum_{nm} (b_{nm} - a_{nm}) \leq \frac{1}{R} \sum_{nm} (f(b_{nm}) - f(a_{nm})) \\ &\leq \frac{1}{R} \sum_n f(b_n) - f(a_n) \leq \frac{r}{R} \sum_n (b_n - a_n) \leq \frac{r}{R} |I_N| \end{aligned}$$

Now  $E \cap I_N$  is contained in  $U_N$ , because if  $x \in E \cap I_N$ , then  $D^+ f(x) > R$  and  $D_- f(x) < r$ , so we can sum in  $N$  to conclude that

$$|E| \leq \sum \frac{r}{R} |I_N| = \frac{r}{R} |U_N| < |E|$$

a contradiction proving the claim.  $\square$

**Corollary 3.18.** *If  $f$  is increasing and continuous, then  $f'$  is measurable, non-negative, and*

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

*Proof.* The fact the  $f'$  is measurable and non-negative results from the fact that the functions  $g_n(x) = \Delta_{1/n} f(x)$  are non-negative and continuous, and  $g_n \rightarrow f'$  almost surely. We know

$$\begin{aligned} \int_a^b f'(x) &\leq \liminf_{n \rightarrow \infty} \int_a^b g_n(x) = \liminf_{n \rightarrow \infty} n \int_a^b [f(x + 1/n) - f(x)] \\ &= \liminf_{n \rightarrow \infty} n \left[ \int_b^{b+1/n} f(x) - \int_a^{a+1/n} f(x) \right] = f(b) - f(a) \end{aligned}$$

where the last equality follows because  $f$  is continuous.  $\square$

*Even for increasing continuous functions, the inequality in the theorem above need not be an equality, as the next example shows, so we need something stronger to obtain our differentiation theorem.*

**Example.** *The Cantor-Lebesgue function is a continuous increasing function  $f$  from  $[0, 1]$  to itself, with  $f(0) = 0$ , and  $f(1) = 1$ , but with  $f'(x) = 0$  almost everywhere. This means*

$$\int_0^1 f'(x) = 0 < 1 = f(1) - f(0)$$

so we cannot obtain equality in general. To construct  $f$ , 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, provided by absolute continuity.

### 3.7 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  $\sum(b_i - a_i) < \delta$ ,  $\sum|f(b_i) - f(a_i)| < \varepsilon$ . Thus the function should be 'essentially constant' over every set of zero measure. 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

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

The Cantor function is not absolutely continuous, since it is constant except on the Cantor set, and we can cover the Cantor set by intervals with total length  $(2/3)^n$  for each  $n$ . Thus it is impossible for the Cantor function to satisfy the fundamental theorem of calculus.

**Theorem 3.19.** If  $g \in L^1(\mathbf{R})$ , and

$$f(x) = \int_a^x g(t) dt$$

then  $f$  is absolutely continuous.

*Proof.* Fix  $\varepsilon > 0$ . We claim that there is  $\delta$  such that if  $|E| < \delta$ , then  $\int_E |g| < \varepsilon$ . Otherwise there are sets  $E_n$  with  $|E_{n+1}| \leq |E_n|/3$  and with  $\int_{E_n} |g| \geq \varepsilon$ . Thus if we define the sets  $E'_m = E_m - \bigcup_{n>m} E_n$  then the  $E'_m$  and we have  $|E_m| \sim |E'_m|$ . Since  $g$  is integrable, we must have  $\sum \int_{E'_n} |g| < \infty$ , so we conclude that as  $N \rightarrow \infty$ ,

$$\sum_{n \geq N} \int_{E'_n} |g| \rightarrow 0$$

Yet for any  $N$ ,

$$\sum_{n \geq N} \int_{E'_n} |g| = \int_{E_N} |g| \geq \varepsilon$$

which is an impossibility. Thus such a  $\delta$  exists for every  $\varepsilon$ , and so if we have disjoint intervals  $(a_n, b_n)$  with  $\sum (b_n - a_n) < \delta$ , then

$$\sum |f(b_n) - f(a_n)| = \sum \left| \int_{a_n}^{b_n} g(t) \right| \leq \sum \int_{a_n}^{b_n} |g| = \int_{\bigcup (a_n, b_n)} |g| < \varepsilon$$

which shows the function is absolutely continuous.  $\square$

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

**Lemma 3.20.** *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  $E_2 = K_1 - \bigcup \overline{B_k^1}$ . If this process repeats itself  $k$  times, then we obtain a family of open sets with total measure greater than or equal to  $k(3^{-d}\delta)$ . But then if we eventually have  $k \geq (|E| - \delta)3^d/\delta$ , then the family of open sets satisfies the requirements of the theorem.  $\square$

**Corollary 3.21.** *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 only 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 3.22.** *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 3.23.** *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.  $\square$

*We now dwell slightly longer on the properties of absolutely continuous functions, which enables us to generalize other properties of integrals found in the calculus. We begin by noting that it is easy to verify that if  $f$  and  $g$  are absolutely continuous functions, then  $fg$  is also absolutely continuous. We know*

$f'$ ,  $g'$ , and  $(fg)'$  exist almost everywhere. But when all three exist simultaneously, the product rule gives  $(fg)' = f'g + fg'$ . The absolute continuity implies that

$$\int_a^b f'g + fg' = \int_a^b (fg)' = f(b)g(b) - f(a)g(a)$$

Thus one can integrate a pair of absolutely continuous functions by parts. Next, we shall show that monotone absolutely continuous functions are precisely those we can use to change variables. One important thing to note is that even if  $f$  is a continuous function, and  $g$  is measurable,  $g \circ f$  need not be measurable. The easy reason to see this is that the inverse image of every open set in  $g$  is measurable, so in order to guarantee  $g \circ f$  is measurable we need the inverse image of every measurable set under  $f$  be measurable.

**Example.** Consider the function  $f(x) = \int_0^x \chi_E(x) dx$ , where  $E$  is a thick Cantor set. Then  $f$  is absolutely continuous, strictly increasing on  $[0, 1]$ , and maps  $E$  to a set of measure zero. This is because  $E = \lim E_n$ , where  $E_n$  is a family of intervals with  $|E_n| \downarrow |E|$ . Then  $f(E_n)$  has total length  $|E_n - E|$ , so as  $n \rightarrow \infty$ , we see  $\lim f(E_n) = f(E)$  has measure zero. This means that  $f(X)$  is measurable for any subset  $X$  of  $E$ , and in particular, if  $X$  is non-measurable, then  $f^{-1}(f(X))$  cannot be measurable, even though  $f(X)$  is measurable. Note that  $f$  is strictly increasing even though its derivatives vanish on a set of positive measure.

The next lemmas will show that even though  $g \circ f$  may not be measurable, this does not really bother us too much when changing variables.

**Lemma 3.24.** *If  $f$  is absolutely continuous, then it maps sets of measure zero to sets of measure zero.*

*Proof.* Let  $E$  be a set of measure zero. Then for each  $\delta > 0$ ,  $E$  is coverable by a family of open intervals with total length  $\delta$ . But if  $\delta$  is taken small enough, this means that  $f(E)$  is coverable by a family of open intervals with total length bounded by  $\varepsilon$ , for any  $\varepsilon$ .  $\square$

This property of absolutely continuous functions is independent of the properties of the Euclidean domain as its domain, and is used in the generalization of absolute continuity to more general domains, or even to measures. If  $f$  is absolutely continuous, then the image of every interval is an interval, and since  $f(\bigcup K_n) = \bigcup f(K_n)$ , this implies that the image of a  $F_\sigma$  set is measurable. But since every measurable set of  $\mathbf{R}$  differs from a  $F_\sigma$  set by a set of measure zero,

the image of every Lebesgue measurable set is Lebesgue measurable. The reverse is almost true.

**Lemma 3.25.** *If  $f$  is absolutely continuous, and  $E$  measurable, then the set*

$$f^{-1}(E) \cap \{x : f'(x) > 0\}$$

*is measurable.*

*Proof.* If  $E$  is an open set, then

$$|E| = \int_{f^{-1}(E)} f'(x) dx$$

It suffices to prove this when  $E$  is an interval, and then this is just the theorem of differentiation for absolutely continuous functions. But then applying the dominated convergence theorem shows that this equation remains true if  $E$  is an  $G_\delta$  set. Furthermore, this means the theorem is true if  $E$  is a closed set, and so by applying the monotone convergence theorem, the theorem is true if  $E$  is an  $F_\sigma$  set. But if  $E$  is an arbitrary measurable set, then for every  $\varepsilon$  there are  $F_\sigma$  and  $G_\delta$  sets  $K \subset E \subset U$  with  $|U - K| = 0$ . But

$$\alpha |f^{-1}(U - K) \cap \{f' \geq \alpha\}| \leq \int_{f^{-1}(U - K)} f'(x) dx = |U - K| = 0$$

Thus  $f^{-1}(U - K) \cap \{f' \geq \alpha\}$  is a set of measure zero, and so in particular by completeness, every set contained in this set is measurable, in particular  $f^{-1}(U - E) \cap \{f' \geq \alpha\}$  is measurable. But now this means

$$\{f' \geq \alpha\} - f^{-1}(U - E) \cap \{f' \geq \alpha\} = f^{-1}(E) \cap \{f' \geq \alpha\}$$

is measurable. Taking  $\alpha \downarrow 0$  completes the proof.  $\square$

*Because of this, even though  $g \circ f$  is not necessarily measurable,  $(g \circ f)f'$  is always measurable if  $f$  is absolutely continuous. Thus the expression  $\int (g \circ f)f'$  makes sense, and thus we can always interpret the change of variables formula.*

**Theorem 3.26.** *If  $f$  is absolutely continuous, and  $g$  is integrable, then*

$$\int g(f(x))f'(x) dx = \int g(y) dy$$

*Proof.* Using the notation in the last proof, if  $E$  is measurable, then

$$|K| = \int_{f^{-1}(K)} f'(x) dx \leq \int_{f^{-1}(E)} f'(x) dx \leq \int_{f^{-1}(U)} f'(x) dx = |U|$$

and  $|U| = |K| = |E|$ , so that for any measurable set  $E$ ,

$$|E| = \int_{f^{-1}(E)} f'(x) dx$$

This implies the theorem we need to prove is true whenever  $g$  is the characteristic function of any measurable set. But then by linearity, it is true for any simple function. By monotone convergence, it is then true for any non-negative function, and then by partitioning  $g$  into the sum of simple functions, we obtain the theorem in general.  $\square$

### 3.8 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 3.27.** *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 3.28.**  *$J'$  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  $\delta = 0$ .  $\square$

*This concludes our argument that every function of bounded variation has a derivative almost everywhere, because every such function can be uniquely written (up to a shift in the range of the functions) as the sum of a continuous function and a jump function. If  $f$  is a function with bounded variation, then the function*

$$F(x) = \int_0^x f'(x)$$

*is absolutely continuous, and  $f - F$  is a continuous function with derivative zero almost everywhere. The fact that this decomposition is unique up to a shift as well (which can easily be seen in the case of an increasing function, from which the general case follows) leads us to refer to this as the **Lebesgue decomposition** of a function of bounded variation on the real line.*

### 3.9 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  $\sqrt{2}$ .

**Theorem 3.29.** *If a curve is parameterized by absolutely continuous functions  $x$  and  $y$  on  $[a, b]$ , then the curve is rectifiable, and has length*

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

*Proof.* This proof can be reworded as saying if  $f$  is complex-valued and absolutely continuous, then its total variation can be expressed as

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

If  $P = \{a \leq t_1 < \cdots < t_n \leq b\}$  is a partition, then

$$\sum |f(t_{n+1}) - f(t_n)| = \sum \left| \int_{t_n}^{t_{n+1}} f'(t) dt \right| \leq \sum \int_{t_n}^{t_{n+1}} |f'(t)| dt \leq \int_a^b |f'(t)| dt$$

so  $V(f, a, b) \leq \int_a^b |f'(t)| dt$ . To prove the converse inequality, fix  $\varepsilon > 0$ , and find a step function  $g$  with  $f' = g + h$ , with  $\|h\|_1 \leq \varepsilon$ . If  $G(x) = \int_a^x g(t) dt$  and  $H(x) = \int_a^x h(t) dt$ , then  $F = G + H$ , and  $V(f, a, b) \geq V(G, a, b) - V(H, a, b) \geq V(G, a, b) - \varepsilon$ , and if we partition  $[a, b]$  into  $a = t_0 < \cdots < t_N$ , where  $G$  is constant on each  $(t_n, t_{n+1})$ , then

$$\begin{aligned} V(G, a, b) &\geq \sum |G(t_n) - G(t_{n-1})| = \sum \left| \int_{t_{n-1}}^{t_n} g(t) dt \right| \\ &= \sum \int_{t_{n-1}}^{t_n} |g(t)| dt = \int_a^b |g(t)| dt \geq \|f'\|_1 - \varepsilon \end{aligned}$$

Letting  $\varepsilon \rightarrow 0$  now gives the result.  $\square$

*It is interesting to note that any rectifiable curve has a special parameterization by arclength, i.e. a parameterization  $(x(t), y(t))$  such that if  $L$  is the length function associated to the parameterization, then  $L(A, B) = B - A$ . This is obtainable by inverting the length function.*



**Theorem 3.30.** *If  $z = (x, y)$  is a parameterization of a rectifiable curve by arclength, then  $x$  and  $y$  are absolutely continuous, and  $|z'| = 1$  almost everywhere.*

*Proof.* For any  $s < t$ ,

$$t - s = L(s, t) = V(f, s, t) \geq |z(t) - z(s)|$$

so it follows immediately that  $|z|$  is an absolutely continuous function, and  $|z'| \leq 1$  almost surely. But now we know that

$$\int_a^b |z'(t)| = b - a$$

and this equality can now only hold if  $|z'(t)| = 1$  almost surely.  $\square$

### 3.10 Bounded Variation in Higher Dimensions

*Since the higher dimensional Euclidean domains do not have an ordering, it is impossible to define their length by partitioning their domain, and the meaning of a jump discontinuity is no longer clear. However, there are properties equivalent to having bounded variation which are more extendable to higher dimensions.*

**Theorem 3.31.** *The following properties of  $f : \mathbf{R} \rightarrow \mathbf{R}$  are equivalent, for some fixed finite constant  $A$ .*

- *$f$  can be modified on a set of measure zero so that it has bounded variation not exceeding  $A$ .*
- *$\int |f(x+h) - f(x)| \leq A|h|$  for all  $h \in \mathbf{R}$ .*
- *For any  $C^1$  function  $\varphi$  with compact support,  $|\int f(x)\varphi'(x)| \leq A\|\varphi\|_\infty$ .*

*Proof.* If  $V(f) = A$ , where  $A < \infty$ , then we can write  $f = f^+ - f^-$ , where  $f^+$  and  $f^-$  are both increasing functions, and with  $V(f) = V(f^+) + V(f^-)$ . It then follows that  $|f(x+h) - f(x)| \leq (f^+(x+h) - f^+(x)) + (f^-(x) - f^-(x+h))$ , so it suffices to prove the second property by assuming  $f$  is increasing. But

then by the monotone convergence theorem, assuming  $h > 0$  without loss of generality,

$$\int |f(x+h) - f(x)| = \lim_{y \rightarrow \infty} \int_{-y}^y f(x+h) - f(x) = \lim_{y \rightarrow \infty} \int_y^{y+h} f(x) - \int_{-y-h}^{-y} f(x)$$

The first term of the limit converges to  $hV(f)$ , and the second to zero, completing the first part of the theorem. Now assuming the second point, we prove the third point. Then using the second point, we find

$$\begin{aligned} \left| \int f(x) \varphi'(x) \right| &= \left| \lim_{h \rightarrow 0} \int f(x) \frac{\varphi(x+h) - \varphi(x)}{h} \right| \\ &= \left| \lim_{h \rightarrow 0} \int \frac{f(x-h) - f(x)}{h} \varphi(x) \right| \leq A \|\varphi\|_\infty \end{aligned}$$

Finally, we consider the third point being true. The set of all partitions with rational points is countable. Suppose that for each rational  $P = \{t_0 < \dots < t_N\}$  there is a set  $E_P$  of measure zero for each rational partition  $P$  such that

$$\sum_{n=1}^N \sup_{\substack{x \in [t_{n-1}, t_n] \\ x \notin E_P}} f(x) - \inf_{\substack{x \in [t_{n-1}, t_n] \\ x \notin E_P}} f(x) \leq A$$

Then the union of  $E_P$  over all rational  $P$  has measure zero. We can modify  $f$  on  $E_P$  by setting  $f(x) = \liminf_{y \rightarrow 0} f(x+y)$ , and then  $V(f, P) \leq A$  for all rational partitions  $P$ . If  $Q$  is now any partition, we can find a rational partition  $P$  with  $V(f, P) \geq V(f, Q) - \varepsilon$ , and so  $V(f, P) \leq A - \varepsilon$ . Taking  $\varepsilon \rightarrow 0$  completes the argument. Thus if  $f$  cannot be modified to have finite variation  $A$ , there exists a rational partition  $P$  such that for any set  $E$  of measure zero,

$$\sum_{n=1}^N \sup_{\substack{x \in [t_{n-1}, t_n] \\ x \notin E}} f(x) - \inf_{\substack{x \in [t_{n-1}, t_n] \\ x \notin E}} f(x) > A$$

Thus for any  $\varepsilon$ , there exists  $E_n^+, E_n^- \subset [t_{n-1}, t_n]$  of positive measure such that

$$\sum_{n=1}^N \inf_{x \in E_n^+} f(x) - \sup_{x \in E_n^-} f(x) > A$$

If we consider the polygonal function  $\phi$  which

□

### 3.11 Minkowski Content

Given a set  $K \in \mathbf{R}^n$ , we let  $K^\delta$  denote the open set consisting of points  $x$  with  $d(x, K) < \delta$ . The  $m$  dimensional **Minkowski content** of  $K$  is defined to be

$$\lim_{\delta \rightarrow 0} \frac{1}{\alpha(n-m)} \frac{|K^\delta|}{\delta^m}$$

where  $\alpha(d)$  is the volume of the unit ball in  $d$  dimensions. When this limit exists, we denote it by  $M^m(K)$ . In this section, we mainly discuss the one dimensional Minkowski content in two dimensions, i.e. the values of

$$\lim_{\delta \rightarrow 0} \frac{|K^\delta|}{2\delta^m}$$

and it's relation the length of curves. Since we now only care about the one dimensional Minkowski content, we let  $M(K)$  denote the one dimension Minkowski content.

**Lemma 3.32.** If  $\Gamma = \{z(t) : a \leq t \leq b\}$  is a curve, and  $\Delta$  is the distance between the endpoints of the curve, then  $|\Gamma^\delta| \geq 2\delta\Delta$ .

*Proof.* By rotating, we may assume that both endpoints of the curve lie on the  $x$  axis, so  $z(a) = (A, 0)$ ,  $z(b) = (B, 0)$  with  $A < B$ , so  $\Delta = B - A$ . If  $\Delta = 0$ , the theorem is obvious. Otherwise, for each point  $x \in [A, B]$  there is  $t(x)$  such that if  $z_1(t(x)) = x$ , and so  $\Gamma^\delta$  contains  $x \times [z_2(t(x)) - \delta, z_2(t(x)) + \delta]$ , which has length  $2\delta$ . Thus by Fubini's theorem,

$$|\Gamma^\delta| = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi_{\Gamma^\delta}(x, y) dx dy \geq \int_A^B 2\delta = 2\delta\Delta$$

so the theorem is proved.  $\square$

**Theorem 3.33.** If  $\Gamma = \{z(t) : a \leq t \leq b\}$  is a quasi-simple curve (simple except at finitely many points), then the Minkowski content of  $\Gamma$  exists if and only if  $\Gamma$  is rectifiable, and in this case  $M^1(\Gamma)$  is the length of the curve  $L$ .

*Proof.* To prove the theorem, we consider the upper and lower Minkowski contents

$$M^*(\Gamma) = \limsup_{\delta \rightarrow 0} \frac{|\Gamma|^\delta}{\alpha(n-1)\delta} \quad M_*(\Gamma) = \liminf_{\delta \rightarrow 0} \frac{|\Gamma|^\delta}{\alpha(n-1)\delta}$$

First, we prove that  $M^*(\Gamma) \leq L$ . Consider a partition  $P$  of  $[a, b]$ , and let  $L_P$  be the length of the polygonal approximation to the curve. By refining the partition, we may assume that  $\Gamma$  is simple, with the repeated points at the boundaries of the intervals. For each interval  $I_n$  in the partition, we select a closed subinterval  $J_n = [t_n, u_n]$  such that  $\Gamma$  is simple on  $\bigcup J_n$ , and

$$\sum |z(u_n) - z(t_n)| \geq L_P - \varepsilon$$

Since the intervals  $J_n$  are disjoint, for suitably small  $\delta$  the sets  $J_n^\delta$  are disjoint. Applying the previous lemma, we conclude that

$$|\Gamma^\delta| \geq \sum |J_n^\delta| \geq 2\delta \sum |z(u_n) - z(t_n)| = 2\delta(L_P - \varepsilon)$$

First, by letting  $\varepsilon \rightarrow 0$  and then  $\delta \rightarrow 0$ , we conclude that  $M_*(\Gamma) \geq \lim_P L_P$ . In particular, this shows that if  $\Gamma$  has Minkowski content one, then the curve is rectifiable. Conversely, we consider the functions

$$F_n(s) = \sup_{0 < |h| < 1/n} \left| \frac{z(s+h) - z(s)}{h} - z'(s) \right|$$

Because  $z$  is continuous, this supremum can be considered over a countable, dense subset, and so each  $F_n$  is measurable. Since  $F_n(s) \rightarrow 0$  for almost every  $s$ , we can apply Egorov's theorem to show that this limit is uniform except on a singular set  $E$  with  $|E| < \varepsilon$ , so that for some large  $N$ , for  $s \notin E$  and  $|h| < 1/N$ ,  $|z(s+h) - z(s) - hz'(s)| < \varepsilon h$ . We now split the interval  $[a, b]$  into consecutive intervals  $I_1, \dots, I_{M+1}$ , with each interval but  $I_{M+1}$  having length  $1/N$ . We let  $\Gamma_n$  denote the section of the curve travelled along the interval  $I_n$ . Thus  $|\Gamma^\delta| \leq \sum |\Gamma_n^\delta|$ . If an interval  $I_n$  contains an element of  $E^c$ , we say  $I_n$  is a 'good' interval. Then we can pick an element  $x_n \in I_n$  for which for any  $x \in I_n$ ,

$$|z(x) - z(x_n) - (x - x_n)z'(x_n)| < \varepsilon|x - x_n| < \varepsilon/N$$

Thus  $\Gamma_n$  is covered by a  $\varepsilon/N$  thickening of a length  $1/N$  line  $J_n$  in  $\mathbf{R}^2$  through  $z(x_n)$  with slope  $z'(x_n)$ . Thus if  $\varepsilon \leq 1$ , we conclude

$$\begin{aligned} |\Gamma_n^\delta| &\leq J_n^{\varepsilon/N + \delta} \leq (1/N + 2\varepsilon/N + 2\delta)(2\varepsilon/N + 2\delta) \\ &\leq 2\delta/N + O(\delta\varepsilon/N + \delta^2 + \varepsilon/N^2) \end{aligned}$$

Since  $M \leq NL$ , if we take the sum of  $|\Gamma_n^\delta|$  over all ‘good’ intervals we obtain an upper bound of

$$NL(2\delta/N + O(\delta\varepsilon/N + \delta^2 + \varepsilon/N^2)) = 2\delta L + O(\delta\varepsilon + \delta^2 N + \varepsilon/N)$$

On the other hand, if  $I_n$  is contained within  $E$ , or if  $n = M + 1$ , we say  $I_n$  is a bad interval. Since  $E$  has total measure bounded by  $\varepsilon$ , there can be at most  $\varepsilon N + 1$  bad intervals. On these intervals we use the crude estimate  $|z(t) - z(u)| \leq |t - u|$  (true because  $z$  is an arclength parameterization) to show  $\Gamma_n$  is contained in a rectangle with sidelengths  $1/N$ , so we obtain that  $|\Gamma_n^\delta| \leq (1/N + 2\delta)^2 = O(1/N^2 + \delta^2)$ . Thus the sum of  $|\Gamma_n^\delta|$  over the ‘bad intervals’ is bounded by

$$O(\varepsilon/N + 1/N^2 + \varepsilon N \delta^2 + \delta^2)$$

In particular, the sum of the two bounds gives

$$|\Gamma^\delta| \leq 2\delta L + O(\delta\varepsilon + \delta^2 N + \varepsilon/N + 1/N^2)$$

Or

$$\frac{|\Gamma^\delta|}{2\delta} \leq L + O(\varepsilon + \delta N + \varepsilon/N + 1/N^2)$$

If we choose  $N \geq 1/\delta$ , we get that

$$\frac{|\Gamma^\delta|}{2\delta} \leq L + O(\varepsilon + \delta N + \delta\varepsilon) = L + O(\varepsilon + \delta N)$$

Letting  $\delta \downarrow 0$ , we conclude that  $M^*(\Gamma) \leq L + O(\varepsilon)$ , and we can then let  $\varepsilon \downarrow 0$  to conclude  $M^*(\Gamma) \leq L$ . This completes the proof that if  $\Gamma$  is rectifiable, then  $\Gamma$  has one dimensional Minkowski content, and  $M(\Gamma) = L$ .  $\square$

*If  $\Gamma$  is rectifiable, it is parameterizable by a Lipschitz map (the arclength parameterization). If we instead consider a curve parameterizable by a map  $z$  which is Lipschitz of order  $\alpha$ , which may no longer be absolutely continuous, but still has a decay very similar to the Minkowski dimension decay.*

**Theorem 3.34.** *If  $z$  is a planar curve which is Lipschitz of order  $\alpha > 1/2$ , then its trace  $\Gamma$  satisfies  $|\Gamma^\delta| = O(\delta^{2-1/\alpha})$ .*

*Proof.* Since  $|z(t) - z(s)| \leq |t - s|^\alpha$ , we can cover  $z$  by  $O(N)$  radius  $1/N^\alpha$  balls, so  $|\Gamma| \lesssim N^{1-2\alpha}$ , and so  $|\Gamma^\delta| \lesssim N(1/N^\alpha + \delta)^2$ . Setting  $N = \delta^{-1/\alpha} + O(1)$  gives  $|\Gamma^\delta| \lesssim \delta^{2-\alpha-1/\alpha}$ .  $\square$

### 3.12 The Isoperimetric Inequality

We now use our Minkowski content techniques to prove the isoperimetric inequality, which asks us to find the region in the plane with largest area whose boundary has a bounded length  $L$ . We suppose  $\Omega$  is a bounded region of the plane, whose boundary  $\partial\Omega$  is a rectifiable curve with length  $L$ . In particular, we shall find the region with the largest area whose boundary has a fixed length are balls. A key inequality used in the proof is the Brun Minkowski inequality, which lowers bounds the measure of  $A + B$  in terms of  $A$  and  $B$ . If we hope for an estimate  $|A + B|^\alpha \gtrsim |A|^\alpha + |B|^\alpha$ , then taking  $B = \alpha A$ , where  $A$  is convex and, for which  $A + \alpha A = (1 + \alpha)A$ , we find  $(1 + \alpha)^{d\alpha} \gtrsim (1 + \alpha^{d\alpha})$ . Thus  $\alpha \geq 1/d$ .

**Lemma 3.35.** *If  $A$ ,  $B$ , and  $A + B$  are measurable,  $|A + B|^{1/d} \geq |A|^{1/d} + |B|^{1/d}$ .*

*Proof.* Suppose first that  $A$  and  $B$  are rectangles with side lengths  $x_n$  and  $y_n$ . Then the Minkowski inequality becomes

$$\left(\prod (x_n + y_n)\right)^{1/d} \geq \left(\prod x_n\right)^{1/d} + \left(\prod y_n\right)^{1/d}$$

Replacing  $x_n$  with  $\lambda_n x_n$  and  $y_n$  with  $\lambda_n y_n$ , we find that we may assume  $x_n + y_n = 1$ , and so we must prove that for any  $x_n \leq 1$ ,

$$\left(\prod x_n\right)^{1/d} + \left(\prod (1 - x_n)\right)^{1/d} \leq 1$$

But this inequality is an immediate consequence of the arithmetic geometric mean inequality. Thus the case is proved. Next, we suppose  $A$  and  $B$  are unions of disjoint closed rectangles, and we prove the inequality by induction on the number of rectangles. Without loss of generality, by symmetry in  $A$  and  $B$ , we may assume that  $A$  has at least two rectangles  $R_1$  and  $R_2$ . Since the inequality is translation invariant separately in  $A$  and  $B$ , and  $R_1$  and  $R_2$  is disjoint, hence separated by a coordinate axis, we may assume there exists an index  $j$  such that every element  $x$  of  $R_1$  has  $x_j < 0$  and every element  $x$  of  $R_2$  has  $x_j > 0$ . Let  $A^+ = A \cap \{x_j \leq 0\}$  and  $A^- = A \cap \{x_j \geq 0\}$ . Next, we translate  $B$  such that if  $B^\pm$  are defined similarly, then

$$\frac{|B^\pm|}{|B|} = \frac{|A^\pm|}{|A|}$$

Note that  $A + B$  contains the union of  $A^+ + B^+$  and  $A^- + B^-$ , and this union is disjoint. Thus by induction,

$$\begin{aligned}
|A + B| &\geq |A^+ + B^+| + |A^- + B^-| \\
&\geq (|A^+|^{1/d} + |B^+|^{1/d})^d + (|A^-|^{1/d} + |B^-|^{1/d})^d \\
&= |A^+| \left( 1 + \left( \frac{|B^+|}{|A^+|} \right)^{1/d} \right)^d + |A^-| \left( 1 + \left( \frac{|B^-|}{|A^-|} \right)^{1/d} \right)^d \\
&= (|A|^{1/d} + |B|^{1/d})^d
\end{aligned}$$

Thus the proof is completed for unions of rectangles. The proof then passes to open sets by approximating open sets by closed rectangles contained within. Then we can pass to where  $A$  and  $B$  are compact sets, since then  $A + B$  is compact, and so if we consider the open thickenings  $A^\varepsilon$ ,  $B^\varepsilon$ , and  $(A + B)^\varepsilon$ , then

$$|A| = \lim |A^\varepsilon| \quad |B| = \lim |B^\varepsilon| \quad |A + B| = \lim |(A + B)^\varepsilon|$$

and  $(A + B)^\varepsilon \subset A^\varepsilon + B^\varepsilon \subset (A + B)^{2\varepsilon}$ . Finally, we can use inner regularity to obtain the theorem in full.  $\square$

**Theorem 3.36.** *For any region  $\Omega$ ,  $4\pi|\Omega| \leq L^2$ .*

*Proof.* For  $\delta > 0$ , consider

$$\Omega_+(\delta) = \{x : d(x, \Omega) < \delta\} \quad \Omega_-(\delta) = \{x : d(x, \Omega^c) \geq \delta\}$$

Then we have a disjoint union  $\Omega_+(\delta) = \Omega_-(\delta) + \Gamma^\delta$ , where  $\Gamma$  is the boundary curve of  $\Omega$ . Furthermore,  $\Omega_+(\delta)$  contains  $\Omega + B_\delta$ , and  $\Omega$  contains  $\Omega_-(\delta) + B_\delta$ . Applying the Brun Minkowski inequality, we conclude

$$|\Omega_+(\delta)| \geq (|\Omega|^{1/2} + \pi^{1/2}\delta)^2 \geq |\Omega| + 2\pi^{1/2}\delta|\Omega|^{1/2}$$

$$|\Omega| \geq (|\Omega_-(\delta)|^{1/2} + \pi^{1/2}\delta)^2 \geq |\Omega_-(\delta)| + 2\pi^{1/2}\delta|\Omega_-(\delta)|^{1/2}$$

But

$$|\Gamma^\delta| = |\Omega_+(\delta)| - |\Omega_-(\delta)| \geq 2\pi^{1/2}\delta \left( |\Omega|^{1/2} + |\Omega_-(\delta)|^{1/2} \right)$$

Dividing by  $2\delta$  and letting  $\delta \rightarrow 0$ , we conclude  $L \geq 2\pi^{1/2}|\Omega|^{1/2}$ . This is precisely the inequality we need.  $\square$

Using some Fourier analysis, we can prove that the only smooth curves which make this inequality tight are circles. Indeed, if a closed  $C^1$  curve  $\Gamma = \{z(t) : a \leq t \leq b\}$  is given, then Green's theorem implies the area of its interior is given by

$$\frac{1}{2} \left| \int_{\Gamma} x \, dy - y \, dx \right| = \frac{1}{2} \left| \int_a^b x(t)y'(t) - y(t)x'(t) \, dt \right|$$

We then take a Fourier series in  $x$  and  $y$ .

**Theorem 3.37.** *The only curves  $\Gamma$  with rectifiable boundary such that  $A = \pi(L/2)^2$  are circles.*

*Proof.* By normalizing, we may assume  $z$  is an arcline parameterization, and  $\Gamma$  has length  $2\pi$ , so  $z : [0, 2\pi] \rightarrow \mathbf{R}^2$ , and  $z$  is absolutely continuous. If  $x(t) \sim \sum a_n e^{nit}$  and  $y(t) \sim \sum b_n e^{int}$ , then  $x'(t) \sim \sum i n a_n e^{nit}$  and  $y'(t) \sim \sum i n b_n e^{int}$ . Parseval's equality implies

$$\int_0^{2\pi} x(t)y'(t) - y(t)x'(t) \, dt = 2\pi i \sum n(b_n \overline{a_n} - a_n \overline{b_n})$$

Thus the area of the curve is precisely

$$\pi \left| \sum n(b_n \overline{a_n} - a_n \overline{b_n}) \right| \leq \pi \sum 2n |b_n a_n| \leq \pi \sum |n|(|a_n|^2 + |b_n|^2)$$

On the other hand, the length constraint implies that, since  $|z'(t)| = 1$ ,

$$1 = \frac{1}{2\pi} \int_0^{2\pi} x'(t)^2 + y'(t)^2 \, dt = \sum |n|^2(|a_n|^2 + |b_n|^2)$$

If  $A = \pi$ , then

$$\sum |n|(|a_n|^2 + |b_n|^2) \geq 1 = \sum |n|^2(|a_n|^2 + |b_n|^2)$$

This means we cannot have  $|n| < |n|^2$  whenever  $a_n$  or  $b_n$  is nonzero. Thus the Fourier support of  $x$  and  $y$  is precisely  $\{-1, 0, 1\}$ . Since  $x$  is real valued,  $a_1 = \overline{a_{-1}} = a$ ,  $b_1 = \overline{b_{-1}}$ . We thus have  $2(|a_1|^2 + |b_1|^2) = 1$ , and since we must have  $a$  a scalar multiple of  $b$  so the Cauchy Schwarz inequality application becomes an equality, we must have  $|a_1| = |b_1| = 1/2$ . If  $a_1 = e^{i\alpha}/2$  and  $b_1 = e^{i\beta}/2$ , the fact that  $1 = 2|a_1 \overline{b_1} - \overline{a_1} b_1|$  implies  $|\sin(\alpha - \beta)| = 1$ , hence  $\alpha - \beta = k\pi/2$ , where  $k$  is an odd integer. Thus  $x(s) = \cos(\alpha + s)$ , and  $y(s) = \cos(\beta + s)$ , which parameterizes a circle.  $\square$



## Chapter 4

### Measure Extension

*To construct the Lebesgue measure, we began with an intuitive notion of area over the set of intervals on the real line, and then extended the notion of length to a bigger class of sets, the measurable sets. This chapter attempts to generalize this strategy to work on arbitrary measure spaces. The idea is to take a countably additive measure defined on a family of sets satisfying weaker conditions than that of a  $\sigma$  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  $F_2$ , 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 a 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 func-*

tion. 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 4.1.** *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 4.2** (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 4.3.** *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 5

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

*Proof.* Fix  $y \in Y$ . If  $E \times F$  is an elementary rectangle, then if  $y \in F$ ,  $(E \times F)^y = E$ , and if  $y \notin F$ ,  $(E \times F)^y = \emptyset$ , 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 5.2.** *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 premeasure 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 5.3.** *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 5.4 (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 5.5.** *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 5.6** (Fubini). *If  $f \in L^1(X \times Y)$ , then  $f_x \in L^1(Y)$  for almost all  $x$ ,  $f^y \in L^1(X)$  for almost all  $y$ , and*

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

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

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

Hence we conclude that for almost all  $y$ ,

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

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

## 5.1 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 5.7.** *The Lebesgue measure on  $\mathbf{R}^{n+m}$  is the completion of the product of the Lebesgue measures on  $\mathbf{R}^d$  and  $\mathbf{R}^m$ .*

*Proof.* It is easy to see that the Lebesgue measure on  $\mathbf{R}^{n+m}$  agrees with the product of the Lebesgue measure  $\lambda^n \otimes \lambda^m$  on any set  $E \times F$ , where  $E$  and  $F$  are disjoint unions of rectangles in  $\mathbf{R}^d$  and  $\mathbf{R}^m$ . But this is a  $\pi$  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 5.8.** *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 5.9.** *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  $f \geq 0$ . 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 5.10.** *If  $f$  is  $(\Sigma \otimes \Pi)^*$  measurable functions on  $X \times Y$  such that  $f = 0$  almost everywhere with respect to  $\mu \otimes \pi$ , then for almost all  $x \in X$ ,  $f(x, y) = 0$  for almost all  $y$ . In particular,  $f_x$  is  $\Sigma$  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  $\mu \otimes \lambda$ .*

## 5.2 Integration in Polar Coordinates

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

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

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

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

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

is Lebesgue measurable. We then define the surface measure  $\sigma$  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 5.11.** *If  $\pi : (0, \infty) \times S^{d-1} \rightarrow \mathbf{R}^d$  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 5.12.** 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

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

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 6

## Integration

**Lemma 6.1** (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  $\liminf$ s, we can never gain mass via the  $\liminf$  process.*

**Lemma 6.2** (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 6.1 (Tao).** *Use the dominated convergence theorem to prove that the harmonic series diverges, and to understand the rate of divergence of the harmonic series.*

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



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 7

# 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 7.1.** *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  $\bigcup (a_i, b_j]$ . Fix  $\varepsilon > 0$ . If we use the right continuity of  $F$  to pick  $b'_j > b_j$  such that  $F(b'_j) - F(b_j) \leq \varepsilon/2^j$ , then

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

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

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

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 7.2.** If  $F$  is absolutely continuous, then

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

for all  $f$  integrable with respect to  $F$ .

*Proof.* Suppose that  $F$  is increasing, and let  $\mu$  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 7.3** (Integration by Parts). If  $F$  and  $G$  have finite variation, then

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

*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

$$\begin{aligned} \int_a^b [F(x) - F(a)] dG(x) &= \int_a^b \int_a^y dF(x) dG(y) \\ \int_a^b [G(x) - G(a)] dF(x) &= \int_a^b \int_a^x dG(y) dF(x) \end{aligned}$$

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 7.4.** *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

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

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  $f \in C^1(\mathbf{R}^d)$ , 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.

## 7.1 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 7.5.**  $|\mu|$  is a positive measure on  $X$  with  $|\mu|(X) < \infty$ .

*Proof.* Let  $E = \bigcup E_i$ , where the  $E_i$  are disjoint. Let  $t_1, t_2, \dots$  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 7.6.** 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 7.7.**  $|\mu|(X) < \infty$ .

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

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 7.8.** 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 7.9.** *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  $L^1(X)$ . □

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 7.10 (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

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

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

□